Multi-gradient Permutation Survival Analysis Identifies Mitosis and Immune Signatures Steadily Associated with Cancer Patient Prognosis

Xinlei Cai[1], Yi Ye[2], Xiaoping Liu[1], Zhaoyuan Fang[4,5], Luonan Chen[1,2,6], Fei Li[3*], Hongbin Ji[1,2,6*]

[1]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, 310024, China.

[2]State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China.

[3]Department of Pathology and Frontier Innovation Center, School of Basic Medical Sciences, Fudan University, Shanghai, China.

[4]Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Haining, 314400, China

[5]The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310000, China

[6]School of Life Science and Technology, Shanghai Tech University, Shanghai 200120, China.

Running title: Mitosis and immune signatures in predicting cancer patient prognosis

***Corresponding author:**

Hongbin Ji, State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai, 200031, China, Phone: 86-21-54921108, Fax: 86-21-54921101, E-mail address: hbji@sibcb.ac.cn;

Fei Li, Department of Pathology and Frontier Innovation Center, School of Basic Medical Sciences, Fudan University, 131 Dong An Road, Shanghai, 200032, China, E-mail address: li_fei@fudan.edu.cn.

Abstract

The inconsistency of the association between genes and cancer prognosis is often attributed to many variables that contribute to patient survival. Whether there exist the Genes Steadily Associated with Prognosis (GEARs) and what their functions are remain largely elusive. Here we developed a novel method named "Multi-gradient Permutation Survival Analysis " (MEMORY) to screen the GEARs by using RNA-seq data from the TCGA database. We employed a network construction approach to identify hub genes from GEARs, and utilized them for cancer classification. In the case of lung adenocarcinoma (LUAD), the GEARs were found to be related to mitosis. Our analysis suggested that LUAD cell lines carrying *PIK3CA* mutations exhibit increased drug resistance. For breast invasive carcinoma (BRCA), the GEARs were related to immunity. Further analysis revealed that *CDH1* mutation might regulate immune infiltration through the EMT process. Moreover, we explored the prognostic relevance of mitosis and immunity through their respective scores and demonstrated it as valuable biomarkers for predicting patient prognosis. In summary, our study offered significant biological insights into GEARs and highlights their potentials as robust prognostic indicators across diverse cancer types.

Keywords: Multi-gradient Permutation Survival Analysis (MEMORY), Gene Steadily Associated with Prognosis (GEAR), Hub Genes, Mitosis, Immune

Introduction

Cancer is one of the leading causes of death worldwide, with over 200 types identified. Despite major advances in the field, accurately predicting cancer patient prognosis remains a significant challenge[1,2]. Previous studies have highlighted the complexity of this task, which is influenced by various factors, including cancer types, clinical stages, therapeutic interventions, nursing care, unexpected comorbidities, and other non-cancer related illnesses that may interplay[3-6]. Furthermore, the gene expression plays a crucial role in predicting cancer patient prognosis, such as *HER2*, *VEGF*, *Ki67*, etc[7-12]. Nevertheless, there is still a need for further improvement in prognostic accuracy to better inform treatment decisions and patient outcomes.

Survival analysis is commonly used to assess the correlation between genes and cancer patient prognosis[13]. However, inconsistent findings are frequently observed, even within the same type of cancer. For instance, studies exploring the correlation between *CCND1* and prognosis in non-small cell lung cancer (NSCLC) reported contrasting results, including positive correlation, inverse correlation, or negligible influence[14-16]. These discrepancies can be attributed to various influential factors, such as differences in sample size, cohort characteristics (including cancer subtypes and tumor staging) and variations in therapeutic approaches[17]. Such observations raise an important question: whether there exist Genes Steadily Associated with Prognosis (hereafter referred to as GEARs) that correlate with patient outcomes across different conditions, particularly considering the varying sample sizes used in different studies? Affirmative answers to this question would have significant implication not only for the development of more accurate prognostic models, but also for enhancing our understanding of cancer biology.

The Cancer Genome Atlas (TCGA) is a comprehensive cancer genomics project initiated in the United States. It consists of transcriptome data, genomic data, and clinical information pertaining to 33 different cancer types, making it the largest cancer clinical sample database currently available. In this study, we developed a novel method named "Multi-gradient Permutation Survival Analysis" (MEMORY) with the utilization of TCGA RNA-seq database,

and assessed the potential existence of GEARs across 15 cancer types, each comprising a cohort of over 200 patients. Furthermore, we evaluated the prognostic predictive power of these GEARs and explored their potential biological functions in driving cancer progression.

Results

Multi-gradient permutation survival analysis identifies GEARs associated with mitosis and immune across multiple cancer types

GEARs are a group of genes consistently and significantly correlate with cancer patient survival, independent of the sample size. To identify these GEARs, we developed a novel method called "Multi-gradient Permutation Survival Analysis " (MEMORY). This method allows us to assess the correlation between a specific gene and cancer patient prognosis using available transcriptomic datasets (**Figure 1A, Table 1**). Initially, we started with a sampling number at 10% of the cohort size, then gradually increased the sampling number with each 10% increment, which was analyzed with 1000 permutations. By calculating the statistical probability of each gene's association with patient survival, we can identify a group of GEARs for further analyses.

In this study, we utilized the TCGA datasets for several reasons, including the diversity of cancer types, the availability of gene expression profiles and patient prognosis information. We set a minimum cohort size of 200 and included 15 eligible cancer types for analysis. These cancer types include bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon adenocarcinoma (COAD), kidney renal papillary cell carcinoma (KIRC), brain lower grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). As the number of samples increases, the significant probability of certain genes gradually approaches 1. Once this score reaches 0.8, previously employed as an empirical standard for survival analyses, and remained above this value consistent with further sample gradient increase, the gene is considered a GEAR (**Figure 1B**)[18]. Remarkably, we successfully identified a set of GEARs across all 15 cancer types (**Supplementary Figure 1-2 , Table 2)**. The GEAR counts in most cancer types ranged from 100 to1000, with exceptions of CESC,

KIRC, LGG, and PAAD with over 1000 GEARs, and THCA with only 22 GEARS (**Table 2)**. In LUAD, the top 10 GEARs with the highest significance probabilities were *TLE1*, *GNG7*, *ERO1A*, *ANLN*, *DKK1*, *TMEM125*, *S100A16*, *KNL1*, *STEAP1*, and *BEX4*. Most of these genes are known to promote LUAD malignant progression except for S100A16[19-27]. In other cancer types, *BEX4* was identified as a common GEAR in KIRC, LGG, PAAD, and STAD (Table 2). *BEX4* is reported as a proto-oncogene promoting cancer onset and malignant progression in multiple cancers, including LUAD, glioblastoma multiforme and oral squamous cell carcinoma[26,28,29]. We also identified the top 1 GEAR in individual cancer types, such as *TLL1* (BLCA), *PGK1* (BRCA), *RFXANK* (CESC), *DPP7* (COAD), *VWA8* (KIRC), *SCMH1* (LGG), *HILPDA* (LIHC), *TLE1* (LUAD), *CD151* (LUSC), *ANKRD13A* (OV), *SOCS2* (PAAD), *DRG2* (PRAD), *ADAMTS6* (STAD), *PSMB8* (THCA), *ASS1* (UCEC) (**Table 3).** Many of these genes were previously reported to be associate with tumorigenesis[19,30-41]. For example, *TLE1* is known as a transcriptional repressor that promotes cell proliferation, migration, and inhibits apoptosis in LUAD[42,43]. Additionally, *PGK1*, a key enzyme in the glycolytic process, has been shown to promote cell proliferation, migration, and invasion in multiple cancers[44,45]. These findings demonstrate the intricate link between the functionality of GEARs and the initiation and progression of cancer.

To gain deep insights into the biological functions of these identified GEARs, we conducted Gene Ontology (GO) enrichment analysis (**Figure 1C**). Interestingly, we found the mitosis-related pathways were enriched in LICH, LUAD, LGG, and PAAD, and the immune-related pathways were enriched in BRCA and UCEC[46]. Additionally, other cancer types exhibited enrichment in various pathways, such as organic acid metabolism pathway in KIRC, oxidative phosphorylation pathway in CESC, organ development-related pathways in LUSC, neurogenesis-related pathways in BLCA, and sustenance metabolism pathways in THCA. Given the crucial role of mitosis in cancer progression and the significance of the immune system in cancer-host interactions, we specifically focused on the mitosis and immune-related GERAs, which accounted for approximately 40% of all the analyzed cancers.

Identification of hub genes in mitosis and immune-related cancers

To identify crucial genes within the GEARs, we conducted and extracted the higher-ranked

edges to construct the core survival network (CSN) (**Figure 2A, Table 4**). The top 10 GEARs with the highest degree in these networks, selected from those with mean expression >10 TPM, were defined as hub genes[47]. We then classified the samples using the hub genes derived from these networks and evaluated their clinical relevance (**Supplementary Figure 3A-O, Supplementary Figure 4A-O**). A significant correlation with cancer stages (TNM stages) is observed in most cancer types except for LUSC, THCA and PAAD (**Supplementary Figure 5A-K**). Furthermore, we conducted GO enrichment analysis on the hub genes selected from CSNs and found that the results were consistent with the GEARs analysis (**Table 5**). Specifically, the hub genes in LIHC, LGG, and LUAD were enriched in mitosis-related pathways, whereas the hub genes in BRCA and UCEC were enriched in immune-related pathways (**Figure 2B**). For instance, the 9 hub genes in LUAD were associated with functions related to mitosis, whereas the 8 hub genes of BRCA were associated with immune-related functions (**Figure 2C-D**).

Next, we conducted survival-dependent analyses on the hub genes of LGG, LIHC, and LUAD. These analyses revealed that mitosis-related hub genes were closely associated with cancer cell viability, especially those hub genes that were correlated with multiple cancer types, such as *CDC20*, *TOP2A*, *BIRC5* and *TPX2*[48-51] (**Supplementary Figure 6A-C**). The importance of these genes is well established. For instance, *TPX2*, a hub gene in all three cancers, is known to play a crucial role in normal spindle assembly during mitosis and is essential for cell proliferation[52]. The significance of these hub genes for the survival of cancer cells suggests that the expression levels of these hub genes could be used for inhibitors screening of tumor growth. By integrating the Genomics of Drug Sensitivity in Cancer (GDSC) and Connectivity Map (cMAP) databases, we identified 76 individual compounds that were able to effectively suppress the expression of the 10 hub genes in LUAD cell lines (**Figure 2E**, **Supplementary Figure 6D-E**). These compounds also significantly inhibited LUAD cell growth and might serve as potential therapeutic agents for LUAD.

We then focused on our analysis on LUAD dataset, which had a larger sample size compared to LGG and LIHC. According to the expression of the hub genes, we classified the LUAD

samples into three subgroups: mitosis low (ML), mitosis medium (MM), and mitosis high (MH) (**Figure 2F, Table 6**). Previous study has reported three classic subgroups of LUAD known as the terminal respiratory unit (TRU), the proximal-proliferative (PP), and the proximal-inflammatory (PI) subgroups[53]. Our analysis revealed that the ML subgroup was primarily enriched with TRU subgroup and a small number of samples from the PP subgroup, but not the PI subgroup. The MH subgroup showed high expression of mitosis-related genes and predominantly encompassed PI and PP subgroups, whereas the MM group exhibited intermediate levels of mitosis-related gene expression (**Figure 2F-G**). This suggests that the new categorization method may help identify new factors that influence patient prognosis (**Supplementary Figure 6F-G**).

Distinct genetic mutation landscapes characterize different clusters of LUAD

The expression of hub genes provided crucial indicators for distinguishing various subgroups of LUAD. However, the underlying mechanisms driving these variations in expression patterns remain unknown. In our subsequent research, we aimed to explore the genomic differences, particularly in terms of gene mutations, among different LUAD subgroups. Initially, we analyzed the variations of genetic mutation landscapes of different LUAD subgroups by assessing the tumor mutation burden (TMB). Interestingly, we found significant differences in TMB among the various subgroups of LUAD (**Figure 3A**). Common driver genes were heterogeneously distributed across the three molecular subgroups. *ALK* and *ROS1* fusions were most enriched in the prognosis-favorable ML and MM clusters, while *KRAS*, *EGFR*, *BRAF* and *ERBB2* mutations were more evenly distributed or enriched in the less prognosis-favorable MH group (**Figure 3B**). Due to limited counts for some fusion events, formal statistical comparison was not performed, but the distributional patterns suggested distinct subtype preferences across different driver events. Moreover, low tumor mitotic activity was associated with better prognosis in LUAD subgroups with *EGFR* mutations or pan-negative (no oncogenic alteration in genes including *KRAS*, *EGFR*, *BRAF*, *ERBB2*, *PIK3CA*, *ALK* and *ROS1*) but not in those with *KRAS* or *BRAF* mutations (**Figure 3C-D, Supplementary Figure 6H-I**). Besides, we observed unique mutation characteristics in each of these three subgroups (**Figure 3E-G**). For example, *TP53* mutation, a prevalent tumor

suppressor gene mutation, was frequently observed in the MH subgroup with a mutation rate of 73%, compared to mutation rates of 14% in the ML subgroup and 39% in the MM subgroup. While genes, such as *CSMD3*, *RP1L1* and *ZFHX4*, exhibited similar trend in mutation frequency across the three subgroups. These findings indicate that substantial genomic differences among these three LUAD subgroups are based on the hub genes.

We identified gene mutations that showed significant changes through gene dependence analysis (**Figure 4A**). To further explore the functional implications of these mutations, we enriched them using a pathway system called Nested Systems in Tumors (NeST)[54]. The results revealed notable differences in multiple functional pathways, including androgen receptor, cell cycle, *TP53*, *EGFR*, IL-6, and *PIK3CA*-related pathways (**Figure 4B**). To gain further insights into the functional implications of these different mutations, we assessed the gene dependency of cells with these mutations by using the DepMap database. Importantly, we observed that the pathway differences between the MM and MH clusters were primarily associated with *PIK3CA*-related pathway. Previous studies have reported that *PIK3CA* mutation confers resistance in colorectal cancer, lung cancer, and breast cancer[55-57]. Therefore, we aimed to validate the role of *PIK3CA* mutations in drug resistance by using public databases. We analyzed two lung adenocarcinoma cell lines, A549 cells and SW1573, harboring *KRAS* mutation and both *KRAS* and *PIK3CA* mutations, respectively, and identified five potentially effective inhibitors (BMS-345541, Dactinomycin, Epirubicin, Irinotecan, and Topotecan) from the previously screened set of 76 LUAD cell growth inhibitors by using GDSC data. Strikingly, SW1573 cells exhibited increased resistance to all these 5 inhibitors when compared to A549 cells (**Figure 4C**). In line with this, clinical data supported the notion that concurrent *PIK3CA* mutation is a poor prognostic factor for LUAD patients[58]. These findings suggest that *PIK3CA* mutation may contribute to drug resistance in LUAD.

Distinct clusters of BRCA exhibit different immune infiltration landscapes

In the meantime, we conducted a comprehensive analysis on BRCA, which has the largest sample size among immune-related cancers (**Figure 1C**). We hierarchically grouped the BRCA into three distinct immune subgroups based on the expression of hub genes: immune low (IL), immune medium (IM), and immune high (IH) (**Figure 5A, Table 7**). Interestingly,

each PAM50 subtype included tumors from the IL, IM, and IH immune subgroups, with only minor differences in their relative frequencies[59] (**Figure 5B, Supplementary Figure 7A-B**). Next, we investigated the relationship among PAM50 classification, immune subtypes, and patient prognosis. Our data revealed significant prognostic differences among immune subgroups of the luminal B subtypes but not the luminal A, basal or HER2 subtypes (**Supplementary Figure 7C-F**), highlighting that integrating our method with traditional classification may enable a more detailed stratification of breast cancer samples. Mutation analysis showed that distinct patterns of gene mutations among the three subgroups had significant differences in mutation frequencies of genes such as *TP53*, *CDH1*, and *LRP2* (**Figure 5C-E**). We conducted a comparison of the proportions of immune cells across these three subgroups and observed a strong association between the overall immune cell proportion and the clustering results based on hub genes (**Figure 5F, Supplementary Figure 8A-J**). Specifically, the IH subgroup exhibited significantly higher proportions of $CD8^+$ T cells and Treg cells, with median percentages at 3.9% and 5.5%, respectively, compared to 1.2% and 2.8% in the IM subgroup, and 0.2% and 1.6% in the IL subgroup. Together, these findings suggest that distinct immune responses across the three subgroups potentially link genetic mutation patterns to differences in the immune microenvironment.

Next, we explored the specific genomic factors influencing immune infiltration in BRCA. Mutation analysis revealed that *CDH1* gene mutation ranked as the top two genetic alteration in BRCA samples, following *TP53* mutation (**Figure 5C-E**). Previous study has reported that *CDH1* is involved in the mechanisms regulating cell-cell adhesions, mobility, and proliferation of epithelial cells[60]. We found that *CDH1*-mutant samples exhibited significantly lower *CDH1* expression compared to *CDH1*-wildtype samples, and *CDH1* expression showed a close correlation with immune cell infiltration (**Figure 5G, Supplementary Figure 8K**). These results were consistent with clinical observations of *CDH1* mutation and high immune infiltration in invasive lobular carcinoma of the breast[61].

The above result encouraged us to investigate how *CDH1* mutation influenced immune infiltration. In BRCA, we observed a significant correlation between the expression of *CDH1* and EMT marker genes *VIM* and *TWIST2* (**Supplementary Figure 8L-M**)[62,63]. This suggests

that the regulation of *CDH1* is intricately linked with EMT process. EMT score analysis showed that there's a positive correlation between the EMT score and the proportion of immune infiltration (**Figure 5H**), which suggests that *CDH1* might influence immune infiltration through the EMT process.

Mitotic and immune signatures predict patient prognosis at pan-cancer level

Finally, we analyzed the association of *CDH1* and *PIK3CA* with specific biological processes at pan-cancer level. Our data showed that *PIK3CA* level was positively correlated with mitotic scores in most of the cancer types, whereas the correlation between *CDH1* expression and the proportion of immune cell infiltration was positive in PRAD, LGG, OV, Uveal Melanoma (UVM), THCA, LIHC and LUAD but negative in BRCA, Testicular Germ Cell Tumors (TGCT), Thymoma (THYM), LUSC, BLCA, Head and Neck squamous cell carcinoma (HNSC), Sarcoma (SARC), Esophageal carcinoma (ESCA), PAAD and STAD (**Figure 6A-B**). Eventually, we sought to explore the prognostic relevance of mitosis and immune to patient outcomes at the pan-cancer level. The scores for these biological processes were computed using RNA-seq data from the TCGA database, and the median score was used as a cut-off to categorize patients into different subgroups. Out of 33 cancer types, 19 showed a statistically significant correlation ($p < 0.05$) with at least one pathway score (**Figure 6C**). Specifically, 10 cancer types were exclusively associated with the mitosis score, 4 cancer types were exclusively associated with the immune score, and 5 cancer types exhibited a correlation with both mitosis and immune scores simultaneously. Overall, the identification of mitosis and immune-related biological processes as significant prognostic factors at the pan-cancer level suggests their potential utility as valuable biomarkers for predicting patient prognosis.

Discussion

Due to the multifaceted nature of variables affecting cancer patient prognosis, it remains uncertain whether there exist a set of genes steadily associated with cancer prognosis , regardless of sample size and other factors. Here, we utilized the MEMORY method to address this question and discovered that all the cancer types have GEARs. We observed

significant variation in the number of GEARs among different cancer types, indicative of cancer type-specific patterns. The substantial heterogeneity in driver genes and mortality rates among various cancer types could potentially explain this phenomenon. For example, THCA, known for its low malignancy, displays the fewest genetic expression alterations among all the studied cancer types. This observation could be correlated with the relatively high five-year survival rate in THCA, which exceeds 50% even in advanced stage[64]. In contrast, PRAD, despite of a favorable prognosis similar to THCA, exhibits a significantly higher number of genetic expression alterations. We hypothesize that the positive prognosis in PRAD cases might be largely attributed to early diagnosis, which enables timely treatment for the majority of PRAD patients[65]. This discrepancy between the number of genetic alterations and prognosis in THCA and PRAD highlights the intricate nature of cancer genetics and emphasizes the importance of personalized considerations in cancer treatment and prognosis evaluation. Furthermore, certain genes are commonly found in the GEARs of various cancer types, indicating their potential significance in cancer development. For instance, *BEX4* is present in the GEARs of LUAD, KIRC, LGG, PAAD, and STAD, and known to play oncogenic role in inducing carcinogenic aneuploid transformation via modulating the acetylation of α-tubulin[66,67]. Aneuploidy is a hallmark characteristic of cancer and can lead to alterations in the dosage of oncogenes or tumor suppressor genes, thereby influencing tumor initiation and progression[68-71]. The regulation of aneuploid transformation by *BEX4* may represent a common mechanism through which this gene impacts prognosis across different cancer types. Subsequently, we discovered that GEARs across different cancers displayed distinct functional characteristics. Notably, a recurrent theme was the prominence of mitosis-related and immune-related features. Specifically, the GEARs of LGG, LIHC, and LUAD were enriched in mitosis-related pathways, whereas BRCA and UCEC showed the enrichment of immune-related pathways. Mitosis and immune processes have been widely recognized as having a significant impact on patient prognosis[72-76]. However, current clinical guidelines do not yet recommend the use of transcriptomic data to assess scores related to mitotic and immune pathways for predicting patient outcomes, despite the well-established association between these pathways and prognosis in cancer[77-82]. Cancers enriched with mitosis pathways often exhibit heterogeneous tumor growth kinetics across individuals, with tumor size being

one of the crucial factors influencing patient survival[83-85]. For instance, in the case of LGG, the growth rate of tumors directly impacts the patient's prognosis due to the primary location in the brain. Consequently, alterations in the expression levels of genes related to mitosis are often indicative of the prognosis of LGG[86]. Cancer types that are enriched in immune-related pathways, such as BRCA and UCEC, are closely associated with deficiencies in DNA mismatch repair (MMR)[87,88]. Cancers enriched in immune-related pathways often exhibit heterogeneous tumor growth kinetics across individuals, where tumor size is closely correlated with patient survival.

To explore the underlying differences between samples associated with distinct mitotic and immune-related pathways, we employed GEAR analysis to identify hub genes for further molecular classification and mutation analysis. Specifically, we focused on LUAD and BRCA, two representative cancer types exhibiting enriched mitosis and immune signatures in GEARs. By classifying the hub gene, we divided LUAD into three subgroups: ML, MM and MH, which displayed significant difference in survival outcomes. Subsequently, we utilized NeST to analyze the mutations within these subgroups. Notably, there were significant differences observed in cell cycle and signal transduction-related pathways among ML, MM and MH subgroups. Of particular importance, the *PIK3CA*-related pathway emerged as a key differentiating pathway between the MH and MM subgroups. Our analyses suggest that LUAD cell line carrying *PIK3CA* mutation may exhibit increased drug resistance. This aligns with previous studies demonstrating that targeting the PI3K pathway can overcome drug resistance[89,90]. Of course, future research is warranted to elucidate the exact functional role of *PIK3CA* mutations in LUAD.

To investigate the factors influencing immune infiltration in BRCA, we classified the samples based on hub genes and analyzed the differentially mutated genes. Interestingly, we found that *CDH1* mutation occurred at a high frequency in the IH subgroups. This led us to speculate that *CDH1* plays a crucial role in the regulation of immune infiltration in BRCA. Furthermore, previous studies have reported that *CDH1* inactivation promotes immune infiltration in breast cancer[61]. *CDH1* is a vital gene associated with EMT, and various studies have demonstrated the close relationship between EMT and the tumor immune microenvironment in different

cancers[91]. Consistently, our results also supported the correlation between EMT score and immune infiltration in BRCA. These findings suggest that the mutations in *PIK3CA* and *CDH1*, identified through GEAR analysis, have significant impacts on cancer development and hold potential value in improving clinical therapies. This further emphasizes the importance of GEARs in understanding cancer biology and guiding treatment strategies.

Lastly, we investigated the prognostic predictive capabilities of the mitosis score and immune score at the pan-cancer level. Surprisingly, we found that approximately half of the cancer types exhibited significant correlations between these two scores and patient prognosis. Interestingly, even for cancers originating from the same primary location, their correlations with these scores could differ, indicating the potential diversity of mechanisms underlying cancer-related mortality. For instance, both LGG and GBM are brain tumors, but the primary risk factor for patient prognosis in LGG is closely linked to tumor diameter, whereas the main risk factor for GBM patients lies in its high invasiveness and challenge of surgical resection[92,93].

Although the present study defined a preliminary catalogue of GEARs and supported their relevance with extensive biological data, we recognize several outstanding methodological limitations at both the algorithmic and analytical levels. Chief among them is the need for rigorous, large-scale multiple-testing adjustment before any GEAR list can be considered clinically actionable. Because this work was conceived as an exploratory screen, such correction was intentionally deferred; nonetheless, forthcoming versions of MEMORY will incorporate a dedicated false-positive–control module that will be applied to the consolidated GEAR catalogue prior to any translational use. Although GEARs show robust prognostic associations, the CSN edges are undirected, and hub genes serve primarily as stable biomarkers. Future work will explore causality and therapeutic implications. As a result, the hub genes identified from GEAR in the CSN may primarily serve as stable and effective biological markers. Additionally, through multi-omics analysis, we obtained some functional mutations, but the therapeutic significance of these mutations remains to be elucidated. For example, further study is needed to understand the exact role of *PIK3CA* mutations in promoting tumor cell proliferation and drug resistance. Similarly, the association of *CDH1*

mutations with the infiltration of multiple immune cell types also warrants additional experimental investigation. In our future studies, we plan to utilize protein-protein interaction networks and pathway databases to construct a novel network based on the CSN. This approach will allow us to directly screen genes from GEARs that could potentially serve as therapeutic targets. Undoubtedly, future efforts are still required to utilize protein-protein interaction networks and pathway databases to construct a new network based on CSN.

In conclusion, our study utilized the MEMORY algorithm to identify GEARs in 15 cancer types, highlighting the significance of mitosis and immunity in cancer prognosis. Our findings demonstrate that GEARs possess substantial biological significance beyond their role as prognostic biomarkers. This study provides valuable guidance for establishing standards for survival analysis evaluation and holds potential for the development of novel therapeutic strategies.

Methods

Datasets

The gene expression profiles and clinical information of 33 cancers were obtained from the TCGA database and downloaded by the GDC data website (https: //portal.gdc.cancer.gov/). All gene expression data and survival data were integrated and normalized.

Multi-gradient permutation survival analysis

RNA-seq count data were normalized by TPM form the gene expression matrix (genes × samples) for each cancer type. We randomly sampled the gene expression data according to the gradient. The sampling strategy was as follows: Ten gradient increases in sample size were pre-set, ranging from approximately 10% to about 100%, with intervals of 10%. Random samples were taken from total samples of each cancer 1000 times, based on each pre-set sample size. Multiple sampling matrix was obtained after the sampling strategy was performed in each gradient. Then survival analysis was performed using the R package “survival”. Gene expression values were dichotomized based on the median expression level

of each gene in the sampled matrix. The survival analysis method was as follows according to the median expression value of every gene, every sampling matrix was divided into a high and low expression group. This approach divides the samples into high and low expression groups with equal size, providing a standardized grouping strategy for survival analysis. We used 1 for significant survival analysis results ($P < 0.05$) and 0 for non-significant results ($P > 0.05$). The survival analysis matrices were obtained after these processes. This generated a binary matrix of shape (genes × 1000) per gradient, representing the significance profile across permutations. The binary matrix was integrated to a significant-probability matrix by a formula:

$$A_{ij} = \frac{\sum_{n=1}^{1000} k_{j \times i \times n}}{1000} \qquad (1)$$

where the A_{ij} is the value from row i and column j in the significant-probability matrix. We defined the sampling size k_j reached saturation when the max value of column j was equal to 1 in a significant-probability matrix. The least value of k_j was selected, and the genes with their corresponding A_{ij} greater than 0.8 were extracted as GEAR.

Construction of core survival network

We also defined survival analysis similarity (*SAS*) as the similarity of the effect on patient prognosis in two genes. For each gene, the 1 000 permutation tests performed at the first sampling gradient yield a binary vector $A = (A_1, \cdots, A_{1000})$, $B = (B_1, \cdots, B_{1000})$ (1 = significant, 0 = non-significant). Vectors from the first sampling gradient (k□, approximately 10% of the cohort) were used as input. At this sample size, significance patterns exhibited high variability, with many genes not yet consistently significant. This variability was utilized to enable SAS to differentiate gene pairs with concordant versus discordant behavior.

Let $a = \sum_{i=1}^{1000} A_i$, $b = \sum_{i=1}^{1000} B_i$, $c = \sum_{i=1}^{1000} A_i B_i$

where c is the number of permutations in which both genes are significant and a, b are the total significant counts for each gene.

SAS is calculated as

$$SAS(A,B) = \frac{c}{a+b-2c+1} \qquad (2)$$

a Jaccard-like metric bounded between 0 (no overlap) and 1 (complete overlap); the "+1" term prevents division by zero.

Pair-wise SAS values were computed between every GEAR and all other genes. The 1 000 SAS were selected to construct a core survival network (CSN) in Cytoscape[94]. Node degrees were calculated, and the top 10 GEARs with the highest degree in these networks, selected from those with mean expression >10 TPM, were defined as hub genes. We constructed the Core Survival Network (CSN) using the top 1000 gene pairs ranked by SAS values. As the CSN was designed as a heuristic similarity network to prioritize genes for exploration, edge selection was based on empirical ranking rather than formal statistical thresholds.

Gene ontology enrichment analysis

Gene ontology has been used to classify genes based on functions. The gene functions were divided three types, includes molecular function (MF), biological process (BP), and cellular component (CC). ClusterProfiler is an R package for gene set enrichment analysis[95]. The gene ontology (GO) enrichment analysis was processed in GEARs and hub genes by ClusterProfiler.

Hub gene classification

Tumor samples were genotyped using RNA-seq data. The following steps outline the genotyping and clustering methods used in this study. The expression matrix of hub genes corresponding to various cancer types was extracted from the RNA-seq data. This involved filtering the RNA-seq data to retain only the expression levels of the identified hub genes. To normalize the expression data and reduce the impact of extreme values, a pseudo count of 1 was added to each expression value, and the resulting matrix M was log-transformed using the formula: $M' = log_2(M+1)$. The log-transformed matrix M' was then subjected to hierarchical clustering. This clustering was performed in R (v4.3) using the Ward's method (hclust, method = "ward.D2") to minimize the variance within clusters. The distance metric used was the Euclidean distance (dist, method = "euclidean"). The number of clusters was set

to 3 using the cutree function (cutree(model, k = 3)), based on visual inspection of the dendrogram structure.

Calculation of tumor mutation burden (TMB) and differential mutation

The TMB and differential mutation gene analysis were carried out to explore the differences between different genotyped samples at the genomic level. The data are from the gene mutation data of TCGA tumor samples, and the grouping information is from the hub gene hierarchical clustering results. Maftools is an R package for the analysis of somatic variant data, which can export results in form of charts and graph[96]. The calculation of TMB and difference mutation analysis were used by the maftools.

Quantifying the effect of gene mutations for tumor cell viability

Gene dependence refers to the extent to which genes are essential for cell proliferation and survival. The Cancer Dependency Map (Depmap, https://depmap.org/portal/) database provides genome-wide gene dependence data for large number of tumor cell lines[97]. Gene mutation data of cell lines were obtained from the CCLE database (https://sites.broadinstitute.org/ccle). The genes appearing in the gene mutation data of cell lines were extracted and sorted into mutation list. Each gene in the list was then analyzed for survival-dependent differences. For each gene in the mutation list, we compared the gene dependence score (S) between cell lines with and without the gene mutation. Specifically, cell lines from the CCLE mutation dataset harboring a mutation in a given gene were classified into one group, while cell lines without the mutation constituted the control group. S for the mutation (S_m) and wild-type (S_{wt}) cell lines were then extracted from the DepMap database.

A two-sample t-test was used to assess the statistical significance of differences between S_m and S_{wt}. The test yielded a P-value, with an alpha level set at 0.05, to determine the significance of the difference in means. D_m was defined as gene mutation dependence, the computational formula was as follows:

$$D_m = |\bar{s}_m - \bar{s}_{wt}| \qquad (3)$$

Subsequently, we conducted a standardization assessment of D_m, employing the following

formula:

$$S_{dm} = \frac{2(\bar{s}_m - \bar{s}_{wt})}{\bar{s}_m + \bar{s}_{wt}} \qquad (4)$$

In this study, we categorized mutations identified in cell lines into three groups based on their impact on gene dependency. Functional mutations refer to the mutations that exhibit significant changes in gene dependency scores, with a P-value < 0.05 and S_{dm} > 0.1. Non-functional mutations were those without significant changes in gene dependency scores compared to wildtype. Subtype-associated functional mutations were a subset of functional mutations that show statistically significant differences in occurrence frequency across different LUAD subtypes.

Identification of drugs downregulating hub gene expression

The IC50 data of 76 compounds of LUAD cell lines were obtained from Genomics of Drug Sensitivity in Cancer (GDSC) database[98]. The hub gene expression data of A549 cell lines after treatment with 76 compounds was obtained from Connectivity MAP (cMAP, https://clue.io/) database, and the data were grouped by hierarchical clustering[99]. The drugs which were presented only in the hub gene expression suppression group were considered to effective against the LUAD cell.

Immune infiltration analysis

Immune infiltration data was calculated by quantiseq method[100]. Immunedeconv was an R package for unified access to computational methods for estimating immune Cell fractions from bulk RNA-seq data[101]. The data were from the gene expression data of TCGA tumor samples after TPM was standardized. Then the immune cell fractions of tumor samples were calculated by quantiseq method invoking immunedeconv.

Biological function score calculation

GSVA is an R package for calculate biological function score based on a single sample[102]. The data was from TCGA in the calculation process, and the software package was GSVA. The immune score was T-cell infiltration rate of that calculated by quantiseq[100]. The gene sets that calculated the mitosis score were obtained from gene ontology (GO:0140014).

Acknowledgment

This work was supported by the National Key Research and Development Program of China (grants 2022YFA1103900 to H.J., F.L.; 2020YFA0803300 to H.J.); the National Natural Science Foundation of China (grants 82341002 to H.J., 32293192 to H.J., 82030083 to H.J., 82173340 to L.H., 82273400 to Y.J., 32100593 to X.T., 82203306 to W.H., 82372763 to X.W., 82303916 to C.G., 82303039 to Z.Q., 82141101 to F.L., 2022hwyq16 to F.L., 82372794 to F.L.); the Basic Frontier Scientific Research Program of Chinese Academy of Science (ZDBS-LY-SM006 to H.J.); the Innovative research team of high-level local universities in Shanghai (SSMU-ZLCX20180500 to H.J.); Science and Technology Commission of Shanghai Municipality (21ZR1470300 to L.H.). We thank Gongwei Wu for his helpful comments and suggestions on the manuscript.

Author Contributions

H.J. and F.L. conceived the idea and designed the analysis. X.C. performed all bioinformatics analyses. Y.Y., X.L., and L.C. provided technical assistance and helpful comments. H.J., F.L. and X.C. wrote the manuscript. All authors approved the final version.

Disclosure of conflicts of interest

The authors declare that there is no conflict of interests.

Online content

Code availability are available at https://github.com/XinleiCai/MEMORY.

References

1. Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) *19*, A68-77. 10.5114/wo.2014.47136.
2. Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet *45*, 1113-1120. 10.1038/ng.2764.
3. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400-416 e411. 10.1016/j.cell.2018.02.052.
4. Smeltzer, M.P., Faris, N.R., Ray, M.A., and Osarogiagbon, R.U. (2018). Association of Pathologic Nodal Staging Quality With Survival Among Patients With Non-Small Cell Lung Cancer After Resection With Curative Intent. JAMA Oncol *4*, 80-87. 10.1001/jamaoncol.2017.2993.
5. Miyauchi, E., Morita, S., Nakamura, A., Hosomi, Y., Watanabe, K., Ikeda, S., Seike, M., Fujita, Y., Minato, K., Ko, R., et al. (2022). Updated Analysis of NEJ009: Gefitinib-Alone Versus Gefitinib Plus Chemotherapy for Non-Small-Cell Lung Cancer With Mutated EGFR. J Clin Oncol, JCO2102911. 10.1200/JCO.21.02911.
6. Teh, S.H., Uong, S., Lin, T.Y., Shiraga, S., Li, Y., Gong, I.Y., Herrinton, L.J., and Li, R.A. (2021). Clinical Outcomes Following Regionalization of Gastric Cancer Care in a US Integrated Health Care System. J Clin Oncol *39*, 3364-3376. 10.1200/JCO.21.00480.
7. Gimotty, P.A., Van Belle, P., Elder, D.E., Murry, T., Montone, K.T., Xu, X., Hotz, S., Raines, S., Ming, M.E., Wahl, P., and Guerry, D. (2005). Biologic and prognostic significance of dermal Ki67 expression, mitoses, and tumorigenicity in thin invasive cutaneous melanoma. J Clin Oncol *23*, 8048-8056. 10.1200/JCO.2005.02.0735.
8. Wu, Y., Lv, M., Qian, T., and Shen, Y. (2021). Correlation analysis of Ki67 and CK7 expression with clinical characteristics and prognosis of postoperative cervical adenocarcinoma patients. Ann Palliat Med *10*, 9544-9552. 10.21037/apm-21-1974.
9. Mouabbi, J.A., Singareeka Raghavendra, A., Bassett, R.L., Jr., Hassan, A., Tripathy, D., and Layman, R.M. (2023). Survival Outcomes in Patients With Hormone Receptor-Positive Metastatic Breast Cancer With Low or No ERBB2 Expression Treated With Targeted Therapies Plus Endocrine Therapy. JAMA Netw Open *6*, e2313017. 10.1001/jamanetworkopen.2023.13017.
10. Kang, Y.J., Oh, S.J., Bae, S.Y., Kim, E.K., Lee, Y.J., Park, E.H., Jeong, J., Park, H.K., Suh, Y.J., and Kim, Y.S. (2023). Predictive biological factors for late survival in patients with HER2-positive breast cancer. Sci Rep *13*, 11008. 10.1038/s41598-023-38200-y.
11. Cai, C., Wang, X., Fu, Q., and Chen, A. (2022). The VEGF expression associated with prognosis in patients with intrahepatic cholangiocarcinoma: a systematic review and meta-analysis. World J Surg Oncol *20*, 40. 10.1186/s12957-022-02511-7.
12. Zhang, C., Wang, L., Xiong, C., Zhao, R., Liang, H., and Luo, X. (2021). The role of vascular endothelial growth factor as a prognostic and clinicopathological marker in osteosarcoma: a systematic review and meta-analysis. J Orthop Surg Res *16*, 738. 10.1186/s13018-021-02888-

3.
13. Lai, Y., OuYang, G., Sheng, L., Zhang, Y., Lai, B., and Zhou, M. (2021). Novel prognostic genes and subclasses of acute myeloid leukemia revealed by survival analysis of gene expression data. BMC Med Genomics *14*, 39. 10.1186/s12920-021-00888-0.
14. Anton, R.C., Coffey, D.M., Gondo, M.M., Stephenson, M.A., Brown, R.W., and Cagle, P.T. (2000). The expression of cyclins D1 and E in predicting short-term survival in squamous cell carcinoma of the lung. Mod Pathol *13*, 1167-1172. 10.1038/modpathol.3880215.
15. Esposito, V., Baldi, A., De Luca, A., Tonini, G., Vincenzi, B., Santini, D., Persichetti, P., Mancini, A., Citro, G., Baldi, F., et al. (2005). Cell cycle related proteins as prognostic parameters in radically resected non-small cell lung cancer. J Clin Pathol *58*, 734-739. 10.1136/jcp.2004.023531.
16. Dworakowska, D., Jassem, E., Jassem, J., Boltze, C., Wiedorn, K.H., Dworakowski, R., Skokowski, J., Jaskiewicz, K., and Czestochowska, E. (2005). Prognostic value of cyclin D1 overexpression in correlation with pRb and p53 status in non-small cell lung cancer (NSCLC). J Cancer Res Clin Oncol *131*, 479-485. 10.1007/s00432-004-0661-9.
17. Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the logrank test. Stat Med *1*, 121-129. 10.1002/sim.4780010204.
18. Gebski, V., Gares, V., Gibbs, E., and Byth, K. (2018). Data maturity and follow-up in time-to-event analyses. Int J Epidemiol *47*, 850-859. 10.1093/ije/dyy013.
19. Yao, X., Ireland, S.K., Pham, T., Temple, B., Chen, R., Raj, M.H., and Biliran, H. (2014). TLE1 promotes EMT in A549 lung cancer cells through suppression of E-cadherin. Biochem Biophys Res Commun *455*, 277-284. 10.1016/j.bbrc.2014.11.007.
20. Zheng, H., Tian, H., Yu, X., Ren, P., and Yang, Q. (2021). G protein gamma 7 suppresses progression of lung adenocarcinoma by inhibiting E2F transcription factor 1. Int J Biol Macromol *182*, 858-865. 10.1016/j.ijbiomac.2021.04.082.
21. Chen, G., Wang, Q., and Wang, K. (2022). MicroRNA-218-5p affects lung adenocarcinoma progression through targeting endoplasmic reticulum oxidoreductase 1 alpha. Bioengineered *13*, 10061-10070. 10.1080/21655979.2022.2063537.
22. Xu, J., Zheng, H., Yuan, S., Zhou, B., Zhao, W., Pan, Y., and Qi, D. (2019). Overexpression of ANLN in lung adenocarcinoma is associated with metastasis. Thorac Cancer *10*, 1702-1709. 10.1111/1759-7714.13135.
23. Zhang, P., Li, S., Lv, C., Si, J., Xiong, Y., Ding, L., Ma, Y., and Yang, Y. (2018). BPI-9016M, a c-Met inhibitor, suppresses tumor cell growth, migration and invasion of lung adenocarcinoma via miR203-DKK1. Theranostics *8*, 5890-5902. 10.7150/thno.27667.
24. Wang, T., Zhai, R., Lv, X., Wang, K., and Xu, J. (2020). LINC02418 promotes malignant behaviors in lung adenocarcinoma cells by sponging miR-4677-3p to upregulate KNL1 expression. BMC Pulm Med *20*, 217. 10.1186/s12890-020-01229-0.
25. Huo, S.F., Shang, W.L., Yu, M., Ren, X.P., Wen, H.X., Chai, C.Y., Sun, L., Hui, K., Liu, L.H., Wei, S.H., et al. (2020). STEAP1 facilitates metastasis and epithelial-mesenchymal transition of lung adenocarcinoma via the JAK2/STAT3 signaling pathway. Biosci Rep *40*. 10.1042/BSR20193169.
26. Zhao, Z., Li, J., Tan, F., Gao, S., and He, J. (2018). mTOR up-regulation of BEX4 promotes lung adenocarcinoma cell proliferation by potentiating OCT4. Biochem Biophys Res Commun *500*, 302-309. 10.1016/j.bbrc.2018.04.064.

27. Fan, T., Liu, Y., Liu, H., Wang, L., Tian, H., Zheng, Y., Zheng, B., Xue, L., Li, C., and He, J. (2022). Transmembrane Protein-Based Risk Model and H3K4me3 Modification Characteristics in Lung Adenocarcinoma. Frontiers in Oncology *12*. 10.3389/fonc.2022.828814.
28. Lee, S., Kang, H., Shin, E., Jeon, J., Youn, H., and Youn, B. (2021). BEX1 and BEX4 Induce GBM Progression through Regulation of Actin Polymerization and Activation of YAP/TAZ Signaling. Int J Mol Sci *22*. 10.3390/ijms22189845.
29. Gao, W., Li, J.Z., Chen, S.Q., Chu, C.Y., Chan, J.Y., and Wong, T.S. (2016). Decreased brain-expressed X-linked 4 (BEX4) expression promotes growth of oral squamous cell carcinoma. J Exp Clin Cancer Res *35*, 92. 10.1186/s13046-016-0355-6.
30. He, Y., Luo, Y., Zhang, D., Wang, X., Zhang, P., Li, H., Ejaz, S., and Liang, S. (2019). PGK1-mediated cancer progression and drug resistance. Am J Cancer Res *9*, 2280-2302.
31. Matsuura, K., Sawai, H., Ikeo, K., Ogawa, S., Iio, E., Isogawa, M., Shimada, N., Komori, A., Toyoda, H., Kumada, T., et al. (2017). Genome-Wide Association Study Identifies TLL1 Variant Associated With Development of Hepatocellular Carcinoma After Eradication of Hepatitis C Virus Infection. Gastroenterology *152*, 1383-1394. 10.1053/j.gastro.2017.01.041.
32. Poplawski, P., Alseekh, S., Jankowska, U., Skupien-Rabian, B., Iwanicka-Nowicka, R., Kossowska, H., Fogtman, A., Rybicka, B., Boguslawska, J., Adamiok-Ostrowska, A., et al. (2023). Coordinated reprogramming of renal cancer transcriptome, metabolome and secretome associates with immune tumor infiltration. Cancer Cell Int *23*, 2. 10.1186/s12935-022-02845-y.
33. Guo, M., Li, F., Zhao, L., Fang, Z., Yu, H., Songyang, Z., and Xiong, Y. (2023). Pan-cancer investigation of C-to-U editing reveals its important role in cancer development and new targets for cancer treatment. Front Oncol *13*, 1097667. 10.3389/fonc.2023.1097667.
34. Povero, D., Johnson, S.M., and Liu, J. (2020). Hypoxia, hypoxia-inducible gene 2 (HIG2)/HILPDA, and intracellular lipolysis in cancer. Cancer Lett *493*, 71-79. 10.1016/j.canlet.2020.06.013.
35. Erfani, S., Hua, H., Pan, Y., Zhou, B.P., and Yang, X.H. (2021). The Context-Dependent Impact of Integrin-Associated CD151 and Other Tetraspanins on Cancer Development and Progression: A Class of Versatile Mediators of Cellular Function and Signaling, Tumorigenesis and Metastasis. Cancers (Basel) *13*. 10.3390/cancers13092005.
36. Won, M., Park, K.A., Kim, S., Ju, E., Ko, Y., Yoo, H., Ro, H., Lee, J., Oh, J., Lee, E.G., et al. (2022). ANKRD13a controls early cell-death checkpoint by interacting with RIP1 independent of NF-kappaB. Cell Death Differ *29*, 1152-1163. 10.1038/s41418-021-00906-9.
37. Xu, J., Chen, Q., Tian, K., Liang, R., Chen, T., Gong, A., Mathy, N.W., Yu, T., and Chen, X. (2020). m6A methyltransferase METTL3 maintains colon cancer tumorigenicity by suppressing SOCS2 to promote cell proliferation. Oncol Rep *44*, 973-986. 10.3892/or.2020.7665.
38. Xu, C., Li, H., Zhang, L., Jia, T., Duan, L., and Lu, C. (2016). MicroRNA□1915□3p prevents the apoptosis of lung cancer cells by downregulating DRG2 and PBX2. Mol Med Rep *13*, 505-512. 10.3892/mmr.2015.4565.
39. Mead, T.J. (2022). ADAMTS6: Emerging roles in cardiovascular, musculoskeletal and cancer biology. Front Mol Biosci *9*, 1023511. 10.3389/fmolb.2022.1023511.
40. Yang, B.Y., Song, J.W., Sun, H.Z., Xing, J.C., Yang, Z.H., Wei, C.Y., Xu, T.Y., Yu, Z.N.,

Zhang, Y.N., Wang, Y.F., et al. (2018). PSMB8 regulates glioma cell migration, proliferation, and apoptosis through modulating ERK1/2 and PI3K/AKT signaling pathways. Biomed Pharmacother *100*, 205-212. 10.1016/j.biopha.2018.01.170.

41. Keshet, R., Lee, J.S., Adler, L., Iraqi, M., Ariav, Y., Lim, L.Q.J., Lerner, S., Rabinovich, S., Oren, R., Katzir, R., et al. (2020). Targeting purine synthesis in ASS1-expressing tumors enhances the response to immune checkpoint inhibitors. Nat Cancer *1*, 894-908. 10.1038/s43018-020-0106-7.

42. Palaparti, A., Baratz, A., and Stifani, S. (1997). The Groucho/Transducin-like Enhancer of split Transcriptional Repressors Interact with the Genetically Defined Amino-terminal Silencing Domain of Histone H3. Journal of Biological Chemistry *272*, 26604-26610. 10.1074/jbc.272.42.26604.

43. Lin, W., Wang, J., Ge, J., Zhou, R., Hu, Y., Xiao, L., Peng, Q., and Zheng, Z. (2023). The activity of cuproptosis pathway calculated by AUCell algorithm was employed to construct cuproptosis landscape in lung adenocarcinoma. Discover Oncology *14*. 10.1007/s12672-023-00755-7.

44. Li, X., Jiang, Y., Meisenhelder, J., Yang, W., Hawke, D.H., Zheng, Y., Xia, Y., Aldape, K., He, J., Hunter, T., et al. (2016). Mitochondria-Translocated PGK1 Functions as a Protein Kinase to Coordinate Glycolysis and the TCA Cycle in Tumorigenesis. Mol Cell *61*, 705-719. 10.1016/j.molcel.2016.02.009.

45. Fu, Q., and Yu, Z. (2020). Phosphoglycerate kinase 1 (PGK1) in cancer: A promising target for diagnosis and therapy. Life Sci *256*, 117863. 10.1016/j.lfs.2020.117863.

46. Waldman, A.D., Fritz, J.M., and Lenardo, M.J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol *20*, 651-668. 10.1038/s41577-020-0306-5.

47. Zhou, Z., Li, Y., Hao, H., Wang, Y., Zhou, Z., Wang, Z., and Chu, X. (2019). Screening Hub Genes as Prognostic Biomarkers of Hepatocellular Carcinoma by Bioinformatics Analysis. Cell Transplant *28*, 76S-86S. 10.1177/0963689719893950.

48. Hwang, L.H., Lau, L.F., Smith, D.L., Mistrot, C.A., Hardwick, K.G., Hwang, E.S., Amon, A., and Murray, A.W. (1998). Budding yeast Cdc20: a target of the spindle checkpoint. Science *279*, 1041-1044. 10.1126/science.279.5353.1041.

49. Uuskula-Reimand, L., and Wilson, M.D. (2022). Untangling the roles of TOP2A and TOP2B in transcription and cancer. Sci Adv *8*, eadd4920. 10.1126/sciadv.add4920.

50. Lamers, F., van der Ploeg, I., Schild, L., Ebus, M.E., Koster, J., Hansen, B.R., Koch, T., Versteeg, R., Caron, H.N., and Molenaar, J.J. (2011). Knockdown of survivin (BIRC5) causes apoptosis in neuroblastoma via mitotic catastrophe. Endocr Relat Cancer *18*, 657-668. 10.1530/ERC-11-0207.

51. Wittmann, T., Wilm, M., Karsenti, E., and Vernos, I. (2000). TPX2, A novel xenopus MAP involved in spindle pole organization. J Cell Biol *149*, 1405-1418. 10.1083/jcb.149.7.1405.

52. Kufer, T.A., Sillje, H.H., Korner, R., Gruss, O.J., Meraldi, P., and Nigg, E.A. (2002). Human TPX2 is required for targeting Aurora-A kinase to the spindle. J Cell Biol *158*, 617-623. 10.1083/jcb.200204155.

53. Cancer Genome Atlas Research, N. (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543-550. 10.1038/nature13385.

54. Zheng, F., Kelly, M.R., Ramms, D.J., Heintschel, M.L., Tao, K., Tutuncuoglu, B., Lee, J.J.,

Ono, K., Foussard, H., Chen, M., et al. (2021). Interpretation of cancer mutations using a multiscale map of protein systems. Science *374*, eabf3067. 10.1126/science.abf3067.

55. Wang, Q., Shi, Y.L., Zhou, K., Wang, L.L., Yan, Z.X., Liu, Y.L., Xu, L.L., Zhao, S.W., Chu, H.L., Shi, T.T., et al. (2018). PIK3CA mutations confer resistance to first-line chemotherapy in colorectal cancer. Cell Death Dis *9*, 739. 10.1038/s41419-018-0776-6.

56. Shibata, T., Kokubu, A., Tsuta, K., and Hirohashi, S. (2009). Oncogenic mutation of PIK3CA in small cell lung carcinoma: a potential therapeutic target pathway for chemotherapy-resistant lung cancer. Cancer Lett *283*, 203-211. 10.1016/j.canlet.2009.03.038.

57. Hanker, A.B., Pfefferle, A.D., Balko, J.M., Kuba, M.G., Young, C.D., Sanchez, V., Sutton, C.R., Cheng, H., Perou, C.M., Zhao, J.J., et al. (2013). Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies. Proc Natl Acad Sci U S A *110*, 14372-14377. 10.1073/pnas.1303204110.

58. Eng, J., Woo, K.M., Sima, C.S., Plodkowski, A., Hellmann, M.D., Chaft, J.E., Kris, M.G., Arcila, M.E., Ladanyi, M., and Drilon, A. (2015). Impact of Concurrent PIK3CA Mutations on Response to EGFR Tyrosine Kinase Inhibition in EGFR-Mutant Lung Cancers and on Prognosis in Oncogene-Driven Lung Adenocarcinomas. J Thorac Oncol *10*, 1713-1719. 10.1097/JTO.0000000000000671.

59. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The Immune Landscape of Cancer. Immunity *48*, 812-830 e814. 10.1016/j.immuni.2018.03.023.

60. Meigs, T.E., Fedor-Chaiken, M., Kaplan, D.D., Brackenbury, R., and Casey, P.J. (2002). Galpha12 and Galpha13 negatively regulate the adhesive functions of cadherin. J Biol Chem *277*, 24594-24600. 10.1074/jbc.M201984200.

61. An, Y., Adams, J.R., Hollern, D.P., Zhao, A., Chang, S.G., Gams, M.S., Chung, P.E.D., He, X., Jangra, R., Shah, J.S., et al. (2018). Cdh1 and Pik3ca Mutations Cooperate to Induce Immune-Related Invasive Lobular Carcinoma of the Breast. Cell Rep *25*, 702-714 e706. 10.1016/j.celrep.2018.09.056.

62. Zeisberg, M., and Neilson, E.G. (2009). Biomarkers for epithelial-mesenchymal transitions. J Clin Invest *119*, 1429-1437. 10.1172/JCI36183.

63. Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927-939. 10.1016/j.cell.2004.06.006.

64. Egner, J.R. (2010). AJCC cancer staging manual. Jama *304*, 1726-1727.

65. Rebello, R.J., Oing, C., Knudsen, K.E., Loeb, S., Johnson, D.C., Reiter, R.E., Gillessen, S., Van der Kwast, T., and Bristow, R.G. (2021). Prostate cancer. Nat Rev Dis Primers *7*, 9. 10.1038/s41572-020-00243-0.

66. Lee, J.K., Lee, J., Go, H., Lee, C.G., Kim, S., Kim, H.S., Cho, H., Choi, K.S., Ha, G.H., and Lee, C.W. (2016). Oncogenic microtubule hyperacetylation through BEX4-mediated sirtuin 2 inhibition. Cell Death Dis *7*, e2336. 10.1038/cddis.2016.240.

67. Lee, J.K., Ha, G.H., Kim, H.S., and Lee, C.W. (2018). Oncogenic potential of BEX4 is conferred by Polo-like kinase 1-mediated phosphorylation. Exp Mol Med *50*, 1-12. 10.1038/s12276-018-0168-0.

68. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina,

J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899-905. 10.1038/nature08822.

69. Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948-962. 10.1016/j.cell.2013.10.011.

70. Girish, V., Lakhani, A.A., Thompson, S.L., Scaduto, C.M., Brown, L.M., Hagenson, R.A., Sausville, E.L., Mendelson, B.E., Kandikuppa, P.K., Lukow, D.A., et al. (2023). Oncogene-like addiction to aneuploidy in human cancers. Science *381*, eadg4521. 10.1126/science.adg4521.

71. Bosco, N., Goldberg, A., Zhao, X., Mays, J.C., Cheng, P., Johnson, A.F., Bianchi, J.J., Toscani, C., Di Tommaso, E., Katsnelson, L., et al. (2023). KaryoCreate: A CRISPR-based technology to study chromosome-specific aneuploidy by targeting human centromeres. Cell *186*, 1985-2001 e1919. 10.1016/j.cell.2023.03.029.

72. Ha, S.Y., Choi, M., Lee, T., and Park, C.K. (2016). The Prognostic Role of Mitotic Index in Hepatocellular Carcinoma Patients after Curative Hepatectomy. Cancer Res Treat *48*, 180-189. 10.4143/crt.2014.321.

73. Zhou, C., Wang, Y., Liu, X., Liang, Y., Fan, Z., Jiang, T., Wang, Y., and Wang, L. (2018). Molecular profiles for insular low-grade gliomas with putamen involvement. J Neurooncol *138*, 659-666. 10.1007/s11060-018-2837-1.

74. Makinen, J.M., Laitakari, K., Johnson, S., Makitaro, R., Bloigu, R., Paakko, P., Lappi-Blanco, E., and Kaarteenaho, R. (2017). Histological features of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma. Histopathology *71*, 425-436. 10.1111/his.13236.

75. Byrne, A., Savas, P., Sant, S., Li, R., Virassamy, B., Luen, S.J., Beavis, P.A., Mackay, L.K., Neeson, P.J., and Loi, S. (2020). Tissue-resident memory T cells in breast cancer control and immunotherapy responses. Nat Rev Clin Oncol *17*, 341-348. 10.1038/s41571-020-0333-y.

76. Liu, W., Sun, L., Zhang, J., Song, W., Li, M., and Wang, H. (2021). The landscape and prognostic value of immune characteristics in uterine corpus endometrial cancer. Biosci Rep *41*. 10.1042/BSR20202321.

77. Ducreux, M., Cuhna, A.S., Caramella, C., Hollebecque, A., Burtin, P., Goere, D., Seufferlein, T., Haustermans, K., Van Laethem, J.L., Conroy, T., et al. (2015). Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *26 Suppl 5*, v56-68. 10.1093/annonc/mdv295.

78. Cardoso, F., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rubio, I.T., Zackrisson, S., Senkus, E., and clinicalguidelines@esmo.org, E.G.C.E.a. (2019). Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger. Ann Oncol *30*, 1194-1220. 10.1093/annonc/mdz173.

79. Oaknin, A., Bosse, T.J., Creutzberg, C.L., Giornelli, G., Harter, P., Joly, F., Lorusso, D., Marth, C., Makker, V., Mirza, M.R., et al. (2022). Endometrial cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol *33*, 860-877. 10.1016/j.annonc.2022.05.009.

80. Vogel, A., Cervantes, A., Chau, I., Daniele, B., Llovet, J.M., Meyer, T., Nault, J.C., Neumann, U., Ricke, J., Sangro, B., et al. (2018). Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *29*, iv238-iv255.

10.1093/annonc/mdy308.

81. Stupp, R., Brada, M., van den Bent, M.J., Tonn, J.C., Pentheroudakis, G., and Group, E.G.W. (2014). High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *25 Suppl 3*, iii93-101. 10.1093/annonc/mdu050.

82. Postmus, P.E., Kerr, K.M., Oudkerk, M., Senan, S., Waller, D.A., Vansteenkiste, J., Escriu, C., Peters, S., and Committee, E.G. (2017). Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol *28*, iv1-iv21. 10.1093/annonc/mdx222.

83. Gui, C., Kosteniuk, S.E., Lau, J.C., and Megyesi, J.F. (2018). Tumor growth dynamics in serially-imaged low-grade glioma patients. J Neurooncol *139*, 167-175. 10.1007/s11060-018-2857-x.

84. Alvarez, M., Benhammou, J.N., Darci-Maher, N., French, S.W., Han, S.B., Sinsheimer, J.S., Agopian, V.G., Pisegna, J.R., and Pajukanta, P. (2022). Human liver single nucleus and single cell RNA sequencing identify a hepatocellular carcinoma-associated cell-type affecting survival. Genome Med *14*, 50. 10.1186/s13073-022-01055-5.

85. Infante, M., Berghmans, T., Heuvelmans, M.A., Hillerdal, G., and Oudkerk, M. (2013). Slow-growing lung cancer as an emerging entity: from screening to clinical management. Eur Respir J *42*, 1706-1722. 10.1183/09031936.00186212.

86. Hoshino, T. (1984). A commentary on the biology and growth kinetics of low-grade and high-grade gliomas. J Neurosurg *61*, 895-900. 10.3171/jns.1984.61.5.0895.

87. Sajjadi, E., Venetis, K., Piciotti, R., Invernizzi, M., Guerini-Rocco, E., Haricharan, S., and Fusco, N. (2021). Mismatch repair-deficient hormone receptor-positive breast cancers: Biology and pathological characterization. Cancer Cell Int *21*, 266. 10.1186/s12935-021-01976-y.

88. Doghri, R., Houcine, Y., Boujelbene, N., Driss, M., Charfi, L., Abbes, I., Mrad, K., and Sellami, R. (2019). Mismatch Repair Deficiency in Endometrial Cancer: Immunohistochemistry Staining and Clinical Implications. Appl Immunohistochem Mol Morphol *27*, 678-682. 10.1097/PAI.0000000000000641.

89. Zhang, Q., Sun, T.Y., Kang, P.M., Qian, K., Deng, B., Zhou, J.H., Wang, R.W., Jiang, B., Li, K., Liu, F., et al. (2016). Combined analysis of rearrangement of ALK, ROS1, somatic mutation of EGFR, KRAS, BRAF, PIK3CA, and mRNA expression of ERCC1, TYMS, RRM1, TUBB3, EGFR in patients with non-small cell lung cancer and their clinical significance. Cancer Chemoth Pharm *77*, 583-593. 10.1007/s00280-016-2969-y.

90. Donev, I.S., Wang, W., Yamada, T., Li, Q., Takeuchi, S., Matsumoto, K., Yamori, T., Nishioka, Y., Sone, S., and Yano, S. (2011). Transient PI3K inhibition induces apoptosis and overcomes HGF-mediated resistance to EGFR-TKIs in EGFR mutant lung cancer. Clin Cancer Res *17*, 2260-2269. 10.1158/1078-0432.CCR-10-1993.

91. Wang, G., Xu, D., Zhang, Z., Li, X., Shi, J., Sun, J., Liu, H.Z., Li, X., Zhou, M., and Zheng, T. (2021). The pan-cancer landscape of crosstalk between epithelial-mesenchymal transition and immune evasion relevant to prognosis and immunotherapy response. NPJ Precis Oncol *5*, 56. 10.1038/s41698-021-00200-4.

92. Brown, T.J., Bota, D.A., van Den Bent, M.J., Brown, P.D., Maher, E., Aregawi, D., Liau, L.M., Buckner, J.C., Weller, M., Berger, M.S., and Glantz, M. (2019). Management of low-grade glioma: a systematic review and meta-analysis. Neurooncol Pract *6*, 249-258.

10.1093/nop/npy034.

93. Chen, L., Ma, J., Zou, Z., Liu, H., Liu, C., Gong, S., Gao, X., and Liang, G. (2022). Clinical characteristics and prognosis of patients with glioblastoma: A review of survival analysis of 1674 patients based on SEER database. Medicine (Baltimore) *101*, e32042. 10.1097/MD.0000000000032042.

94. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504. 10.1101/gr.1239303.

95. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.

96. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res *28*, 1747-1756. 10.1101/gr.239244.118.

97. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576 e516. 10.1016/j.cell.2017.06.010.

98. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res *41*, D955-961. 10.1093/nar/gks1111.

99. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437-1452 e1417. 10.1016/j.cell.2017.10.049.

100. Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., Krogsdam, A., Loncova, Z., Posch, W., Wilflingseder, D., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med *11*, 34. 10.1186/s13073-019-0638-6.

101. Sturm, G., Finotello, F., and List, M. (2020). Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. Methods Mol Biol *2120*, 223-232. 10.1007/978-1-0716-0327-7_16.

102. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7. 10.1186/1471-2105-14-7.

Figure Legends

Figure 1: MEMORY uncovers the enrichment of mitosis and immune signatures in multiple cancers. (**A**) Algorithm of MEMORY. Sample sizes, ranging from 10% to 100% with 10% intervals ("gradient"), were used for 1000 permutations of survival analyses of all 15 cancer types. Each matrix was divided into high and low expression groups based on median gene expression values, and survival analyses were performed. Log-rank test significance results were coded as 1 for significant ($P < 0.05$) and 0 for non-significant ($P > 0.05$) outcomes, forming survival analysis matrices and the summarized significant probability matrix, which allowed the identification of GEARs. (**B**) The maximum of significant probability for each gradient sample size of every cancer type. Sample number rate refers to the percentage of samples in each sampling gradient compared to the total number of samples. (**C**) The pathways were enriched by GOEA based on the GEARs of each cancer type. The displayed pathways represent the top 5 most significant pathways for each cancer type. Mitosis-related pathways were marked in red whereas immune-related pathways were marked in blue. Only pathways with a Benjamini–Hochberg–adjusted $p < 0.05$ are displayed.

Figure 2: Identification of hub genes in mitosis or immune-related cancers. (**A**) GEARs and core survival networks. The SAS of all GEARs was calculated, and CSN was constructed. (**B**) Cancers and their hub gene pathways. The left nodes represent five different cancers, while the right nodes detail the functional types of hub-gene pathways. The height of the edges in the middle represents the proportion of hub-gene pathways corresponding to a specified functional type. Only pathways with a Benjamini–Hochberg–adjusted $p < 0.05$ are displayed. (**C-D**) The CSN of LUAD and BRCA. The enlarged section represents the hub gene. (**E**) Heat map showing the hub gene expression of A549 after compounds treatment from cMAP database[99]. Blue annotations meant 76 compounds that can inhibit hub gene expression. (**F-G**) LUAD clustering based on hub gene expression in comparison with the classic classification.

Figure 3: Different LUAD subgroups were characterized by unique genetic mutation profiles. (**A**) TMB analysis in three groups of LUAD samples. ***$P < 0.0005$. (**B**) Proportion of

different oncogenic drivers, including *KRAS*, *EGFR*, *BRAF* and *ERBB2* mutations, and *ALK* and *ROS1* fusions in three LUAD subgroups. (**C**) Kaplan-Meyer overall survival curves for EGFR-mutation samples of LUAD by hub gene subgroups. (**D**) Kaplan-Meyer overall survival curves for pan-negative LUAD samples by hub gene subgroups. (**E-G**) Comparison of top gene mutations in three LUAD clusters, including ML vs. MM (**C**), MM vs. MH (**D**) and ML vs MH (**E**).

Figure 4: PIK3CA mutation associates with mitosis and drug resistance in LUAD. (**A**) Cancer-related gene mutations were annotated based on gene dependency scores data from Depmap database. Functional mutations were indicated in green; functional (subtype-associated) mutations were highlighted in orange; non-functional mutations were indicated in gray. (**B**) The analysis of the NeST differential pathways across three groups including ML vs. MM, MM vs. MH, ML vs MH. Only pathways with a Benjamini–Hochberg–adjusted $p < 0.05$ are displayed. (**C**) Comparison of the IC50 z-score of five tumor cell growth inhibitors for A549 (*KRAS* mutation) and SW1573 (*KRAS* and *PIK3CA* mutation).

Figure 5: The hub gene classification of BRCA revealed a significant association between EMT and immune infiltration. (**A**) Hub gene expression heatmap. The heatmap contains the result of hub gene classification and PAM50 classification of BRCA[59]. (**B**) Comparison of molecular classification based on hub genes with PAM50 classification. (**C-E**) Comparison of gene mutations in three groups of BRCA samples, including IL vs. IM (**C**), IM vs. IH (**D**), IL vs. IH (**E**). (**F**) Comparison of the total immune cell rate in three BRCA clusters. ***$P < 0.0005$. (**G**) Comparison of *CDH1* expression in BRCA samples with wildtype or mutant *CDH1*. ***$P < 0.0005$. (**H**) A schematic diagram illustrating the correlation between the EMT score and immune cell rate. The red line represents the fitting curve. The correlation analysis was performed using Pearson correlation coefficient. Statistical significance for all pairwise group comparisons was assessed with the Wilcoxon test.

Figure 6: Mitosis and immune signatures predict patient prognosis at the pan-cancer level. (**A**) The correlation of mitosis scores of 33 TCGA cancer types with *PIK3CA* expression. (**B**) The correlation of immune cell infiltration rate of 33 TCGA cancer types with *CDH1* expression.

(**C**) The mitosis and immune-related pathway scores of 33 cancer types. Median score was utilized as a threshold to categorize patients for survival analysis. Among the 33 cancer types examined, 10 cancer types were exclusively associated with the mitosis score, 2 cancer types were exclusively associated with the immune score, and 5 cancer types showed a correlation with both mitosis and immunity scores concurrently. Dots represent $-log_{10}(p)$ values from log-rank tests comparing high vs. low score groups for each cancer type. The dashed line marks the nominal significance threshold (p = 0.05).

Figure 1

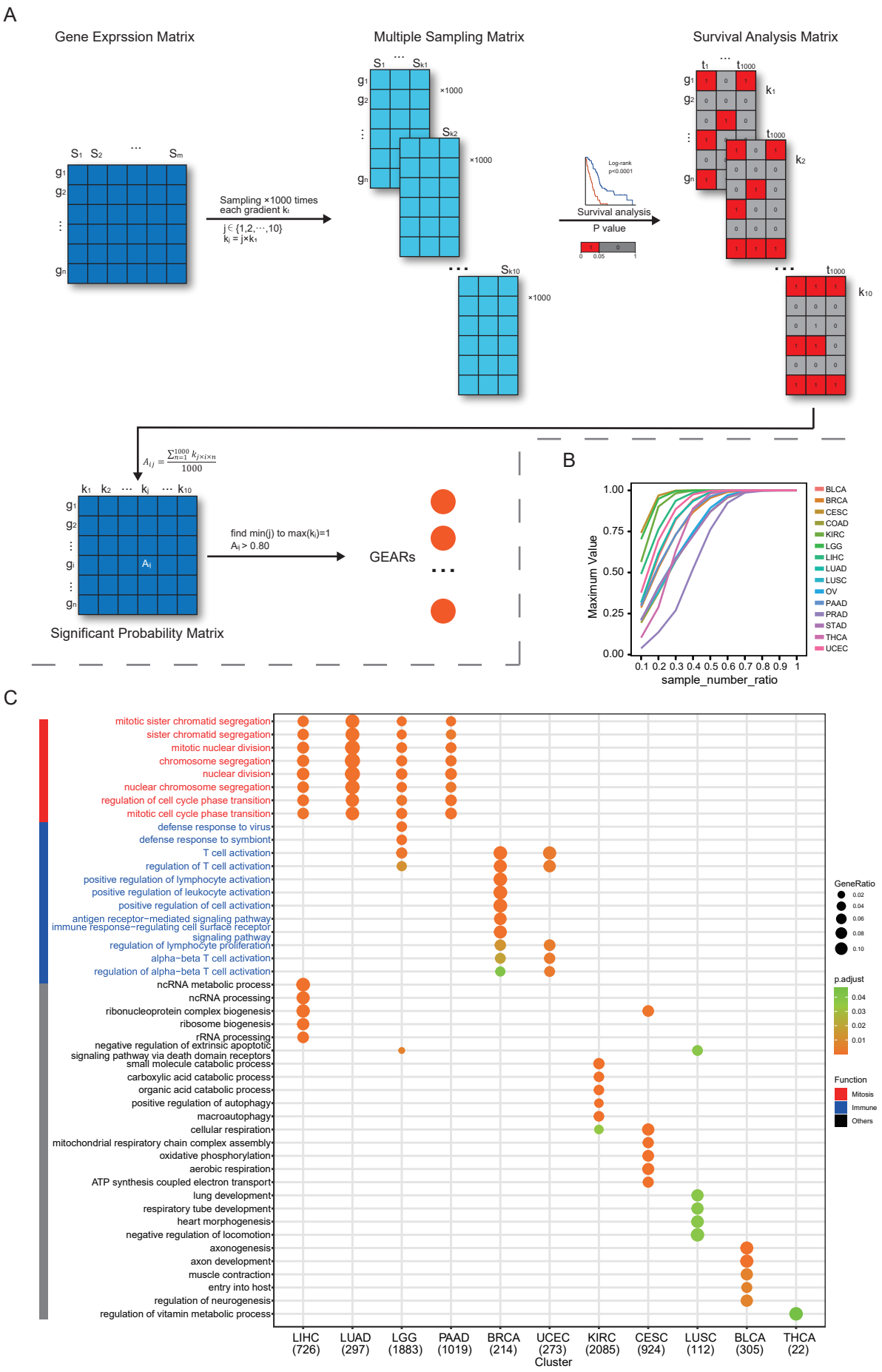

A
Gene Exprssion Matrix
Multiple Sampling Matrix
Survival Analysis Matrix
Sampling ×1000 times each gradient kᵢ
j ∈ {1,2,⋯,10}
kⱼ = j×k₁
×1000
Log-rank p<0.0001
Survival analysis
P value
Significant Probability Matrix
find min(j) to max(kⱼ)=1
Aᵢⱼ > 0.80
GEARs
B
Maximum Value
sample_number_ratio
BLCA
BRCA
CESC
COAD
KIRC
LGG
LIHC
LUAD
LUSC
OV
PAAD
PRAD
STAD
THCA
UCEC
C
mitotic sister chromatid segregation
sister chromatid segregation
mitotic nuclear division
chromosome segregation
nuclear division
nuclear chromosome segregation
regulation of cell cycle phase transition
mitotic cell cycle phase transition
defense response to virus
defense response to symbiont
T cell activation
regulation of T cell activation
positive regulation of lymphocyte activation
positive regulation of leukocyte activation
positive regulation of cell activation
antigen receptor−mediated signaling pathway
immune response−regulating cell surface receptor signaling pathway
regulation of lymphocyte proliferation
alpha−beta T cell activation
regulation of alpha−beta T cell activation
ncRNA metabolic process
ncRNA processing
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
negative regulation of extrinsic apoptotic signaling pathway via death domain receptors
small molecule catabolic process
carboxylic acid catabolic process
organic acid catabolic process
positive regulation of autophagy
macroautophagy
cellular respiration
mitochondrial respiratory chain complex assembly
oxidative phosphorylation
aerobic respiration
ATP synthesis coupled electron transport
lung development
respiratory tube development
heart morphogenesis
negative regulation of locomotion
axonogenesis
axon development
muscle contraction
entry into host
regulation of neurogenesis
regulation of vitamin metabolic process
LIHC (726)
LUAD (297)
LGG (1883)
PAAD (1019)
BRCA (214)
UCEC (273)
KIRC (2085)
CESC (924)
LUSC (112)
BLCA (305)
THCA (22)
Cluster
GeneRatio
p.adjust
Function
Mitosis
Immune
Others

Figure 2

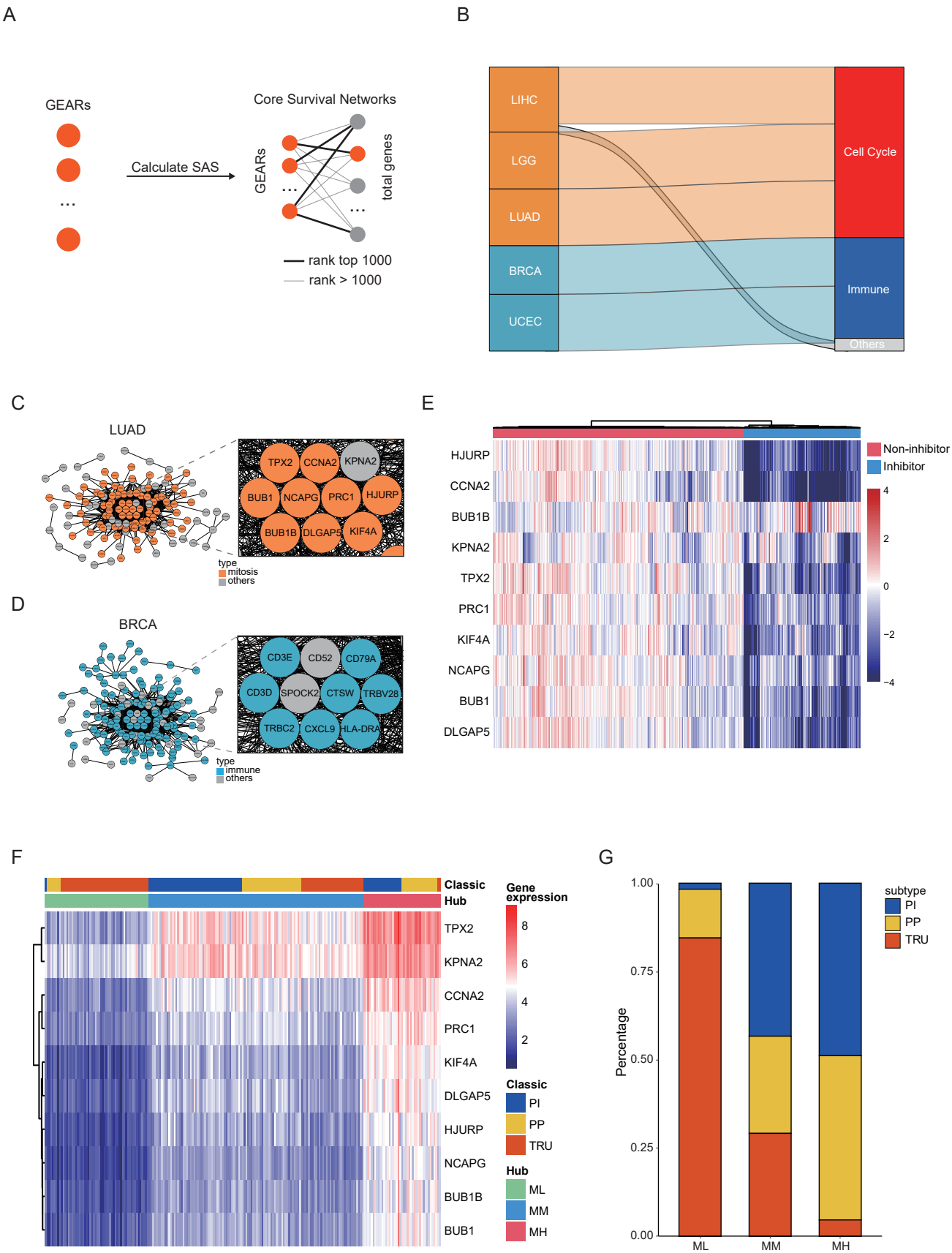

A
GEARs
Calculate SAS
Core Survival Networks
GEARs
total genes
rank top 1000
rank > 1000
B
LIHC
LGG
LUAD
BRCA
UCEC
Cell Cycle
Immune
Others
C
LUAD
TPX2
CCNA2
KPNA2
BUB1
NCAPG
PRC1
HJURP
BUB1B
DLGAP5
KIF4A
type
mitosis
others
D
BRCA
CD3E
CD52
CD79A
CD3D
SPOCK2
CTSW
TRBV28
TRBC2
CXCL9
HLA-DRA
type
immune
others
E
HJURP
CCNA2
BUB1B
KPNA2
TPX2
PRC1
KIF4A
NCAPG
BUB1
DLGAP5
Non-inhibitor
Inhibitor
4
2
0
−2
−4
F
Classic
Hub
TPX2
KPNA2
CCNA2
PRC1
KIF4A
DLGAP5
HJURP
NCAPG
BUB1B
BUB1
Gene expression
8
6
4
2
Classic
PI
PP
TRU
Hub
ML
MM
MH
G
1.00
0.75
0.50
0.25
0.00
Percentage
ML
MM
MH
subtype
PI
PP
TRU

Figure 3

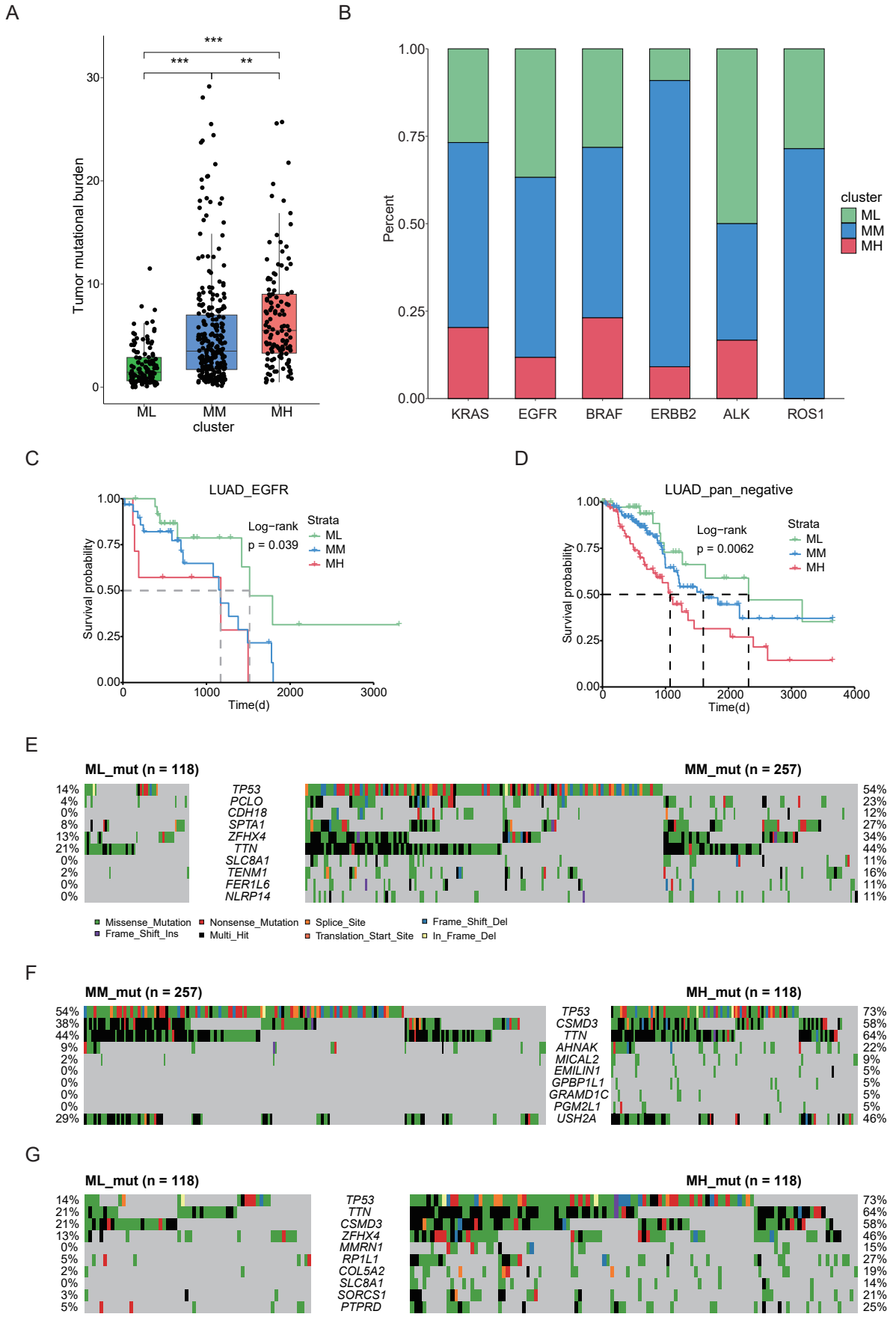

A
Tumor mutational burden
cluster
ML
MM
MH
B
Percent
cluster
ML
MM
MH
KRAS
EGFR
BRAF
ERBB2
ALK
ROS1
C
LUAD_EGFR
Survival probability
Log−rank
p = 0.039
Strata
ML
MM
MH
Time(d)
D
LUAD_pan_negative
Survival probability
Log−rank
p = 0.0062
Strata
ML
MM
MH
Time(d)
E
ML_mut (n = 118)
MM_mut (n = 257)
TP53
PCLO
CDH18
SPTA1
ZFHX4
TTN
SLC8A1
TENM1
FER1L6
NLRP14
Missense_Mutation
Nonsense_Mutation
Splice_Site
Frame_Shift_Del
Frame_Shift_Ins
Multi_Hit
Translation_Start_Site
In_Frame_Del
F
MM_mut (n = 257)
MH_mut (n = 118)
TP53
CSMD3
TTN
AHNAK
MICAL2
EMILIN1
GPBP1L1
GRAMD1C
PGM2L1
USH2A
G
ML_mut (n = 118)
MH_mut (n = 118)
TP53
TTN
CSMD3
ZFHX4
MMRN1
RP1L1
COL5A2
SLC8A1
SORCS1
PTPRD

Figure 4

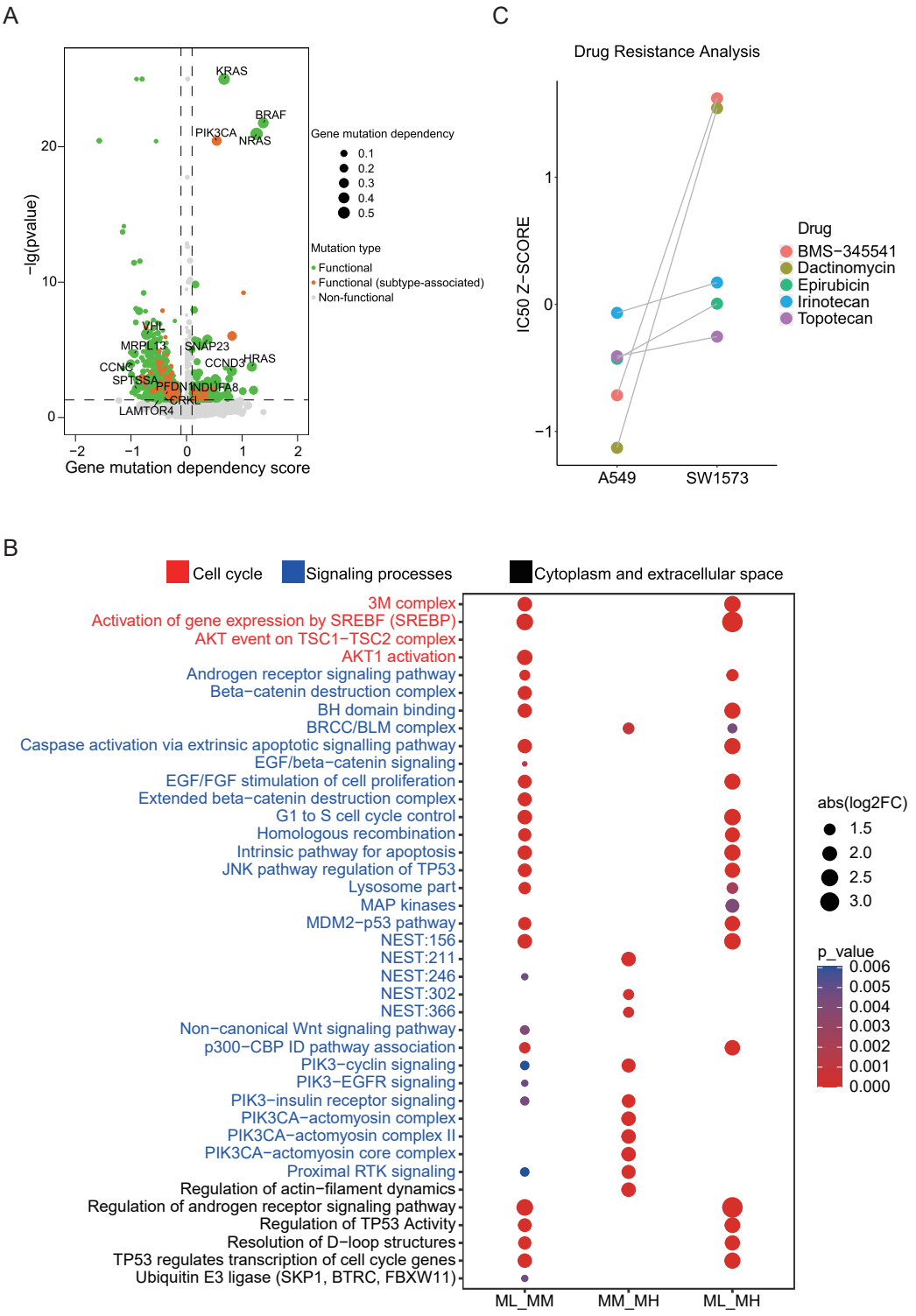

A
KRAS
BRAF
PIK3CA
NRAS
VHL
MRPL13
SNAP23
CCND3
HRAS
CCNC
SPTSSA
PFDN1
NDUFA8
CRKL
LAMTOR4
−lg(pvalue)
Gene mutation dependency score
Gene mutation dependency
0.1
0.2
0.3
0.4
0.5
Mutation type
Functional
Functional (subtype-associated)
Non-functional
C
Drug Resistance Analysis
IC50 Z−SCORE
A549
SW1573
Drug
BMS−345541
Dactinomycin
Epirubicin
Irinotecan
Topotecan
B
Cell cycle
Signaling processes
Cytoplasm and extracellular space
3M complex
Activation of gene expression by SREBF (SREBP)
AKT event on TSC1−TSC2 complex
AKT1 activation
Androgen receptor signaling pathway
Beta−catenin destruction complex
BH domain binding
BRCC/BLM complex
Caspase activation via extrinsic apoptotic signalling pathway
EGF/beta−catenin signaling
EGF/FGF stimulation of cell proliferation
Extended beta−catenin destruction complex
G1 to S cell cycle control
Homologous recombination
Intrinsic pathway for apoptosis
JNK pathway regulation of TP53
Lysosome part
MAP kinases
MDM2−p53 pathway
NEST:156
NEST:211
NEST:246
NEST:302
NEST:366
Non−canonical Wnt signaling pathway
p300−CBP ID pathway association
PIK3−cyclin signaling
PIK3−EGFR signaling
PIK3−insulin receptor signaling
PIK3CA−actomyosin complex
PIK3CA−actomyosin complex II
PIK3CA−actomyosin core complex
Proximal RTK signaling
Regulation of actin−filament dynamics
Regulation of androgen receptor signaling pathway
Regulation of TP53 Activity
Resolution of D−loop structures
TP53 regulates transcription of cell cycle genes
Ubiquitin E3 ligase (SKP1, BTRC, FBXW11)
ML_MM
MM_MH
ML_MH
abs(log2FC)
1.5
2.0
2.5
3.0
p_value
0.006
0.005
0.004
0.003
0.002
0.001
0.000

Figure 5

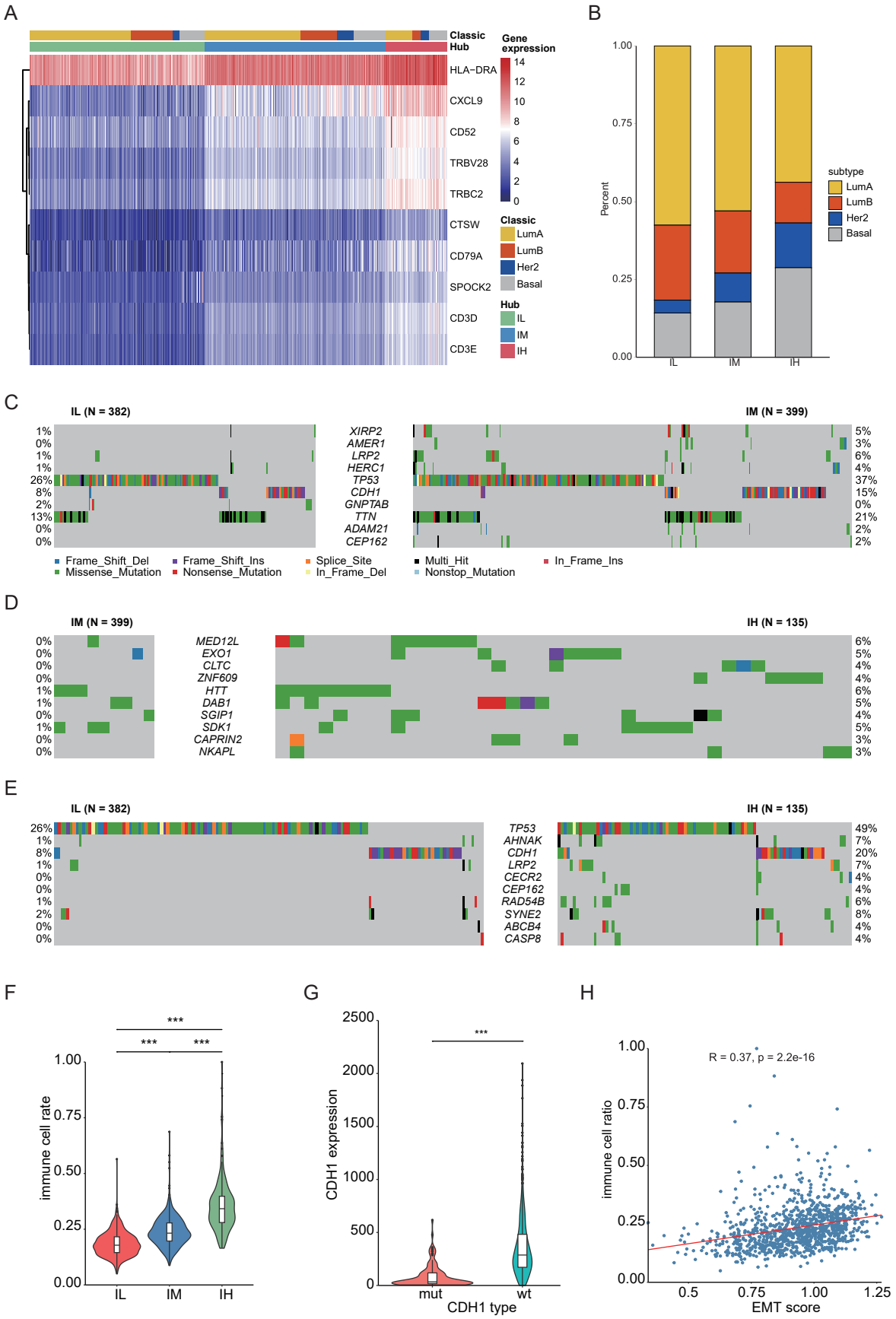
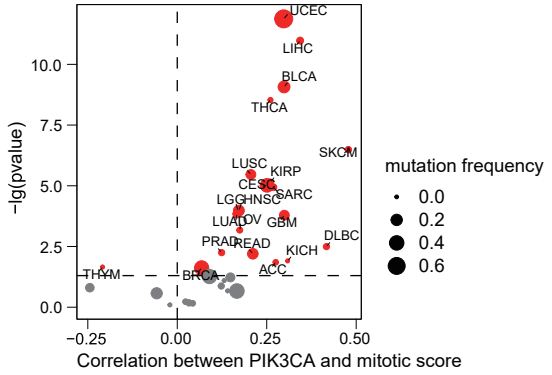
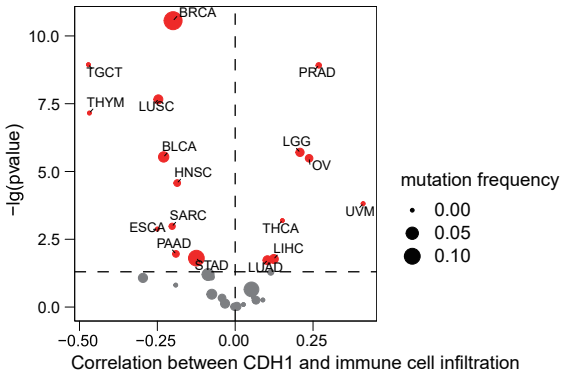

A
Classic
Hub
Gene expression
HLA-DRA
CXCL9
CD52
TRBV28
TRBC2
CTSW
CD79A
SPOCK2
CD3D
CD3E
Classic
LumA
LumB
Her2
Basal
Hub
IL
IM
IH
B
Percent
subtype
LumA
LumB
Her2
Basal
IL
IM
IH
C
IL (N = 382)
IM (N = 399)
XIRP2
AMER1
LRP2
HERC1
TP53
CDH1
GNPTAB
TTN
ADAM21
CEP162
Frame_Shift_Del
Frame_Shift_Ins
Splice_Site
Multi_Hit
In_Frame_Ins
Missense_Mutation
Nonsense_Mutation
In_Frame_Del
Nonstop_Mutation
D
IM (N = 399)
IH (N = 135)
MED12L
EXO1
CLTC
ZNF609
HTT
DAB1
SGIP1
SDK1
CAPRIN2
NKAPL
E
IL (N = 382)
IH (N = 135)
TP53
AHNAK
CDH1
LRP2
CECR2
CEP162
RAD54B
SYNE2
ABCB4
CASP8
F
immune cell rate
IL
IM
IH
G
CDH1 expression
mut
wt
CDH1 type
H
R = 0.37, p = 2.2e-16
immune cell ratio
EMT score

Figure 6

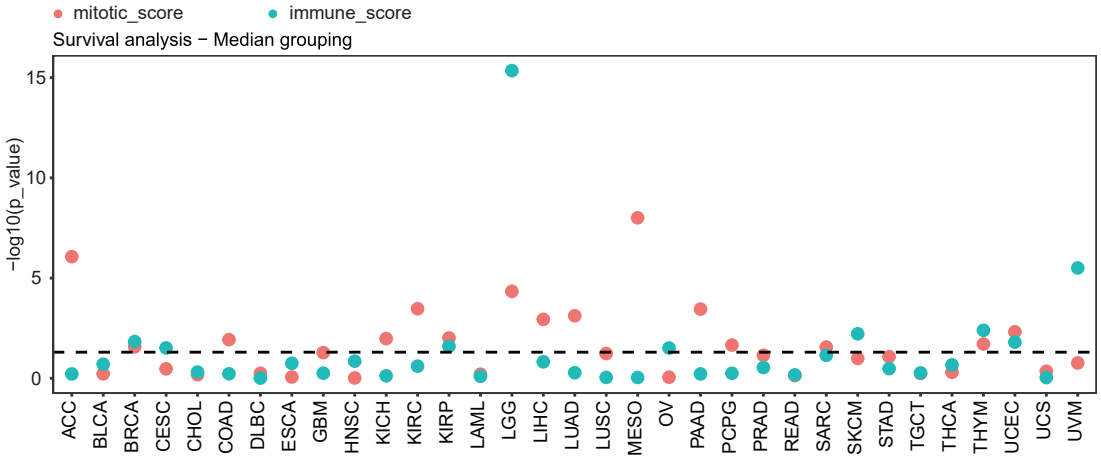
A


UCEC
LIHC
BLCA
THCA
SKCM
LUSC
KIRP
CESC
SARC
LGG
HNSC
LUAD
OV
GBM
DLBC
PRAD
READ
KICH
ACC
THYM
BRCA
−lg(pvalue)
10.0
7.5
5.0
2.5
0.0
−0.25
0.00
0.25
0.50
Correlation between PIK3CA and mitotic score
mutation frequency
0.0
0.2
0.4
0.6

B


BRCA
TGCT
PRAD
THYM
LUSC
BLCA
LGG
OV
HNSC
UVM
SARC
ESCA
THCA
PAAD
LIHC
STAD
LUAD
−lg(pvalue)
10.0
7.5
5.0
2.5
0.0
−0.50
−0.25
0.00
0.25
Correlation between CDH1 and immune cell infiltration
mutation frequency
0.00
0.05
0.10

C


mitotic_score
immune_score
Survival analysis − Median grouping
−log10(p_value)
15
10
5
0
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM

Multi-gradient Permutation Survival Analysis Identifies Mitosis and Immune Signatures Steadily Associated with Cancer Patient Prognosis

Xinlei Cai[1], Yi Ye[2], Xiaoping Liu[1], Zhaoyuan Fang[4,5], Luonan Chen[1,2,6], Fei Li[3*], Hongbin Ji[1,2,6*]

[1]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, 310024, China.

[2]State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China.

[3]Department of Pathology and Frontier Innovation Center, School of Basic Medical Sciences, Fudan University, Shanghai, China.

[4]Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Haining, 314400, China

[5]The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310000, China

[6]School of Life Science and Technology, Shanghai Tech University, Shanghai 200120, China.

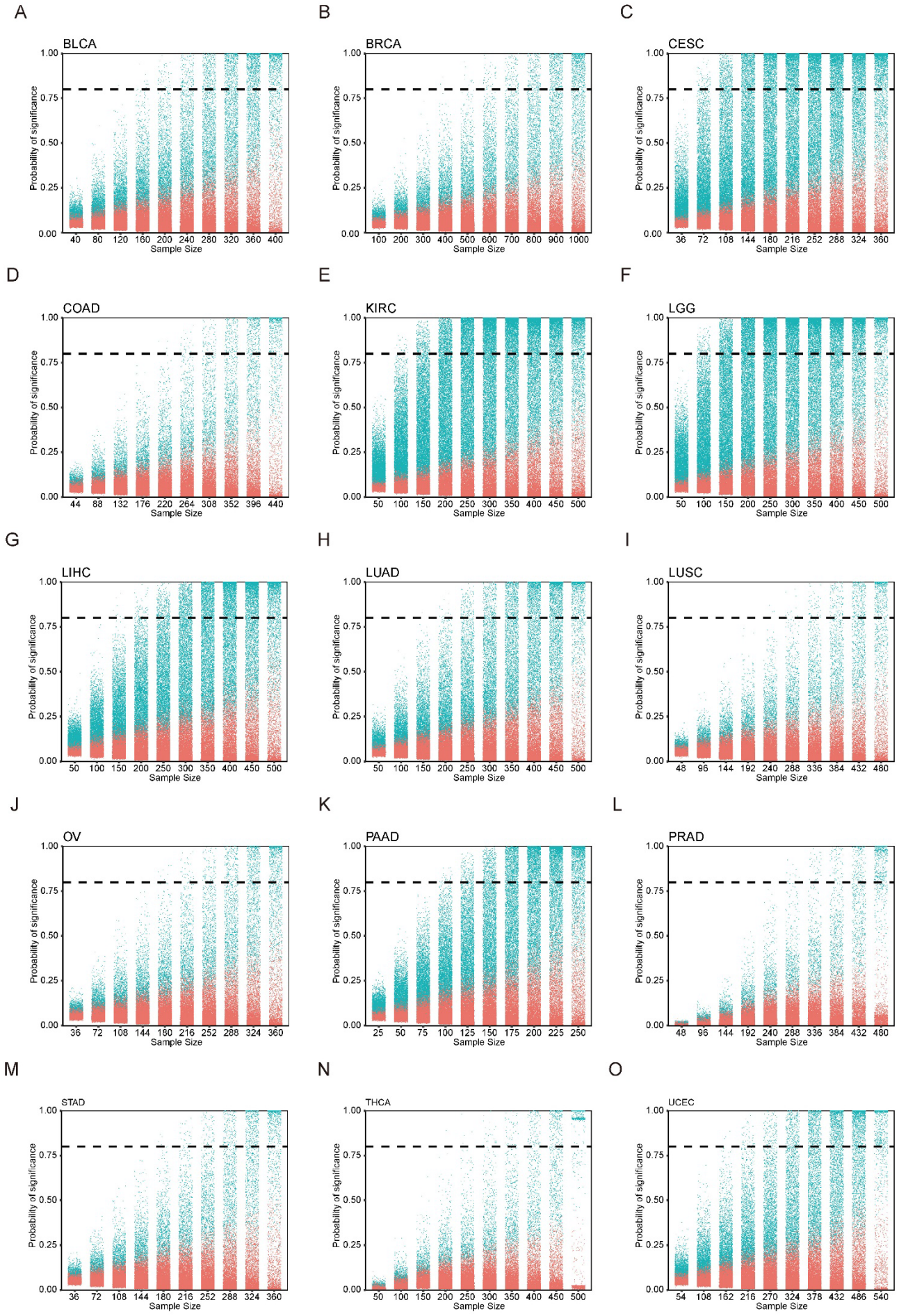
Running title: Mitosis and immune signatures in predicting cancer patient prognosis

***Corresponding author:**

Hongbin Ji, State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai, 200031, China, Phone: 86-21-54921108, Fax: 86-21-54921101, E-mail address: hbji@sibcb.ac.cn;

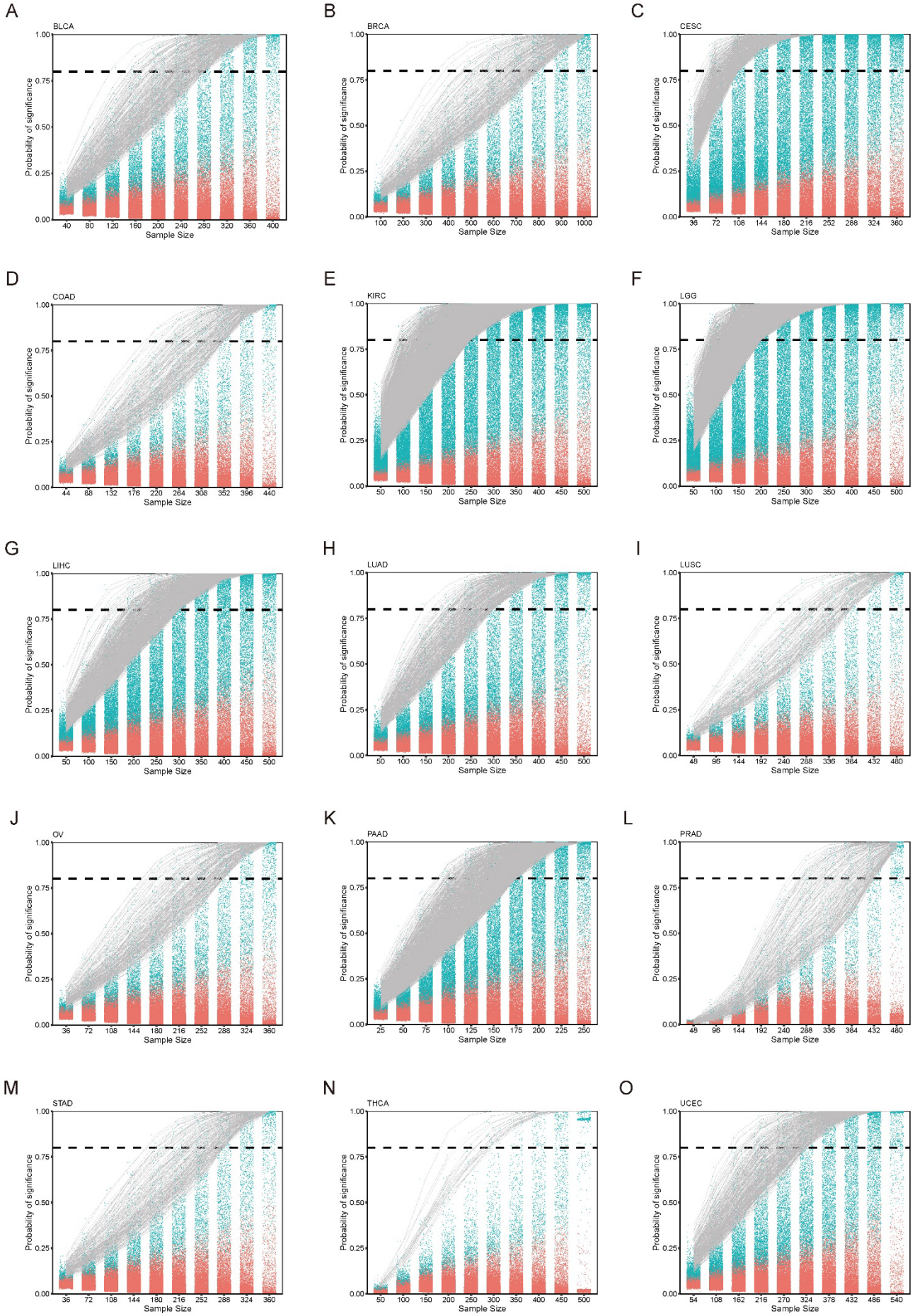
Fei Li, Department of Pathology and Frontier Innovation Center, School of Basic Medical Sciences, Fudan University, 131 Dong An Road, Shanghai, 200032, China, E-mail address: li_fei@fudan.edu.cn.

Supplementary Figure 1



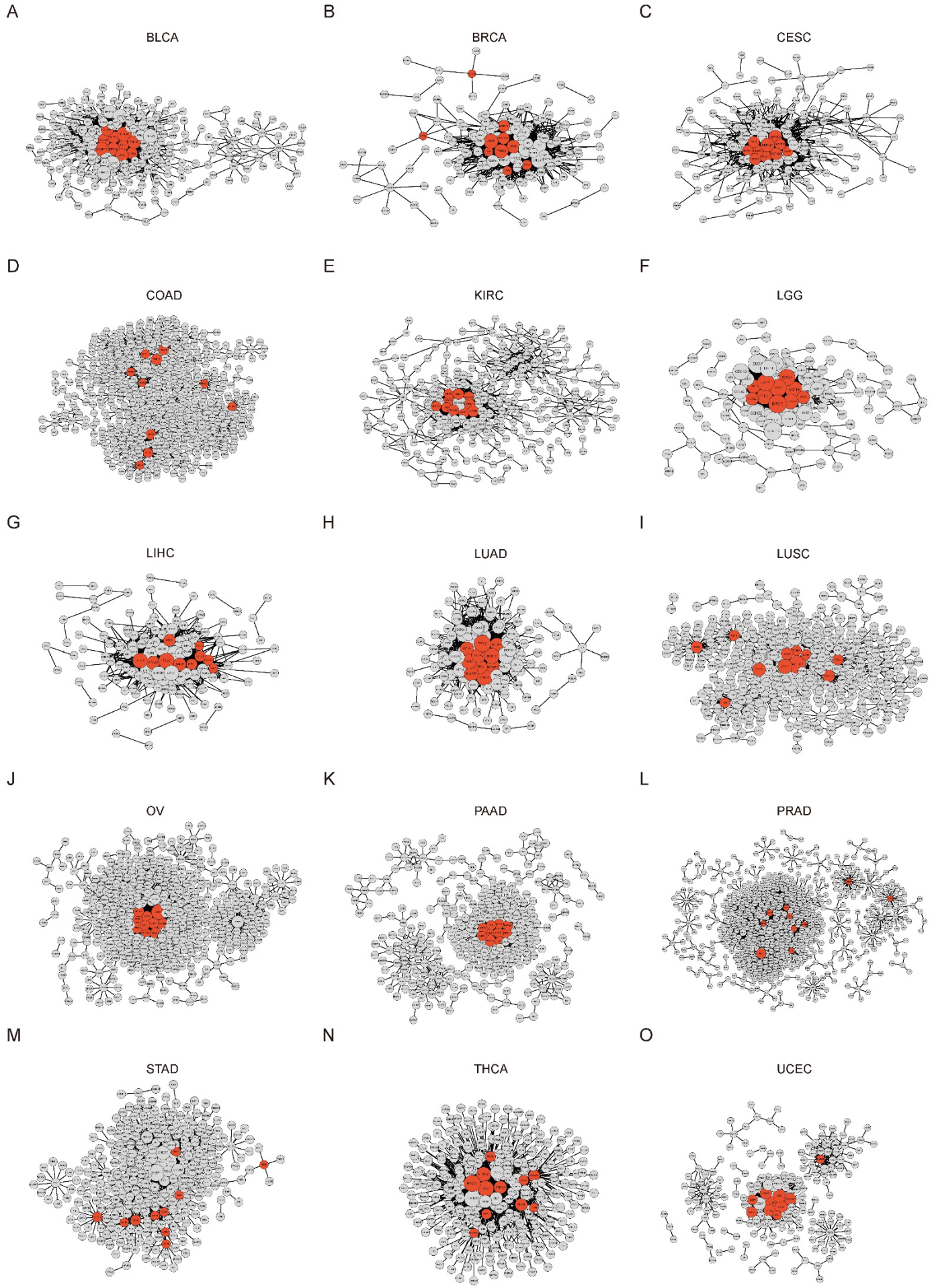
Supplementary Figure 1: (A) to (O) The number of genes significantly related to prognosis which were obtained through MEMORY was positively correlated with sample size. The vertical value of the dotted line is 0.8.

Supplementary Figure 2

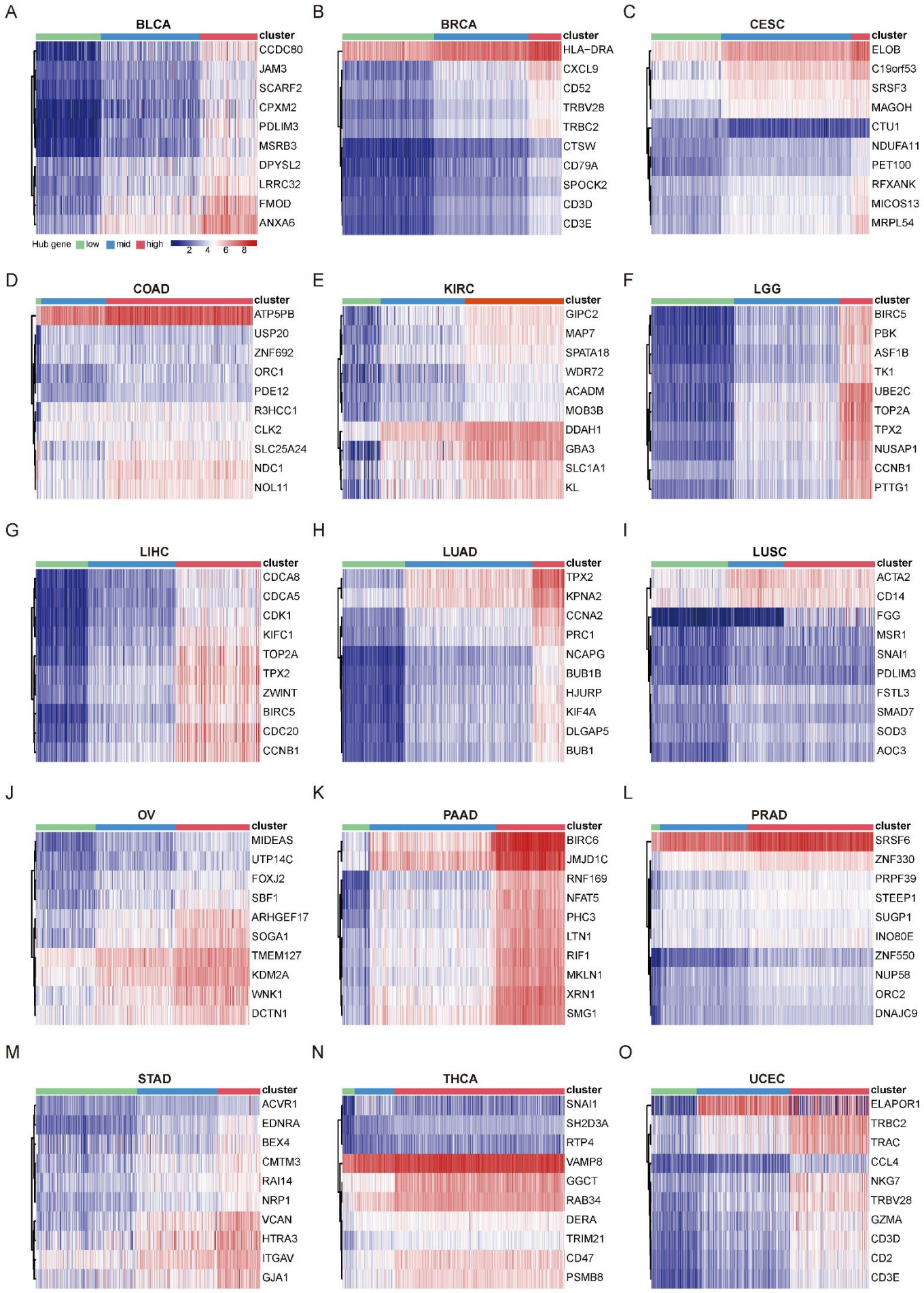


Supplementary Figure 2: (A) to (O) The significant probability of GEAR at each gradient. The vertical value of the dotted line is 0.8.

Supplementary Figure 3



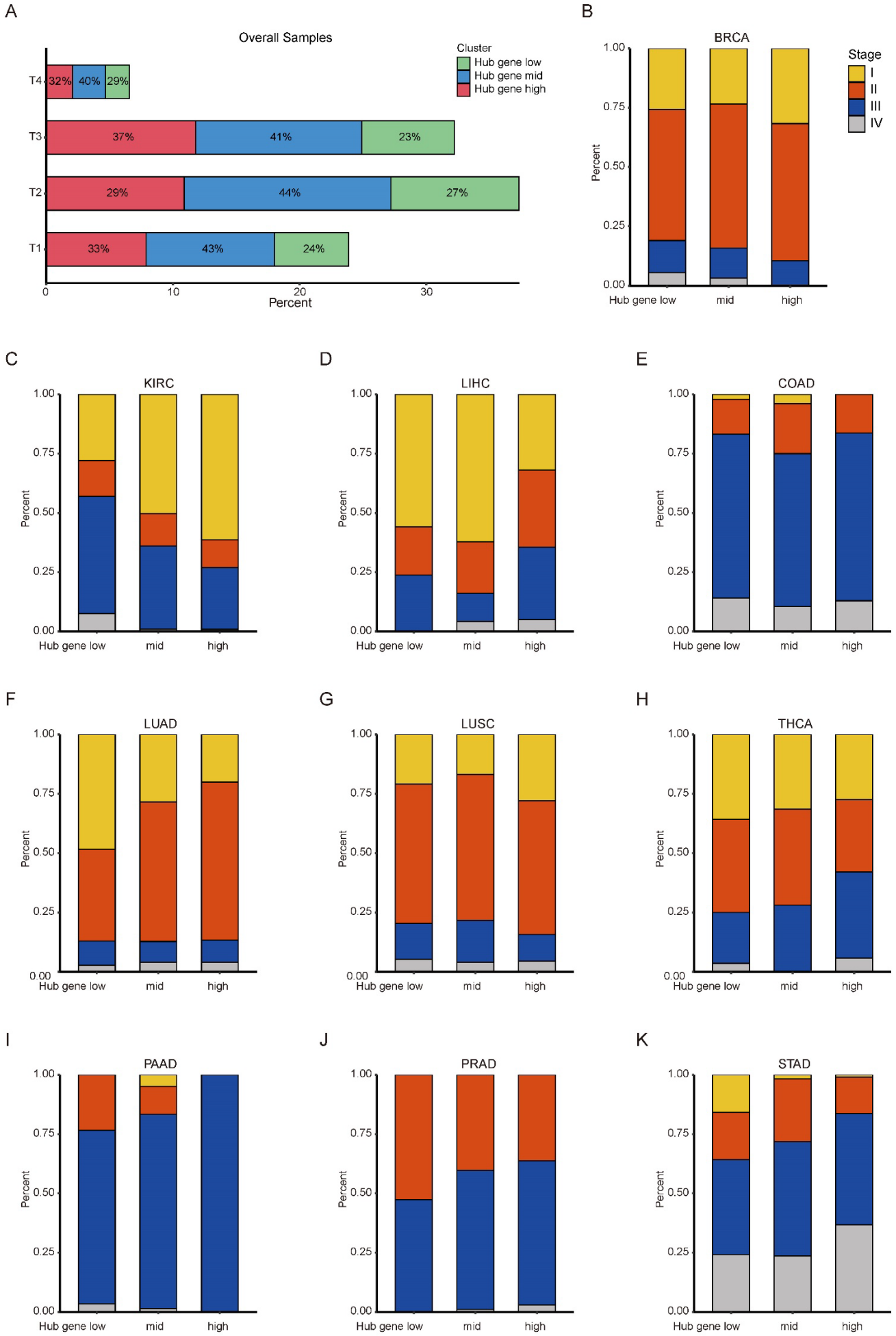
Supplementary Figure 3: (A) to (O) The core survival networks of 15 cancer types. Red nodes are hub genes.



Supplementary Figure 4

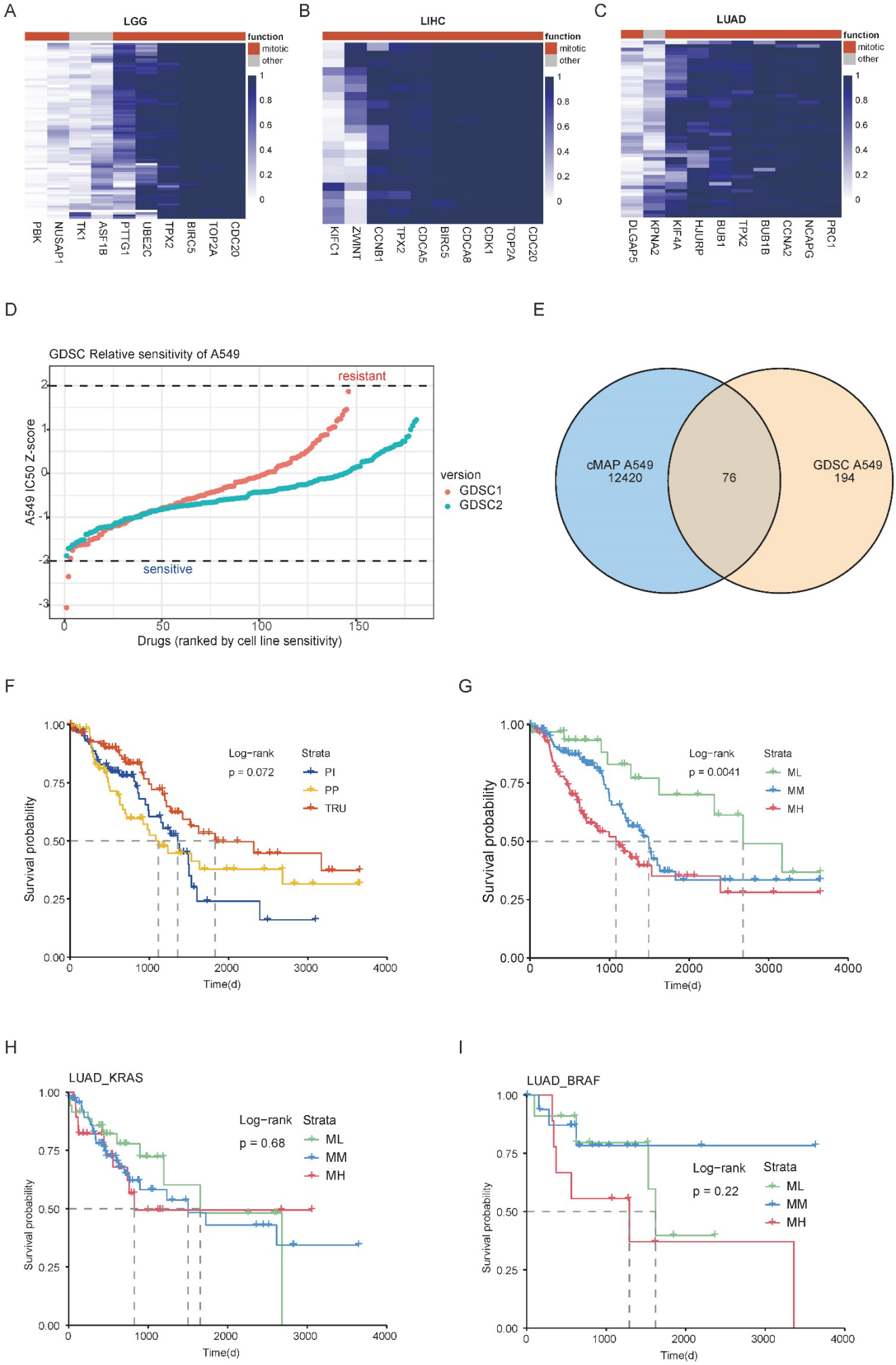
Supplementary Figure 4: (A) to (O) 15 cancer types were genotyped through hierarchical clustering based on the expression of hub genes, and the samples of each cancer were divided into three clusters

Supplementary Figure 5

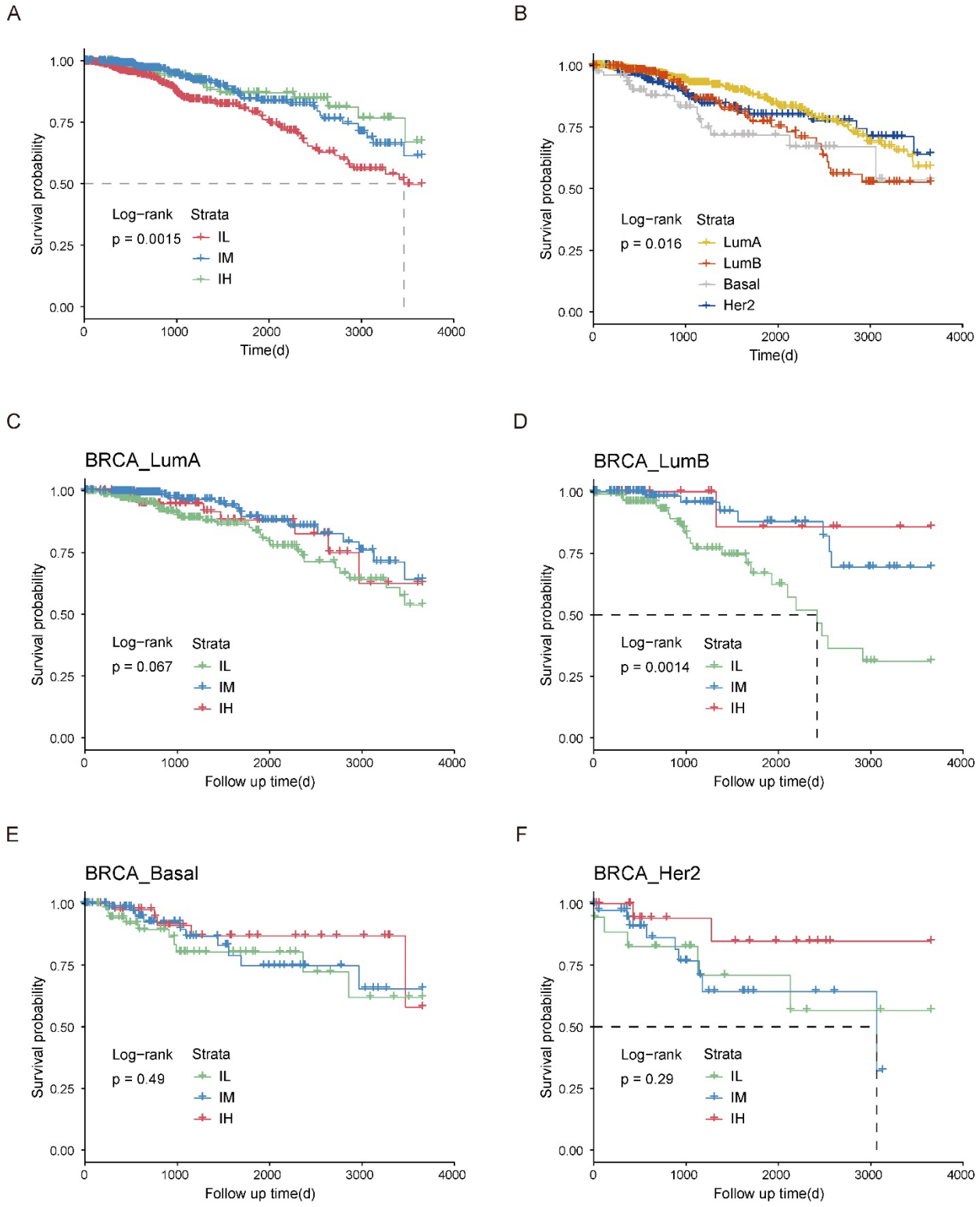


Supplementary Figure 5: (A) The proportion of hub gene classifications contained in different stages of pan-cancer level. (B) to (K) The proportion of cancers at different stages contained in different hub gene classifications.

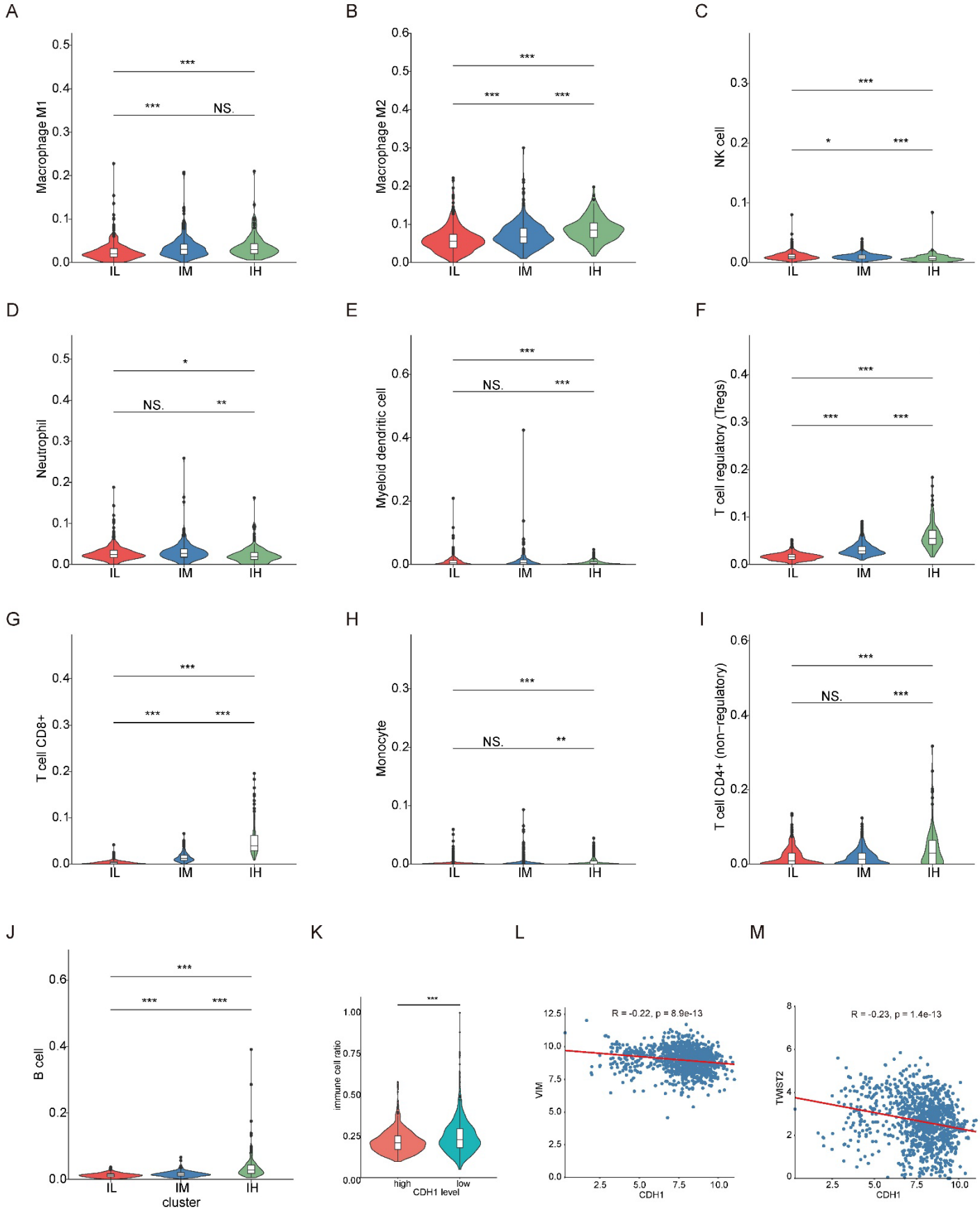
Supplementary Figure 6



Supplementary Figure 6: (A) to (C) The gene dependency analysis of hub genes for LGG (A), LIHC (B), LUAD (C). (D) The number of compounds associated with A549 in the cMAP and GDSC databases. (E) Relative sensitivity of A549 cell line to all compounds from the GDSC database, calculated based on IC50 values. (F) Kaplan-Meyer overall survival curves for LUAD classic subgroups. (G) Kaplan-Meyer overall survival curves for LUAD hub gene subgroups. (H) Kaplan-Meyer overall survival curves for *KRAS*-mutation samples of LUAD by hub gene subgroups. (I) Kaplan-Meyer overall survival curves for *BRAF*-mutation samples of LUAD by hub gene subgroups.



Supplementary Figure 7: (A) Kaplan-Meyer overall survival curves for BRCA classic subgroups. (B) Kaplan-Meyer overall survival curves for BRCA PAM50 subgroups. (C-F) Kaplan-Meyer overall survival curves for BRCA samples by hub gene subgroups in LumA (C), LumB (D), Basal (E), Her2 (F).



Supplementary Figure 8

Supplementary Figure 8: (A) to (J) Comparison of infiltration of different types of immune cells in three group BRCA samples. ***P < 0.0005, **P < 0.005 and *P < 0.05. (K) Comparison

of the proportion of immune cells in the total cells in BRCA samples with high *CDH1* expression and low *CDH1* expression. ***P < 0.0005. (L) Schematic diagram of the correlation between *CDH1* expression and *VIM* expression. The red line was a fitting curve. Pearson correlation coefficient was used for correlation analysis. (M) Schematic diagram of the correlation between *CDH1* expression and *TWIST2* expression. The red line was a fitting curve. Pearson correlation coefficient was used for correlation analysis.