



OPEN

## Investigating the role of Kinesin family in lung adenocarcinoma via integrated bioinformatics approach

Gulnaz Tabassum<sup>1,7</sup>, Prithvi Singh<sup>2,7</sup>, Rishabh Gurung<sup>2</sup>, Mohammed Ageeli Hakami<sup>3</sup>, Nada Alkhorayef<sup>3</sup>, Ahad Amer Alsaiani<sup>4</sup>, Leena S. Alqahtani<sup>5</sup>, Mohammad Raghiful Hasan<sup>3</sup>, Summya Rashid<sup>6</sup>, Atul Kumar<sup>1</sup>, Kapil Dev<sup>1</sup>✉ & Ravins Dohare<sup>2</sup>✉

Lung cancer is the leading cause of mortality from cancer worldwide. Lung adenocarcinoma (LUAD) is a type of non-small cell lung cancer (NSCLC) with highest prevalence. Kinesins a class of motor proteins are shown to be involved in carcinogenesis. We conducted expression, stage plot and survival analyses on kinesin superfamily (KIF) and scrutinized the key prognostic kinesins. Genomic alterations of these kinesins were studied thereafter via cBioPortal. A protein–protein interaction network (PPIN) of selected kinesins and 50 closest altering genes was constructed followed by gene ontology (GO) term and pathway enrichment analyses. Multivariate survival analysis based on CpG methylation of selected kinesins was performed. Lastly, we conducted tumor immune infiltration analysis. Our results found *KIF11/15/18B/20A/2C/4A/C1* to be significantly upregulated and correlated with poor survival in LUAD patients. These genes also showed to be highly associated with cell cycle. Out of our seven selected kinesins, *KIFC1* showed the highest genomic alteration with highest number of CpG methylation. Also, CpG island (CGI) cg24827036 was discovered to be linked to LUAD prognosis. Therefore, we deduced that reducing the expression of *KIFC1* could be a feasible treatment strategy and that it can be a wonderful individual prognostic biomarker. CGI cg24827036 can also be used as a therapy site in addition to being a great prognostic biomarker.

Lung cancer (LC) is a prevalent and deadly disease that ranks first among cancers in terms of death and 2nd most diagnosed cancer in both genders globally<sup>1</sup>. The etiological and molecular heterogeneity of LC contributes greatly to treatment failure and adverse survival outcomes<sup>2,3</sup>. Most LCs diagnosed are malignant epithelial tumours, which can be further classified as small-cell lung carcinoma (SCLC) or non-small cell lung carcinoma (NSCLC). NSCLC accounts for 85–90% of lung malignancies, with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) being the most frequent subtypes<sup>4,5</sup>. LUAD and LUSC can be classified into four stages, referred to as I, II, III, and IV, as per the tumor node metastasis (TNM) taxonomy<sup>6</sup>. The early, non-metastatic stage is referred to as stage I. Stages II and III typically represent the intermediate, regional lymphatic metastatic phases, with stage III exhibiting more significant metastasis in the lymphatic region than stage II. Meanwhile, stage IV often denotes a late stage with distant metastases<sup>5</sup>.

Despite evidence that smoking increases the risk of LUAD, it is currently the most common subgroup of LC among non-smokers and women<sup>7,8</sup>. Patients with LUAD typically have a poor prognosis and frequently show local progression or metastasis when diagnosed<sup>9</sup>. However, LUSC is more prevalent in men than in women and has been strongly linked to smoking<sup>10</sup>. Although chemotherapy, radiation, and targeted medicines are widely

<sup>1</sup>Department of Biotechnology, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India. <sup>2</sup>Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India. <sup>3</sup>Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Al-Quwayyah, Shaqra University, Riyadh 13343, Saudi Arabia. <sup>4</sup>Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taif University, Taif 21944, Saudi Arabia. <sup>5</sup>Department of Biochemistry, College of Science, University of Jeddah, Jeddah 23445, Saudi Arabia. <sup>6</sup>Department of Pharmacology and Toxicology, College of Pharmacy, Prince Sattam Bin Abdulaziz University, Alkharij 16278, Saudi Arabia. <sup>7</sup>These authors contributed equally: Gulnaz Tabassum and Prithvi Singh. ✉email: kdev@jmi.ac.in; ravinsdohare@gmail.com

employed, therapeutic resistance to these treatments is a primary cause of treatment failure. Understanding the underlying molecular pathways of carcinogenesis is thus critical for developing effective LC therapies.

Human kinesin superfamily members (KIFs) consist of 14 kinesin family members, kinesin-1 to kinesin-14, according to the standardized nomenclature adopted by the kinesin research group<sup>11</sup>. There are 45 members in the KIFs superfamily, including 39 N-kinesins, three M-kinesins, and three C-kinesins<sup>12</sup>. KIF proteins are a family of motor proteins that move molecules and depend on microtubules. They have ATPase activity as well as motion characteristics. They bind to microtubules and then move along the microtubules, carrying protein complexes, organelles, and messenger RNAs (mRNAs)<sup>12–14</sup>. In recent years, it has come to light that several KIFs contribute uniquely to the process of mitosis, also known as cell division, by taking part in the motion of chromosomes and spindles<sup>15,16</sup>. Additionally, individual kinesins are also essential for a number of other cellular processes, such as endocytosis and transcytosis, intracellular transport<sup>14</sup>.

Mitosis, the process by which eukaryotic cells divide, creates two daughter cells with approximately equal amounts of the cell's nucleus, cytoplasm, organelles, and membrane. It is possible that mistakes in this process could lead to the death of cell, abnormalities (including deletion of gene, translocation of chromosome, or the duplication of chromosomes), and even cancer. Since mitosis is so intricately controlled therefore, any alteration or changes in KIF expression or function could potentially cause cancer. Kinesins and motor proteins with abnormal expression are crucial mitotic process regulators and potential targets in human malignancies<sup>17–19</sup>. Human cancer is a genetic disorder characterized by uncontrolled cell development, hence inhibiting kinesins may provide a unique approach to managing this disease.

Thus, identifying anomalous kinesin gene expression could be utilized as a biomarker for early tumor diagnosis and targeting kinesins could also be a novel approach for cancer therapy. Therefore, in the current study, we conducted a comprehensive bioinformatics analysis to identify the key kinesins influencing the prognosis of LUAD cancer patients. We performed expression and stage plot analyses of KIFs across the cancer genome atlas (TCGA)-LUAD patient samples and reported only significant ones. Next, we proceeded with overall survival (OS) analysis followed by mutational, enrichment, and protein–protein interaction network (PPIN) analyses. At last, we obtained *KIFC1* as final prognostic biomarker responsible for LUAD pathogenesis. *KIFC1* can be further used for early detection of LUAD patients and targeted therapy or personalized medicine.

## Materials and methods

**Kinesins expression and stage analysis across LUAD cohort.** Gene expression profiling interactive analysis v2 (GEPIA 2) web-based tool<sup>20</sup> (<http://gepia2.cancer-pku.cn/>) was accessed for comparing the relative mRNA expression level of all kinesin family members across TCGA-LUAD cohort and matched TCGA normal and GTEx data. The expression values from GEPIA were already transformed into  $\log_2(\text{TPM} + 1)$  values followed by differential analysis. Pathological stage plot analysis was also done with GEPIA 2 to investigate the kinesin family members' expression with respect to different pathological stages in LUAD. The threshold used in GEPIA for mRNA expression level comparison across LUAD and normal samples were as follows:  $p\text{-value} < 0.05$  and  $|\log_2(\text{foldchange})| > 1$ . Kinesins statistically significant in both expression and stage plot analyses were selected for further analyses.

**Prognostic analysis of kinesins across LUAD cohort.** Kaplan–Meier (KM) plotter<sup>21,22</sup> (<https://kmpplot.com/analysis/>) was queried for prognostic analysis of kinesins having significance in expression and stage plot analyses. We generated KM plots of only those kinesins which showed significant OS across LUAD patient samples. The microarray LUAD patients were bifurcated into higher and lower expression groups based on their median values. The redundant samples were removed in the quality control section, and biased arrays were excluded. Hazard ratio (HR) with the corresponding 95% confidence interval (CI), logrank  $p$  value and median survival were calculated. logrank  $p$  value  $< 0.05$  was considered as a statistically significant threshold for assessing the prognosis of kinesins between two expression groups.

**Validation of prognostic kinesins using cBioPortal.** We queried the cBioPortal for Cancer Genomics<sup>23</sup> (<https://www.cbioportal.org/>) for investigating the mutations and putative copy number alterations (CNAs) of prognostically significant kinesins. The LUAD dataset (TCGA, Firehose Legacy) was chosen to perform our analysis.

**Validation of prognostic kinesins using GEO and correlation analysis.** We queried the NCBI-GEO<sup>24</sup> (<https://www.ncbi.nlm.nih.gov/geo/>) using “LUAD” and “Lung Adenocarcinoma” as suitable keywords for extracting LUAD-associated mRNA expression profile. All the search results were further trimmed down in accordance with the following inclusion criteria: (1) the samples present in dataset(s) must belong to ‘Homo Sapiens’; (2) dataset(s) type must be ‘expression profiling by array’; (3) both preprocessed and raw files of the dataset(s) must be available; (4) the dataset(s) submission date to GEO must be within last 10 years (i.e. 2012–2022); (5) the dataset(s) must be comprising both tumor and healthy control tissue samples; (6) the dataset(s) must comprise at least 25 samples. Any abstracts, case reports, review-based articles, cell-line-based experimental study designs, and studies devoid of healthy controls or non-human samples were excluded. Sequential steps of batch correction, probe ID to gene mapping, and duplicacy removal were performed as discussed previously<sup>25</sup>. The DEGs were screened corresponding to a Benjamini-Hochberg (BH)— $p$  value  $< 0.05$  and  $|\log_2(\text{fold change})| > 0.5$  utilizing limma<sup>26</sup>. The presence of key prognostic kinesins was checked in the DEGs list. Next, we accessed GEPIA 2 to perform pairwise correlation analysis of key prognostic kinesins across TCGA-LUAD and normal patients.  $p$  value  $< 0.05$  was considered as the cutoff for statistical significance.

**PPIN construction and enrichment analysis.** A PPIN was constructed between the prognostically significant kinesins and top 50 frequently altered genes corresponding to a default confidence (i.e., interaction score > 0.4) using Search Tool for the Retrieval of Interacting Genes (STRING) v1.1.5 web-based tool<sup>27</sup> (<https://string-db.org/>) and visualized via Cytoscape v3.9.1<sup>28</sup>. Top 10 significant (i.e.,  $p$  - value < 0.05) pathway and gene ontology (GO) terms for the constructed PPIN items were compiled using Enrichr web server<sup>29</sup> (<https://maayanlab.cloud/Enrichr>). Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>30–32</sup>, GO-Biological Process (BP), GO-Molecular Function (MF), and GO-Cellular Compartment (CC) libraries were used for pathway and GO terms.

**Tumor infiltration analysis.** We looked into the relationship between mRNA expression levels of prognostically significant kinesins with tumor-infiltrating immune cells such as B cells, CD8<sup>+</sup>T cell, macrophage, and neutrophils across TCGA-LUAD patients using TIMER 2.0<sup>33</sup> (<http://timer.cistrome.org/>). To assess the statistical significance, Spearman correlation was used.

**Methylation analysis.** Prognostic analysis of single CpG methylation of selected genes of kinesin family in LUAD patients was conducted using MethSurv<sup>34</sup> (<https://biit.cs.ut.ee/methsurv>), a web tool for multivariate survival analysis based on CpG methylation data.

## Results

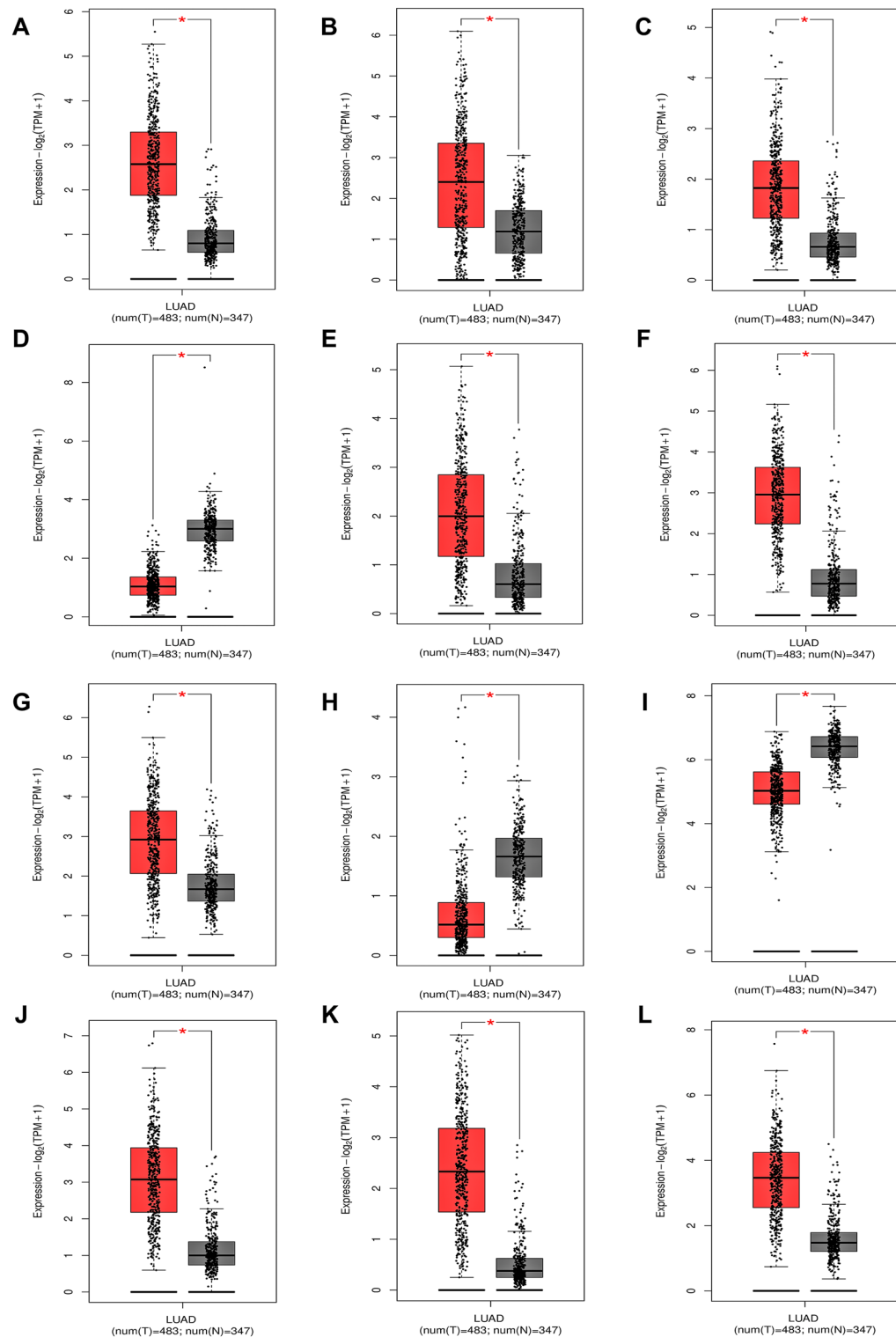
**Kinesins expression and stage plot analysis across LUAD cohort.** All kinesins' relative mRNA expression distribution across TCGA-LUAD cohort (483 tumor and 347 normal) was compiled utilizing GEPIA. *KIF11*, *KIF12*, *KIF15*, *KIF23*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, *KIFC1* expression levels were significantly upregulated while *KIF17*, *KIF26A*, *KIF1C* expressions were significantly downregulated in tumor samples as shown by the box-and-whisker plots in Fig. 1A–L. All these significantly expressed kinesins were carried further to stage plot analysis. The pathological sub-stage analysis as shown by violin plots in Fig. 2A–H revealed that over-expressed levels of *KIF11*, *KIF15*, *KIF23*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, *KIFC1* significantly correlated with advanced TNM stages across TCGA-LUAD cohort.

**Prognostic analysis of kinesins across LUAD cohort.** Using KM plotter, prognostic analysis was performed on *KIF11*, *KIF15*, *KIF23*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, *KIFC1* to determine the correlation between their mRNA expression levels and risk of 513 LUAD patient samples. The KM plots as shown in Fig. 3A–G revealed significantly poor OS of LUAD patients when mRNA expression levels of *KIF11*, *KIF15*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, and *KIFC1* were high. The low and high expression cohort median survival time, HR, 95%CI, and logrank  $p$  - value of each kinesin is detailed in Supplementary Table S1, respectively.

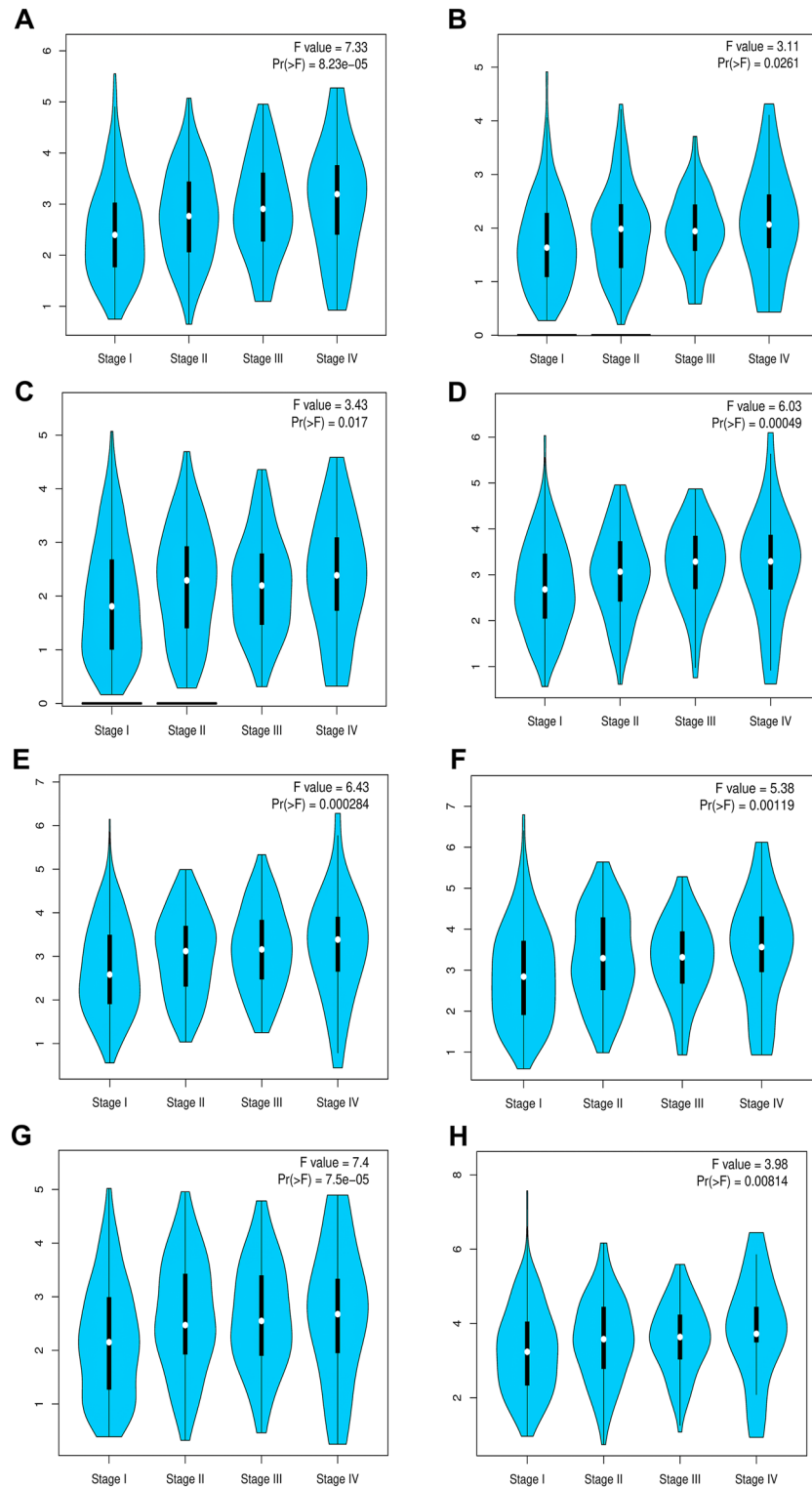
**Validation of key prognostic kinesins using cBioPortal.** We used cBioPortal to validate the specific genetic modifications associated with key prognostic kinesins (i.e., *KIF11*, *KIF15*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, *KIFC1*) across LUAD dataset (TCGA, Firehose legacy) comprising 584 tumor patient samples. OncoPrint results for these queried genes as represented in Fig. 4 revealed genetic alterations in 8% (49/584) patient samples. As observed, *KIFC1* showed maximum mutation frequency (2.3%) as compared to others. The cancer type summary analysis revealed the overall alteration frequency of these genes as shown in Supplementary Figure S1. We observed 0.78% (4/516 cases) missense mutation and 0.39% (2/516 cases) deep deletion in case of *KIF11*. In case of *KIF15*, we observed 0.58% (3/516 cases) missense mutation and 0.19% (1/516 case) deep deletion. In case of *KIF18B*, we observed 0.78% (4/516 cases) amplification and 0.58% (3/516 cases) missense mutation. In case of *KIF20A*, we observed 0.58% (3/516 cases) missense mutation, 0.39% (2/516 cases) deep deletion, and 0.19% (1/516 case) amplification. In case of *KIF2C*, we observed 0.19% (1/516 case) truncating mutation and 1.55% (8/516 cases) amplification. In case of *KIF4A*, we observed 0.19% (1/516 case) deep deletion, 0.39% (2/516 cases) amplification, and 1.36% (7/516 cases) missense mutation. In case of *KIFC1*, we observed 1.55% (8/516 cases) amplification and 0.78% (4/516 cases) missense mutation.

**Validation using GEO and correlation analysis.** As per the specified inclusion and exclusion criteria we chose GSE43458 (30 healthy control + 80 tumor tissues) and GSE116959 (11 healthy control + 57 tumor tissues) LUAD-associated mRNA expression profiles. A total of 2861 and 5128 DEGs were screened corresponding to GSE43458 and GSE116959 as per the specified threshold. The lists of DEGs are shown in Supplementary Tables S2 and S3. All the key prognostic kinesins (i.e., *KIF11*, *KIF15*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, and *KIFC1*) were present in the DEGs lists of both datasets, thus confirming their validation in external GEO datasets. Strikingly, all the prognostic kinesins were upregulated among DEGs list and matched with the primary results obtained from GEPIA 2. Scatterplots showing pairwise correlations among these key prognostic kinesins are demonstrated in Supplementary Figures S2–S5. Significantly highest correlation between *KIF4A* and *KIF2C* ( $R = 0.95$ ,  $p$  value =  $6.9 \times 10^{-269}$ ) was observed.

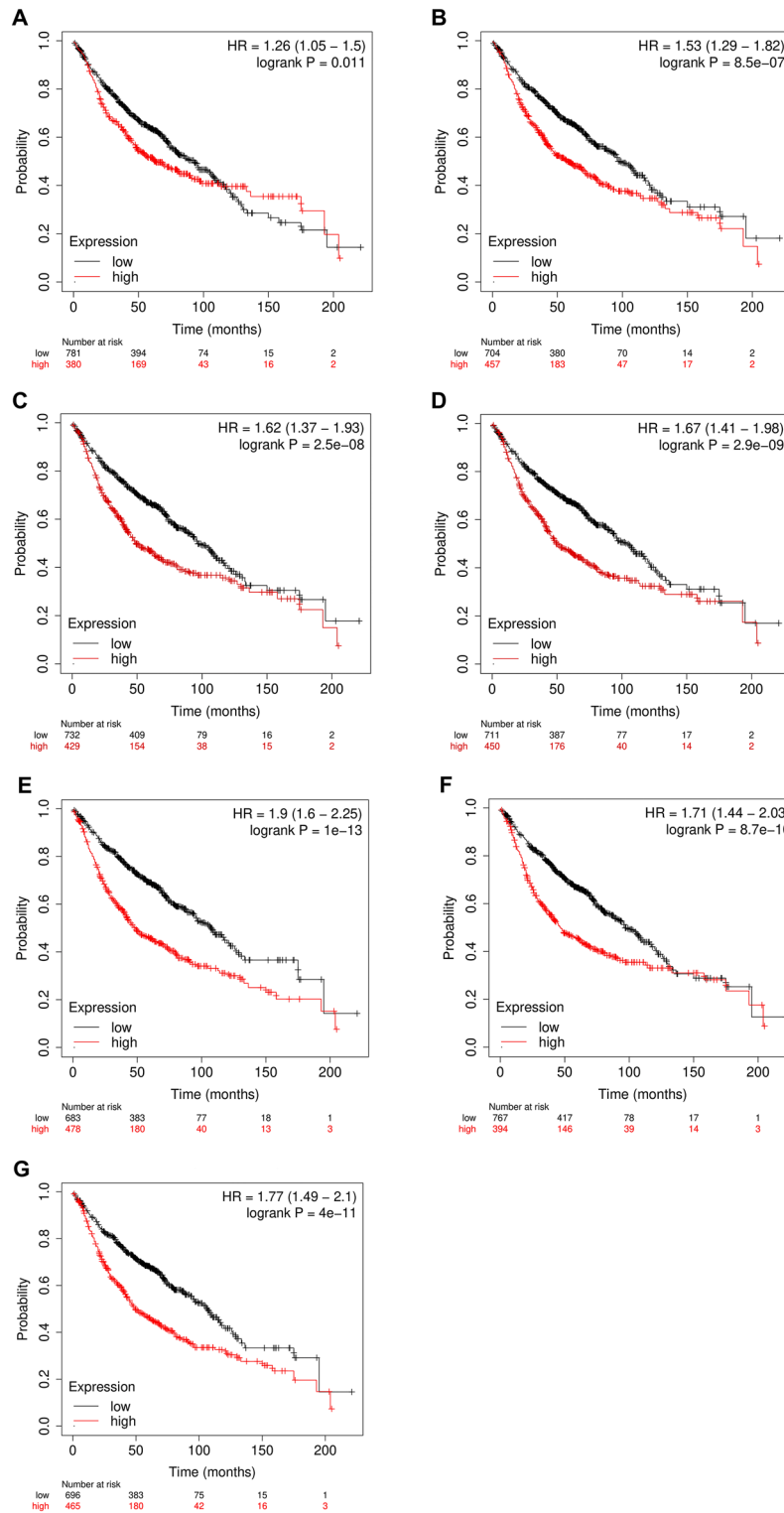
**PPIN construction and enrichment analysis.** Our PPIN comprised a total of 57 nodes and 1455 edges as shown in Fig. 5. Within PPIN, degree, betweenness, and closeness values ranged from 4 to 56, 0.07 to 40.44, and 0.51 to 1. The average degree, betweenness, and closeness of PPIN were 51.05, 4.94, and 0.931. Topological/centrality measures like node degree, betweenness, closeness, clustering coefficient, neighborhood connectivity, and average shortest path length of PPIN are demonstrated in Supplementary Figure S6. Subsequently, we performed pathway and GO term enrichment analysis on key prognostic kinesins and associated top 50



**Figure 1.** Box-and-whisker plots displaying the relative mRNA expression levels of (A) KIF11, (B) KIF12, (C) KIF15, (D) KIF17, (E) KIF18B, (F) KIF20A, (G) KIF23, (H) KIF26A, (I) KIF1C, (J) KIF2C, (K) KIF4A, (L) KIF1C1 across TCGA-LUAD and normal samples. Grey-and red-colored box areas signify normal and tumor patient samples. The top and bottom of the boxes signify 75th and 25th percentile of distribution. Horizontal lines within the boxes represent the median values while minimum and maximum values label the axes endpoints. \* $p$  value < 0.05.

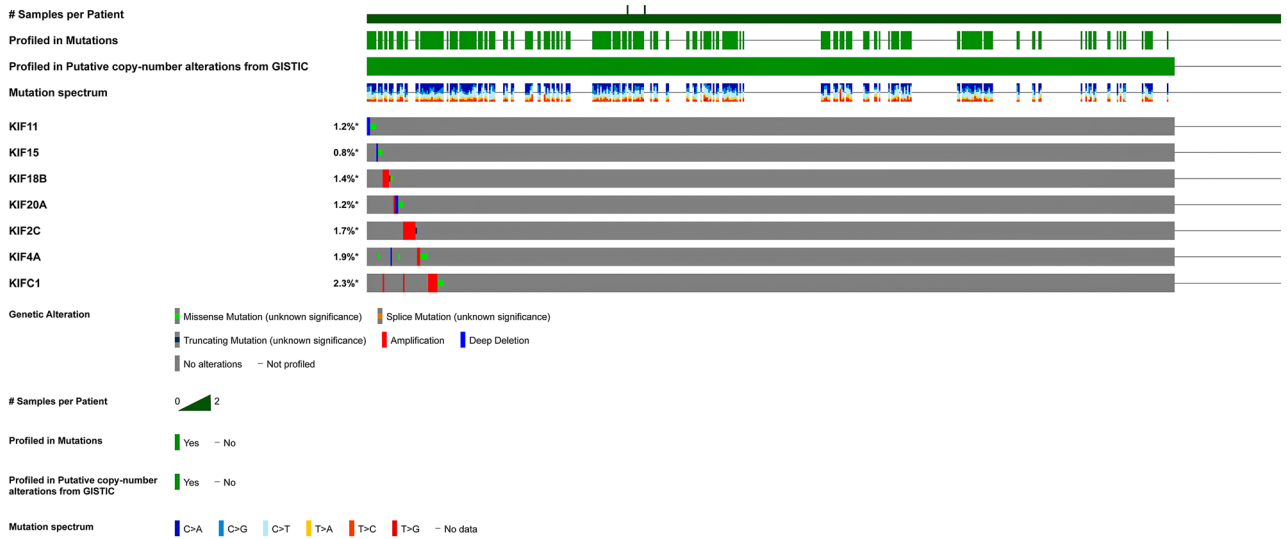


**Figure 2.** Violin plots displaying association between significant TNM sub-stages and mRNA expression levels of (A) KIF11 (B) KIF15, (C) KIF18B, (D) KIF20A, (E) KIF23, (F) KIF2C, (G) KIF4A, (H) KIF1C across TCGA-LUAD cohort. The black-colored vertical bars and white-colored dots signify interquartile ranges and median, respectively. The ordinate and abscissa depict expression levels of these genes and various stages. Distribution density is represented by the width of turquoise-colored shapes, respectively.

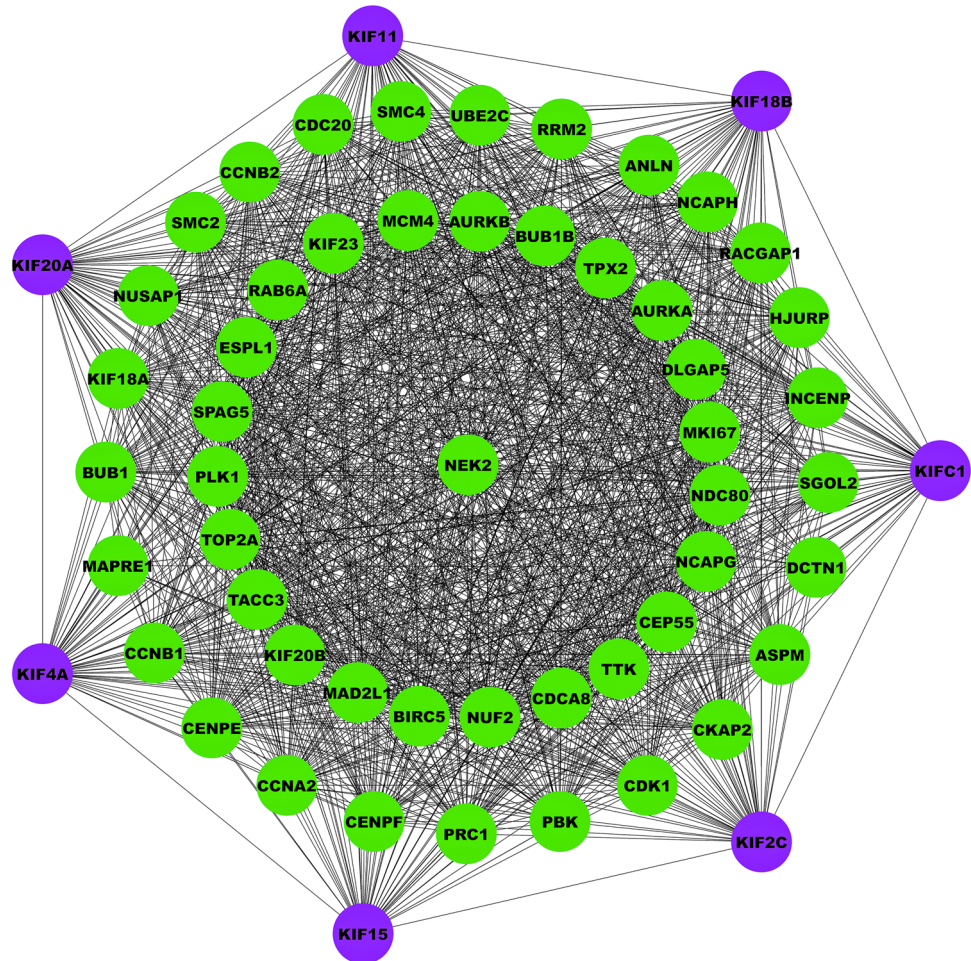


**Figure 3.** KM plots showing the OS of (A) KIF11 (B) KIF15, (C) KIF18B, (D) KIF20A, (E) KIF2C, (F) KIF4A, (G) KIFC1 across LUAD microarray cohort. Red and black colors signify higher and lower expression groups.

frequently altered genes. Barplots showing top 10 significantly enriched pathway and GO terms is shown in Fig. 6. The most significant pathway, GO-BP, GO-MF, GO-CC terms were cell cycle ( $p$  value =  $4.8 \times 10^{-14}$ ), microtubule cytoskeleton organization involved in mitosis ( $p$  value =  $1.44 \times 10^{-38}$ ), microtubule binding ( $p$  value =  $1.54 \times 10^{-21}$ ), spindle ( $p$  value =  $5.46 \times 10^{-36}$ ). Most number of genes corresponding to pathway,



**Figure 4.** OncoPrint summarizing genomic alterations of key prognostic kinesins across TCGA-LUAD cohort comprising 584 patient samples. The bottom row represents frequency of genomic alterations in KIF11, KIF15, KIF18B, KIF20A, KIF2C, KIF4A, KIFC1 with red, blue, green, orange, and grey bars signifying amplifications, deep deletions, missense, splice, and truncating mutations, respectively. First, second, third, fourth, and fifth rows depicts the clinical annotation bars such as profiled in putative copy-number alterations from GISTIC, mutation spectrum, sex, tissue source site, and mutation count, respectively.

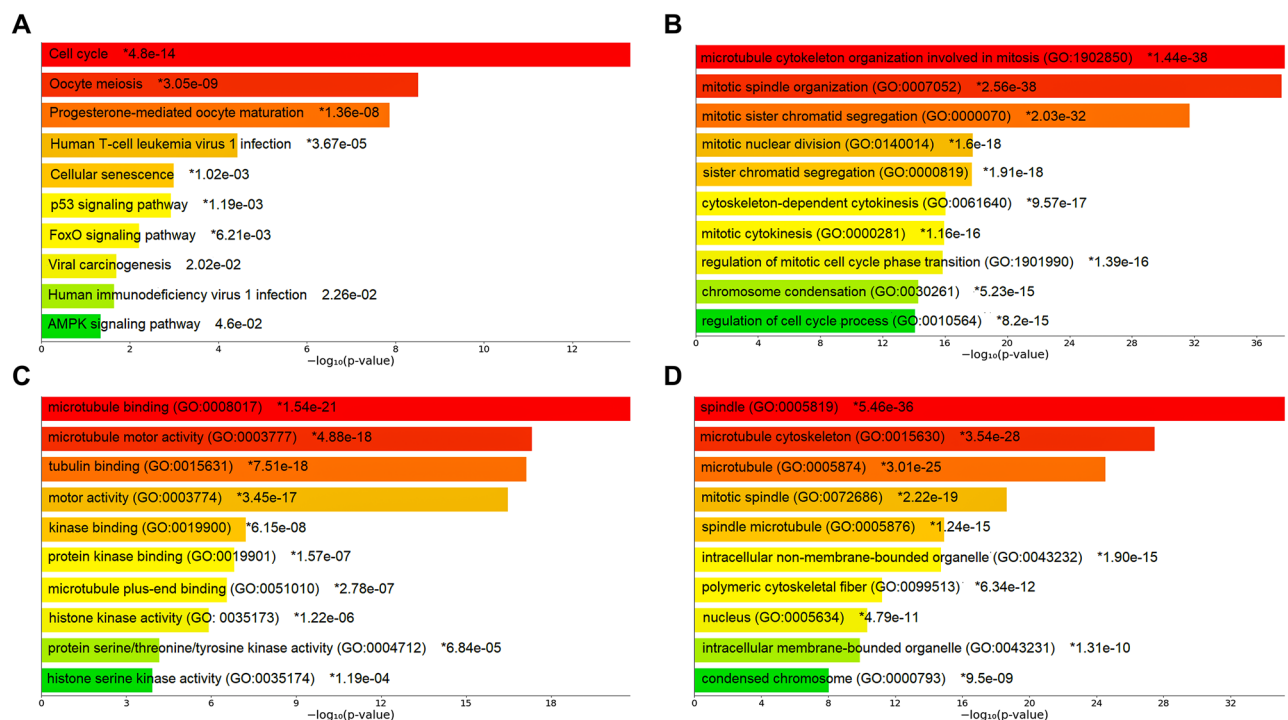


**Figure 5.** PPIN comprising 57 nodes and 1455 edges. Magenta-colored nodes represent prognostic kinesins and green-colored nodes represent top 50 frequently altered genes.

GO-BP, GO-MF, GO-CC terms were 11, 25, 18, 38 for cell cycle, mitotic spindle organization, microtubule binding, intracellular membrane-bounded organelle.

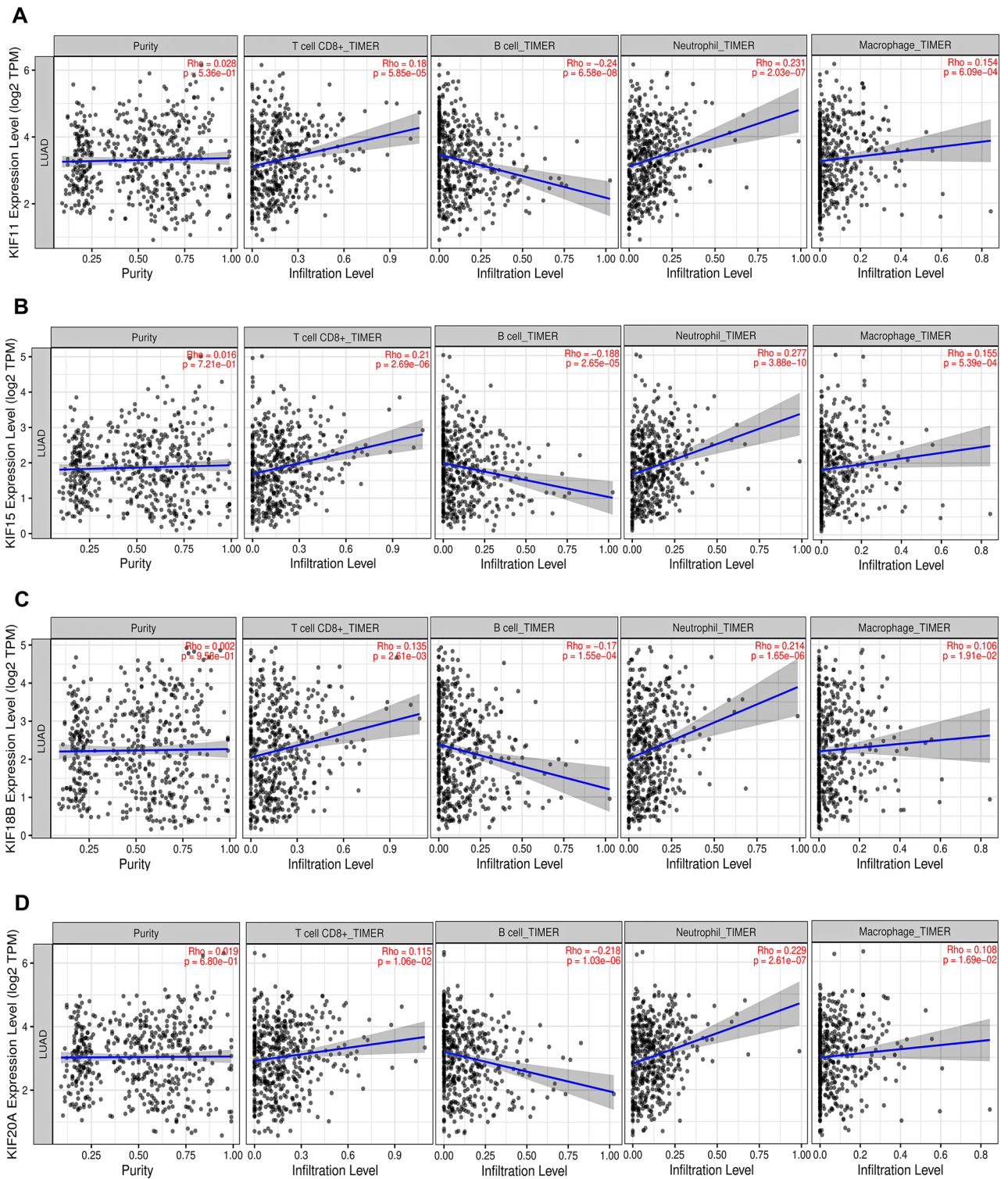
**Tumor infiltration analysis.** Correlation of *KIF11*, *KIF15*, *KIF18B*, *KIF20A*, *KIF2C*, *KIF4A*, *KIFC1* mRNA expression levels with tumor purity and infiltrating levels of neutrophils, macrophages, B cells, and CD8<sup>+</sup>T cell across TCGA-LUAD cohort are shown by scatterplots in Fig. 7. *KIF11* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.18$ ,  $p = 5.85 \times 10^{-5}$ ), neutrophils ( $r = 0.231$ ,  $p = 2.03 \times 10^{-7}$ ), and macrophages ( $r = 0.154$ ,  $p = 6.09 \times 10^{-4}$ ). *KIF15* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.21$ ,  $p = 2.69 \times 10^{-6}$ ), neutrophils ( $r = 0.277$ ,  $p = 3.88 \times 10^{-10}$ ), and macrophages ( $r = 0.155$ ,  $p = 5.39 \times 10^{-4}$ ). *KIF18B* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.135$ ,  $p = 2.61 \times 10^{-3}$ ), neutrophils ( $r = 0.214$ ,  $p = 1.65 \times 10^{-6}$ ), and macrophages ( $r = 0.106$ ,  $p = 1.91 \times 10^{-2}$ ). *KIF20A* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.115$ ,  $p = 1.06 \times 10^{-2}$ ), neutrophils ( $r = 0.229$ ,  $p = 2.61 \times 10^{-7}$ ), and macrophages ( $r = 0.108$ ,  $p = 1.69 \times 10^{-2}$ ). *KIF2C* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.161$ ,  $p = 3.19 \times 10^{-4}$ ), neutrophils ( $r = 0.207$ ,  $p = 3.64 \times 10^{-6}$ ), and macrophages ( $r = 0.146$ ,  $p = 1.15 \times 10^{-3}$ ). *KIF4A* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.192$ ,  $p = 1.71 \times 10^{-5}$ ), neutrophils ( $r = 0.262$ ,  $p = 3.38 \times 10^{-9}$ ), and macrophages ( $r = 0.195$ ,  $p = 1.31 \times 10^{-5}$ ). *KIFC1* displayed significant positive correlations with infiltrating levels of CD8<sup>+</sup>T cell ( $r = 0.137$ ,  $p = 2.28 \times 10^{-3}$ ), neutrophils ( $r = 0.193$ ,  $p = 1.60 \times 10^{-5}$ ), and macrophages ( $r = 0.128$ ,  $p = 4.27 \times 10^{-3}$ ). *KIF11* ( $r = -0.24$ ,  $p = 6.58 \times 10^{-8}$ ), *KIF15* ( $r = -0.188$ ,  $p = 2.65 \times 10^{-5}$ ), *KIF18B* ( $r = -0.17$ ,  $p = 1.55 \times 10^{-4}$ ), *KIF20A* ( $r = -0.218$ ,  $p = 1.03 \times 10^{-6}$ ), *KIF2C* ( $r = -0.221$ ,  $p = 6.92 \times 10^{-7}$ ), *KIF4A* ( $r = -0.22$ ,  $p = 7.76 \times 10^{-7}$ ), *KIFC1* ( $r = -0.164$ ,  $p = 2.58 \times 10^{-4}$ ) showed significant negative correlations with infiltrating levels of B cells. In addition, *KIF11* ( $r = 0.028$ ,  $p = 5.36 \times 10^{-1}$ ), *KIF15* ( $r = 0.016$ ,  $p = 7.21 \times 10^{-1}$ ), *KIF18B* ( $r = 0.002$ ,  $p = 9.58 \times 10^{-1}$ ), *KIF20A* ( $r = 0.019$ ,  $p = 6.80 \times 10^{-1}$ ), *KIF2C* ( $r = 0.007$ ,  $p = 8.77 \times 10^{-1}$ ), *KIF4A* ( $r = 0.01$ ,  $p = 8.24 \times 10^{-1}$ ), *KIFC1* ( $r = 0.031$ ,  $p = 4.97 \times 10^{-1}$ ) showed nonsignificant positive correlations with tumor purity across TCGA-LUAD cohort.

**Prognostic analysis based on single CpG methylation of selected kinesins in LUAD patients.** We obtained the heatmaps of DNA methylation of selected kinesins using MethSurv. Among which cg04344917 CpG island (CGI) of *KIF11*, cg09053247 CGI of *KIF15*, cg01838385 CGI of *KIF18B*, cg07632946 CGI of *KIF20A*, cg20487572 CGI of *KIF2C*, cg27286863 CGI of *KIF4A*, cg2390442 CGI of *KIFC1* showed the highest methylation levels (Fig. 8). Furthermore, we studied KM plots which revealed that cg24827036 CGI of *KIFC1* were significantly associated with survival of LUAD patients (Fig. 9). A total of 461 patients were split into higher and lower expression groups. Higher methylated expression of *KIFC1* worsened the OS of LUAD patients.



**Figure 6.** Barplots showing top 10 significantly enriched (A) pathways, (B) GO-BP, (C) GO-MF, (D) GO-CC terms with respect to  $p$  values. The color of bars varies in accordance with  $p$  values with red signifying lowest  $p$  values and green signifying highest  $p$  values. Asterisk signs represent the terms are also significant according to FDR.

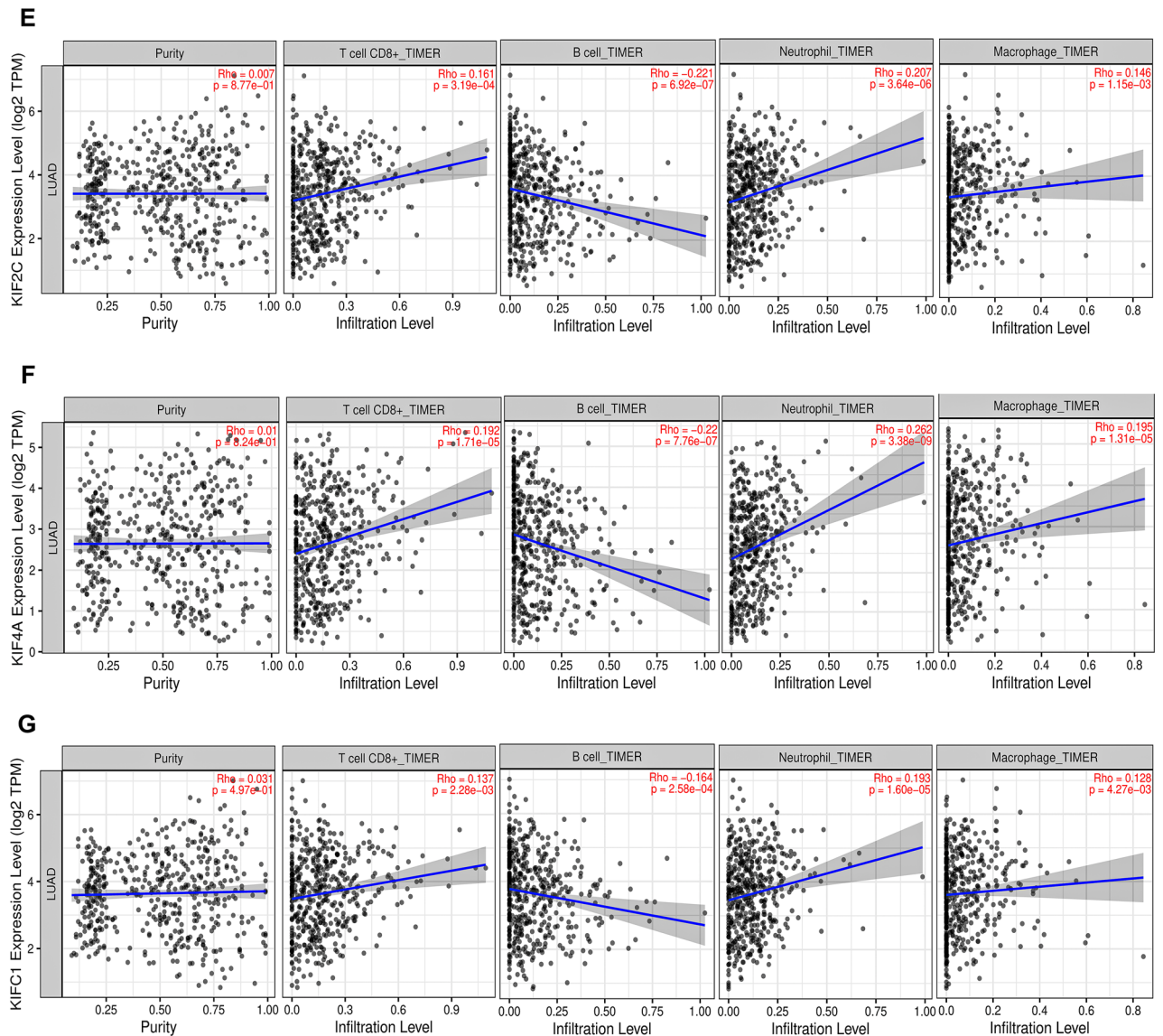




**Figure 7.** Scatterplots showing significant correlations of (A) KIF11, (B) KIF15, (C) KIF18B, (D) KIF20A, (E) KIF2C, (F) KIF4A, (G) KIFC1 with infiltrating levels of CD8<sup>+</sup> T cell, B cells, neutrophils, and macrophages across TCGA-LUAD cohort. Spearman's correlation value and estimated statistical significance were shown as the legends for each scatter plot.

### Discussion

LUAD's malignancy results in high morbidity and fatality rate<sup>35,36</sup>. Despite advances in surgery, radiation, and chemotherapy, which have improved tumor patients' clinical prognosis and survival<sup>37</sup>, LUAD is still hard to

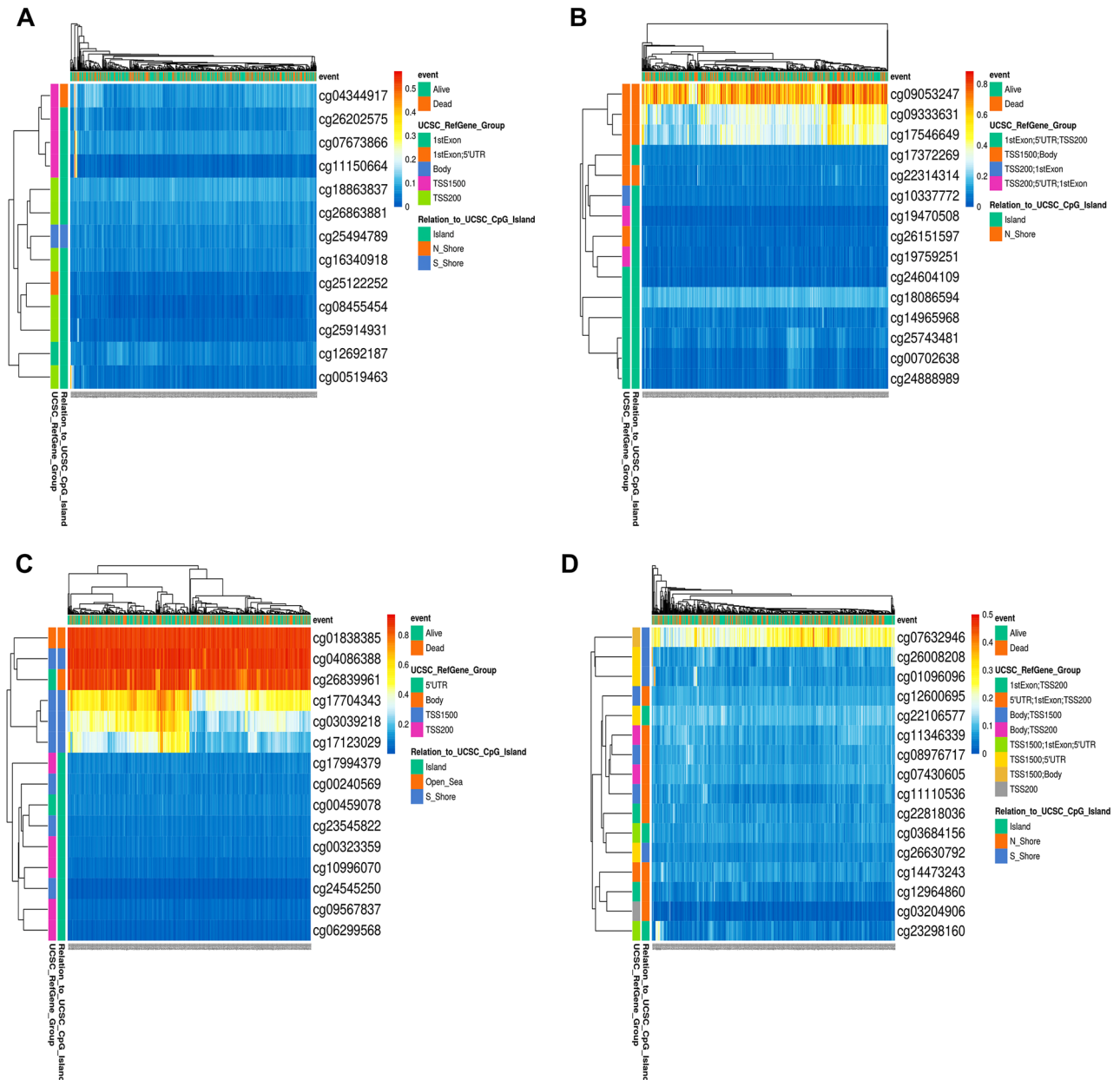


**Figure 7.** (continued)

treat because scientists don't fully understand the molecular mechanisms and basic signaling pathways in how LC works. It is expected that molecule-targeted therapy will be a revolutionary treatment technique for solid tumors, however, its efficacy and advantages remain restricted<sup>38,39</sup>. Because of chemoresistance and recurrence, the currently available therapeutic choices are limited. Therefore, a new and effective molecular target must be identified to cure LUAD.

Members of the KIF gene family are mostly found in eukaryotic cells, namely microtubules. Experiments conducted in vitro have shown that the transport of proteins occurs in only one direction, along the microtubule's negative pole and in the direction of the positive pole. Therefore, the genes that make up the KIF family are responsible for controlling the movement of mass proteins both inside of cells and outside of cells. This control encompasses a variety of functions, such as moving organelles and vesicles that contain material and taking part in the process of cell mitosis<sup>15,40,41</sup>. There have been reports that several genes in the kinesin family are linked to different kinds of cancer<sup>42-44</sup>. KIF family member genes have been demonstrated in various cancer types to establish their prognostic and diagnostic capacities. In our current study, we performed expression analysis of kinesin family in LUAD which revealed overexpression of *KIF11/12/15/23/18B/20A/2C/4A/C1* in tumor samples whereas *KIF17/26A/1C* were underexpressed in LUAD. Furthermore, we also studied mRNA expression based on cancer stage which showed overexpression of *KIF11/15/23/18B/20A/2C/4A/C1* in tumor tissues. Furthermore, we evaluated the prognostic value of selected kinesins in LUAD patients. Our results showed that an increased *KIF11/15/18B/20A/2C/4A/C1* expression is associated with poor OS in LUAD patients. So, by targeting these kinesins and decreasing their effect can be of therapeutic importance and patients' survival can be increased.

Our findings corroborate with multiple previous findings that showed overexpression of *KIF11*, *KIF15*, *KIF18B*, *KIF20A*, *KIF2C*, and *KIF4A* in LUAD tissues and when LUAD patients have high expression of these KIFs, their chances of survival are lower<sup>38,45-49</sup>. Next, we studied the genomic alterations of key kinesins which

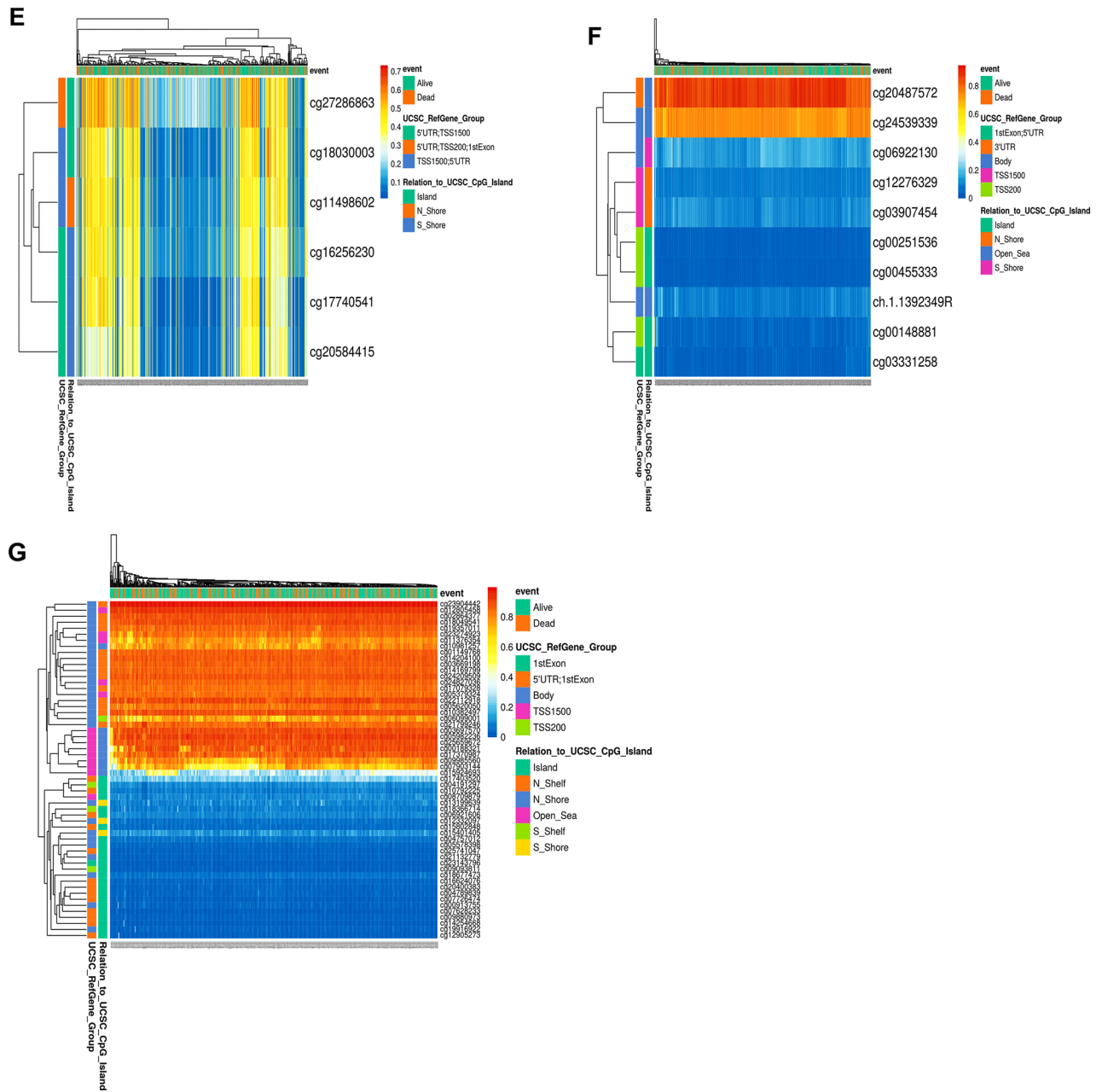


**Figure 8.** Heatmaps of CpG methylation levels of (A) KIF11, (B) KIF15, (C) KIF18B, (D) KIF20A, (E) KIF4A, (F) KIF2C, (G) KIFC1 across LUAD patients. Rows indicates the CpGs and columns indicates the patients. Methylation levels (1 = fully methylated; 0 = fully unmethylated) are shown as a continuous variable from red to blue color, high expression to low expression. Various colorful side boxes were used to represent the event, relation to UCSC\_CpG\_island and UCSC\_refGene\_Group.

showed the highest alteration in *KIFC1* (2.3%) as amplification being the most prominent type of alteration. Following that we constructed a PPIN of key kinesins and top 50 frequently altered genes and performed enrichment analysis. Our results showed high enrichment of kinesins in cell cycle and oocyte meiosis pathway, in biological processes named microtubule cytoskeleton organization involved in mitosis and mitotic spindle organization, microtubule binding and microtubule motor activity molecular functions and spindle and microtubule cytoskeleton cellular components.

For cells to divide and multiply, they go through a series of events known as the cell cycle, and abnormalities in the control of the genes involved in the cell cycle have been linked to the development of tumours. Mutations in upstream signal transduction pathways or genetic abnormalities within genes that encode cell cycle proteins cause cancer<sup>50</sup>. Our result showed high enrichment of kinesins in cell cycle processes hence kinesins are involved in controlling these processes somewhat and regulating LUAD.

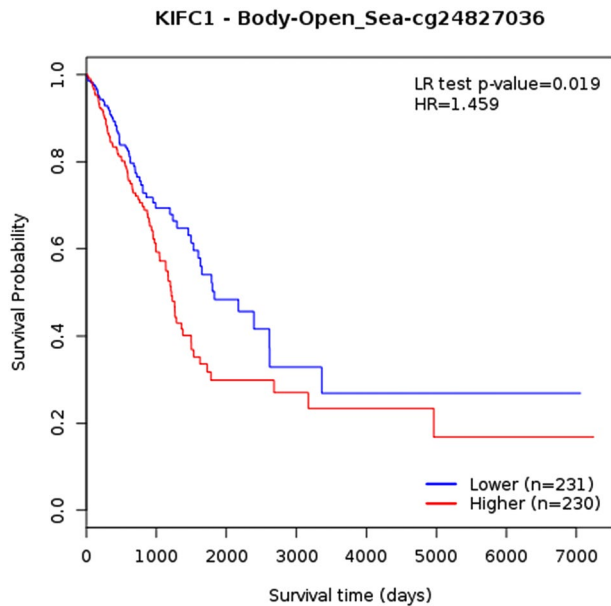
The other two typically active mechanisms in LC were oocyte meiosis and progesterone-mediated oocyte maturation. One cycle of DNA replication in meiosis is followed by two cycles of chromosomal segregation (Meiosis I and Meiosis II). Normally, oocytes are stopped during the G2 stage of meiosis I. Progesterone exposure releases



**Figure 8.** (continued)

them from this natural lock, allowing the two meiotic division cycles to resume and the oocyte to mature<sup>51,52</sup>. So, it makes sense that dysregulations in oocyte maturation and meiosis would impact the cell cycle process, and further cell cycle changes would impact normal bodily functions, increasing the likelihood that one would develop cancer. So, cell cycle, progesterone-mediated oocyte maturation, and oocyte meiosis play prominent roles in the progression of LUAD<sup>53</sup> and these two processes come under top 10 in the pathway enrichment analysis we did in our study showing the importance of kinesins in controlling these pathways in LUAD. Microtubules are  $\alpha$ - and  $\beta$ -tubulin heterodimers polymers. They exhibit highly dynamic behaviour, continuously engaging through processes of polymerization and de-polymerization, as well as lengthening and shortening. The fundamental components of the cytoskeleton are actin, intermediate filaments, and microtubules. They are crucial for various cell processes, including mitosis, the movement of vesicles and organelles inside cells, cell signaling, migration through cilia and flagella, cell shape and morphology<sup>54</sup>. So it can be said that any alteration in the function of kinesin from normal can impact these important pathways involved in LUAD progression.

We also performed tumor immune infiltration analysis on our key kinesins as tumor-infiltrating immune cells are critical components of the tumour microenvironment (TME), influencing tumor growth and survival depending upon their type and interaction. LC clearly displays an invasion of a wide variety of immune cell types comprising neutrophils, natural killer (NK) cells, macrophages, dendritic cells, T cells & B cells<sup>55</sup>. These cells perform multiple purposes and combine or oppose one another, producing the LC TME. Neutrophils make up



**Figure 9.** KM plot showing single CpG methylation of *KIFC1* across LUAD patients. It's location relative to CpG island, gene sub-region, CpG ID, and gene ID are also shown. Red and blue colors signify higher and lower expression groups.

50–70% of all white blood cells in the bloodstream and serve as the body's initial defence against infections. Neutrophils have been found to enhance tumor growth through various clinically relevant mechanisms. Tumor growth, angiogenesis, tumor cell migration, and metastasis are all facilitated by neutrophils. Still, a subclass of TANs known as N1 can have anticancer properties<sup>56</sup>. Our data revealed the highest correlation of our key kinesins with infiltration abundances of neutrophils in LUAD patients. CD8<sup>+</sup>T cell, also known as cytotoxic T lymphocytes, play a crucial role in mounting an efficient antitumor response. These cells can identify specific tumor-associated antigens (TAA) that are presented on major histocompatibility complex (MHC) class I molecules on the surface of cancer cells. Furthermore, they possess the ability to eliminate cancer cells directly<sup>57</sup>. Our study found that infiltrations of CD8<sup>+</sup>T cell and neutrophil correlated most with *KIF15* expression levels. Also, B cell and macrophage infiltrations correlated most with *KIF11* and *KIF4A* expression levels.

Cancer is often caused by the inactivation of several tumor-suppressor genes through point mutations and deletion of chromosomes<sup>58</sup>. Recent research has shown that epigenetic changes are key to cancer development. Many genes have CGIs in their promoter regions, and abnormal methylation of these sites in cancer leads to transcriptional suppression. Epigenetic alterations are passed down through cell division, resulting in gene activity change but no changes in the sequence of DNA<sup>59,60</sup>. Changes in DNA methylation patterns are a key feature of many types of cancer, including LC.

So further we conducted a single CpG methylation-based prognostic analysis on key kinesins which showed that CpG methylation in *KIFC1* was associated with poor prognosis in LUAD patients. *KIFC1* is believed to be an oncogene in various types of cancers as it plays a crucial role in clustering multiple centrosomes to sustain tumor survival<sup>61,62</sup>.

## Conclusions

Our research revealed a significant function for the kinesin family in initiating and progressing LUAD. *KIF11/15/18B/20A/2C/4A/C1* mRNA expression levels were significantly upregulated and correlated with poor OS across LUAD patients. They were highly associated with cell cycle. Our results revealed the highest genomic alteration in *KIFC1* with highest number of CpG methylation. cg24827036 CGI of *KIFC1* was associated with poor prognosis across LUAD. We concluded that *KIFC1* can be a great individual prognostic biomarker, and that inhibiting its expression could be a potential therapeutic approach. Additionally, CpG island cg24827036 can serve as a great prognostic biomarker and treatment site.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 11 January 2023; Accepted: 11 June 2023

Published online: 17 June 2023

## References

- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**(3), 209–249 (2021).
- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018).
- Zhang, L. *et al.* Integrated single-cell RNA sequencing analysis reveals distinct cellular and transcriptional modules associated with survival in lung cancer. *Signal Transduct. Target. Ther.* **7**(1), 9 (2022).
- Devesa, S. S. *et al.* International lung cancer trends by histologic type: Male:female differences diminishing and adenocarcinoma rates rising. *Int. J. Cancer* **117**(2), 294–299 (2005).
- Chang, J. T., Lee, Y. M. & Huang, R. S. The impact of the Cancer Genome Atlas on lung cancer. *Transl. Res.* **166**(6), 568–585 (2015).
- Detterbeck, F. C., Boffa, D. J. & Tanoue, L. T. The new lung cancer staging system. *Chest* **136**(1), 260–271 (2009).
- Subramanian, J. & Govindan, R. Lung cancer in never smokers: A review. *J. Clin. Oncol.* **25**(5), 561–570 (2007).
- Couraud, S. *et al.* Lung cancer in never smokers: A review. *Eur. J. Cancer* **48**(9), 1299–1311 (2012).
- Molina, J. R. *et al.* Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* **83**(5), 584–594 (2008).
- Kenfield, S. A. *et al.* Comparison of aspects of smoking among the four histological types of lung cancer. *Tob. Control* **17**(3), 198–204 (2008).
- Vale, R. D., Reese, T. S. & Sheetz, M. P. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* **42**(1), 39–50 (1985).
- Miki, H. *et al.* All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci. USA* **98**(13), 7004–7011 (2001).
- DeBoer, S. R. *et al.* Conventional kinesin holoenzymes are composed of heavy and light chain homodimers. *Biochemistry* **47**(15), 4535–4543 (2008).
- Hirokawa, N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279**(5350), 519–526 (1998).
- Hirokawa, N. *et al.* Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.* **10**(10), 682–696 (2009).
- Sharp, D. J., Rogers, G. C. & Scholey, J. M. Microtubule motors in mitosis. *Nature* **407**(6800), 41–47 (2000).
- Yu, Y. & Feng, Y. M. The role of kinesin family proteins in tumorigenesis and progression: Potential biomarkers and molecular targets for cancer therapy. *Cancer* **116**(22), 5150–5160 (2010).
- Hirokawa, N. & Takemura, R. Kinesin superfamily proteins and their various functions and dynamics. *Exp. Cell Res.* **301**(1), 50–59 (2004).
- Zhu, C. *et al.* Functional analysis of human microtubule-based motor proteins, the kinesins and dyneins, in mitosis/cytokinesis using RNA interference. *Mol. Biol. Cell* **16**(7), 3187–3199 (2005).
- Tang, Z. *et al.* GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**(W1), W98–W102 (2017).
- Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**(3), 725–731 (2010).
- Lanczky, A. & Gyorffy, B. Web-based survival analysis tool tailored for medical research (KMplot): Development and implementation. *J. Med. Internet Res.* **23**(7), e27633 (2021).
- Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**(5), 401–404 (2012).
- Barrett, T. *et al.* NCBI GEO: Mining millions of expression profiles—database and tools. *Nucleic Acids Res* **33**(Database issue), D562–D563 (2005).
- Gupta, S., *et al.*, *Integrative Multiomics and Regulatory Network Analyses Uncovers the Role of OAS3, TRAFD1, miR-222-3p, and miR-125b-5p in Hepatitis E Virus Infection.* *Genes* (Basel), 2022. **14**(1).
- Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
- Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**(D1), D607–D613 (2019).
- Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504 (2003).
- Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**(W1), W90–W97 (2016).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**(11), 1947–1951 (2019).
- Kanehisa, M. *et al.* KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**(D1), D587–D592 (2023).
- Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**(W1), W509–W514 (2020).
- Modhukur, V. *et al.* MethSurv: A web tool to perform multivariable survival analysis using DNA methylation data. *Epigenomics* **10**(3), 277–288 (2018).
- Kleczyk, E. K. *et al.* Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front. Immunol.* **10**, 954 (2019).
- Sun, Y. *et al.* Combined evaluation of the expression status of CD155 and TIGIT plays an important role in the prognosis of LUAD (lung adenocarcinoma). *Int. Immunopharmacol.* **80**, 106198 (2020).
- Arora, S. *et al.* miR-495-3p regulates sphingolipid metabolic reprogramming to induce Sphk1/ceramide mediated mitophagy and apoptosis in NSCLC. *Free Radic. Biol. Med.* **189**, 71–84 (2022).
- Li, Z. *et al.* KIF11 serves as an independent prognostic factor and therapeutic target for patients with lung adenocarcinoma. *Front. Oncol.* **11**, 670218 (2021).
- Kiyuna, L. A. *et al.* Targeting mitochondrial dysfunction and oxidative stress in heart failure: Challenges and opportunities. *Free Radic. Biol. Med.* **129**, 155–168 (2018).
- Lu, W. & Gelfand, V. I. Moonlighting motors: Kinesin, dynein, and cell polarity. *Trends Cell Biol.* **27**(7), 505–514 (2017).
- Rath, O. & Kozielski, F. Kinesins and cancer. *Nat. Rev. Cancer* **12**(8), 527–539 (2012).
- Sheng, N. *et al.* Overexpression of KIF2A is suppressed by miR-206 and associated with poor prognosis in ovarian cancer. *Cell Physiol. Biochem.* **50**(3), 810–822 (2018).
- Zhang, Y. *et al.* Overexpression of a novel candidate oncogene KIF14 correlates with tumor progression and poor prognosis in prostate cancer. *Oncotarget* **8**(28), 45459–45469 (2017).
- Xu, H. *et al.* Silencing of KIF14 interferes with cell cycle progression and cytokinesis by blocking the p27(Kip1) ubiquitination pathway in hepatocellular carcinoma. *Exp. Mol. Med.* **46**(5), e97 (2014).
- Qiao, Y. *et al.* Increased KIF15 expression predicts a poor prognosis in patients with lung adenocarcinoma. *Cell Physiol. Biochem.* **51**(1), 1–10 (2018).
- Zhong, Y. *et al.* Clinical significance and integrative analysis of kinesin family member 18B in lung adenocarcinoma. *OncoTargets Ther.* **12**, 9249–9264 (2019).

47. Zhao, X. *et al.* Overexpression of KIF20A confers malignant phenotype of lung adenocarcinoma by promoting cell proliferation and inhibiting apoptosis. *Cancer Med.* **7**(9), 4678–4689 (2018).
48. Song, Y., W. Tang, and H. Li, *Identification of KIF4A and its effect on the progression of lung adenocarcinoma based on the bioinformatics analysis.* *Biosci Rep.* 2021. **41**(1).
49. Bai, Y. *et al.* Co-expression network analysis identified KIF2C in association with progression and prognosis in lung adenocarcinoma. *Cancer Biomark.* **24**(3), 371–382 (2019).
50. Otto, T. & Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **17**(2), 93–115 (2017).
51. Mahrous, E., Yang, Q. & Clarke, H. J. Regulation of mitochondrial DNA accumulation during oocyte growth and meiotic maturation in the mouse. *Reproduction* **144**(2), 177–185 (2012).
52. Shao, H. *et al.* Xenopus oocyte meiosis lacks spindle assembly checkpoint control. *J. Cell Biol.* **201**(2), 191–200 (2013).
53. Yu, B. *et al.* Identification of activated pathways in lung adenocarcinoma based on network strategy. *J. Cancer Res. Ther.* **16**(4), 793–799 (2020).
54. Tagliamento, M. *et al.* Microtubule-targeting agents in the treatment of non-small cell lung cancer: Insights on new combination strategies and investigational compounds. *Expert Opin. Investig. Drugs* **28**(6), 513–523 (2019).
55. Zheng, X., Hu, Y. & Yao, C. The paradoxical role of tumor-infiltrating immune cells in lung cancer. *Intractable Rare Dis. Res.* **6**(4), 234–241 (2017).
56. Uribe-Querol, E. & Rosales, C. Neutrophils in cancer: Two sides of the same coin. *J. Immunol. Res.* **2015**, 983698 (2015).
57. Hiraoka, K. *et al.* Concurrent infiltration by CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells is a favourable prognostic factor in non-small-cell lung carcinoma. *Br. J. Cancer* **94**(2), 275–280 (2006).
58. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**(2), 159–170 (1996).
59. Laird, P. W. Oncogenic mechanisms mediated by DNA methylation. *Mol. Med. Today* **3**(5), 223–229 (1997).
60. Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nat. Genet.* **21**(2), 163–167 (1999).
61. Xiao, K. H. *et al.* Kinesin family member C1 accelerates bladder cancer cell proliferation and induces epithelial–mesenchymal transition via Akt/GSK3beta signaling. *Cancer Sci.* **110**(9), 2822–2833 (2019).
62. Li, Q. *et al.* Kinesin family member 15 promotes cancer stem cell phenotype and malignancy via reactive oxygen species imbalance in hepatocellular carcinoma. *Cancer Lett.* **482**, 112–125 (2020).

## Acknowledgements

The authors would like to thank Jamia Millia Islamia for providing infrastructure, journal access, and internet facilities. Gulnaz Tabassum is highly thankful to the Ministry of Human Resource and Development (MHRD), Government of India for providing Prime Minister’s Research Fellowship (PMRF ID: 3302519). Prithvi Singh would like to thank the ICMR for awarding him Senior Research Fellowship [Grant Number: BMI/11(89)/2020].

## Author contributions

Conceptualization: P.S., K.D., R.D.; Methodology: G.T., P.S., R.G.; Data Curation: G.T., P.S., R.G., M.A.H., N.A., A.A.A., L.S.A., M.R.H., S.R.; Software: P.S. and R.G.; Formal Analysis: P.S. and R.G.; Writing—Original Draft: G.T., P.S. and R.G.; Writing—Review and Editing: G.T., P.S., A.K., M.A.H., N.A., A.A.A., L.S.A., M.R.H., S.R.; Supervision: K.D. and R.D.; Project Administration: K.D. and R.D. All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-36842-6>.

**Correspondence** and requests for materials should be addressed to K.D. or R.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023