Contents lists available at ScienceDirect

Heliyon



journal homepage: www.cell.com/heliyon

Research article

5²CelPress

A novel prognostic signature of coagulation-related genes leveraged by machine learning algorithms for lung squamous cell carcinoma

Guo-Sheng Li^a, Rong-Quan He^b, Zhi-Guang Huang^c, Hong Huang^d, Zhen Yang^d, Jun Liu^a, Zong-Wang Fu^a, Wan-Ying Huang^c, Hua-Fu Zhou^a, Jin-Liang Kong^d, Gang Chen^{c,*}

^a Department of Cardiothoracic Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, PR China

^b Department of Medical Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, PR China

^c Department of Pathology, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, PR China ^d Division of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang

Autonomous Region, PR China

ARTICLE INFO

Keywords: Cancer Prognosis Immunotherapy mRNA Protein

ABSTRACT

Coagulation-related genes (CRGs) have been demonstrated to be essential for the development of certain tumors; however, little is known about CRGs in lung squamous cell carcinoma (LUSC). In this study, we adopted CRGs to construct a coagulation-related gene prognostic signature (CRGPS) using machine learning algorithms. Using a set of 92 machine learning integrated algorithms, the CRGPS was determined to be the optimal prognostic signature (median C-index = 0.600) for predicting the prognosis of an LUSC patient. The CRGPS was not only superior to traditional clinical parameters (e.g., T stage, age, and gender) and its commutative genes but also outperformed 19 preexisting prognostic signatures for LUSC on predictive accuracy. The CRGPS score was positively correlated with poor prognoses in patients with LUSC (hazard ratio > 1, p < 10.05), indicating its suitability as a prognostic marker for this disease. The CRGPS was observed to be inversely correlated with the degree of infiltration of natural killer cells. For some tumors, patients with lower CRGPS scores are more likely to experience enhanced immunotherapy effects (area under the curve = 0.70), which implies that the CRGPS can potentially predict immunotherapy efficacy. A high CRGPS score is predictive of an LUSC patient being sensitive to several drugs. Collectively, these findings indicate that the CRGPS may be a reliable indicator of the prognoses of patients with LUSC and may be useful for the clinical management of such patients.

1. Introduction

Globally, lung cancer is the second most common type of cancer-after breast cancer-and the most fatal of all cancer types, with

https://doi.org/10.1016/j.heliyon.2024.e27595

Received 5 April 2023; Received in revised form 26 December 2023; Accepted 4 March 2024

Available online 5 March 2024

^{*} Corresponding author. Department of Pathology, The First Affiliated Hospital of Guangxi Medical University, Nanning, 530021, Guangxi Zhuang Autonomous Region, PR China.

E-mail address: chengang@gxmu.edu.cn (G. Chen).

^{2405-8440/© 2024} The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

an estimated 2.1 million new cases and 1.8 million deaths in 2018 [1,2]. Lung squamous cell carcinoma (LUSC) is one of the most common subtypes of lung cancer [3–5], the incidence of which is second only to lung adenocarcinoma [6]. Platinum-based doublet therapy and combination chemotherapy have been the mainstay of treatment for advanced LUSC; however, due to the aggressive nature of LUSC and its late diagnosis, and the therapeutic inefficacy of the current treatment, clinical management of LUSC presents a great challenge, with a five-year overall survival rate of only 18% for LUSC patients [7–11]. This necessitates further research into developing more effective, durable, and potentially personalized clinical management strategies for LUSC patients.

Coagulation abnormality is one of the most common complications of tumors, including LUSC tumors, and it can manifest in many forms—one typical form is venous thromboembolism [12]. The pathogenesis of tumor-associated thrombosis is complex. This may be the result of bleeding and intravascular coagulation due to rupture of a vascular wall or extravascular coagulation due to increased vascular permeability and plasma extravasation. These complex processes may be further facilitated [13] by the entry of metastatic cancer cells into the circulatory system, inducing the recruitment or activation of inflammatory cells such as polymorphonuclear cells and macrophages [14]. Although the pathogenesis of tumor-associated thrombosis is not well understood, the essential roles of a series of coagulation-related genes (CRGs) in the occurrence and development of various cancers (such as stomach cancer [15], breast cancer [16], and lung cancer [17]) have been determined. Dysfunction of the coagulation system is also linked to prognosis in several cancers [12,18,19]. A prior study comprehensively explores the critical roles of CRGs in hepatocellular carcinoma and constructed a prognostic signature for survival prediction and treatment guidance of this disease [20]. However, little is known about CRGs in LUSC, which



Fig. 1. Illustrative flow chart of this study. CRGs, coagulation-related genes; PRGs, prognosis-related genes; DEGs, differential expression genes; CRGPS, coagulation-related gene prognostic signature.

G.-S. Li et al.

needs to be further explored.

Multigene panel construction is a novel way to identify an ideal biomarker for risk stratification and risk reference in cancer treatment [21]. Using the major online databases and employing the fundamentals of machine learning, we constructed a coagulation-related gene prognostic signature (CRGPS) to provide highly comprehensive and persuasive indicators of prognoses and immune therapy effects on LUSC patients.

2. Materials and methods

This study attempts to construct a CRGPS based on dozens of signatures using integration algorithms. The overall design is illustrated in the flow chart in Fig. 1. This study was approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University (Reference No: 2021[KY-E-246]).

2.1. Sampling of LUSC-related public cohorts and sample collection

The cohorts included in this study were recruited from multiple databases, including ArrayExpress, Gene Expression Omnibus [22], The Cancer Genome Atlas (TCGA) [23], and the Genotype-Tissue Expression (GTEx) project [24]. The search strategy employed to find the cohorts involved setting the search term as "lung AND squamous cell AND (tumor OR cancer OR carcinoma OR neoplasms) AND (mRNA OR gene)." The inclusion criteria were as follows: (1) tissues are from LUSC patients, (2) complete messenger RNA (mRNA) expression data are provided, and (3) the sample size of merged datasets is not less than three for both the LUSC group and the control group. Duplicate samples in various datasets were excluded.

The Human Protein Atlas [25,26] was also consulted for the screening of anti-CRG antibody staining information and images, and LUSC tissue and normal lung tissue specimens in the database were used to compare the protein levels of constituent genes of the CRGPS in LUSC sample tissues vis-à-vis normal lung tissues. The protein expression profile data of two LUSC cohorts, PDC000219 [27] and PDC000234 [28], were downloaded from the Proteomic Data Commons (https://pdc.cancer.gov), and were also used to compare the protein levels of commutative genes of the CRGPS in LUSC tissues vis-à-vis control tissues.

The Cancer Single-cell Expression Map (CSEP) [29] can be utilized to analyze and visualize single-cell RNA-sequencing data on human cancers. To investigate the expression distribution of the constituent genes of the CRGPS within various cells, seven LUSC-related single-cell RNA-sequencing data samples were analyzed [30] using the CSEP—the entire process was performed in the CSEP. The clustering visualization method employed for this analysis involved setting a uniform manifold approximation and projection for dimension reduction—detailed parameters can be found in the original study [29] from which the method was adopted.

2.2. Collection of in-house samples and experiments on in-house microarrays

An in-house microarray of three LUSC tissue samples and three samples from noncancerous tissue adjacent to the LUSC tumors was also performed to verify the mRNA expression levels of the CRGPS commutative genes in the LUSC group vis-à-vis the control group. These samples were obtained from the First Affiliated Hospital of Guangxi Medical University, China. The LUSC tissue samples and samples from noncancerous tissue adjacent to the LUSC tumors were obtained from three primary LUSC patients undergoing surgery without chemoradiotherapy. The three patients were all male and were 51, 56, and 58 years old (Supplementary Material 1).

The in-house samples were processed into microarrays. Five minutes after excision, both the LUSC samples and the samples from adjacent cancerous tissue were placed in liquid nitrogen and stored at -80 °C. TRIzol reagent (Invitrogen, USA) was used to extract total RNA from the tissues, following the manufacturer's instructions. An RNasey Mini Kit (Qiagen, p/n 74104, Germany) and an RNeasy mini spin column (Qiagen, Germany) were used to purify the RNA, followed by measurement of the OD230, OD260, and OD280 values using a Nanodrop ND-1000 spectrophotometer (Agilent, USA) to determine the RNA concentration. Sample labeling was completed using a Quick Amp Labeling Kit and a One-Color RNA Spike-in Kit (Agilent, p/n 5190-0442). Hybridization conditions were set for 17 h of hybridization at 65 °C. The microarray was scanned using an Agilent microarray scanner (G2565BA, Agilent, USA).

2.3. Processing of mRNA expression data and score criteria for protein levels

Raw microarray matrixes collated from the ArrayExpress and Gene Expression Omnibus databases were normalized using the *Oligo* [31] and *Limma* [32] software packages for R (version 4.1.0). RNA-sequencing data obtained from TCGA and the GTEx project were in transcripts per kilobase million (TPM) format. Both the microarray and RNA-sequencing data were $log_2(x+1)$ transformed. Datasets obtained from the same platforms (e.g., GPL570) were merged, and the batch effects between these datasets were removed using the *SVA* software package [33] for R. For instance, the GSE84776 and GSE70089 cohorts were detected with the GPL11154 platform, Illumina HiSeq 2000 (Homo sapiens); thus, they were merged into a new cohort called GPL11154.

Tissue samples from the Human Protein Atlas [25,26] had three features that reflected the protein expression levels: staining, intensity, and quality. The score criteria for the three features are as follows: (1) the natural numbers 0, 1, 2, and 3 denote *not detected, low staining, medium staining,* and *high staining,* respectively; (2) the natural numbers 0, 1, 2, and 3 denote *negative intensity, weak intensity, moderate intensity,* and *strong intensity,* respectively; (3) the integers 0, 1, 2, and 3 indicate none, <25%, 25–75%, and >75% positively stained cells, respectively. The ultimate score for the protein levels of each sample was the product of the staining score, intensity score, and quality score.

2.4. Collection of CRGs and selection of PRG, DEG, and CRGPS candidates

CRGs were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [34–37] and the Molecular Signatures Database (MSigDB) [38]. The prognosis-related genes (PRGs) of LUSC were defined using univariate Cox regression analysis, with a hazard ratio (HR) not equaling 1 in at least two of the seven public cohorts in this study (p < 0.05). Differential expression genes (DEGs) from the LUSC group and the control group were selected using standardized mean difference (SMD) values: genes with an SMD >0 or SMD <0 (the 95% confidence interval did not overlap 0) were considered DEGs. When a CRG was both a PRG and a DEG, it was designated as a candidate for establishing the CRGPS.





Fig. 2. Datasets used in this study and in the selection of candidate genes for the CRGPS. Panel A: Datasets and sample sizes used in this study. Panel B: Forest plot of 18 prognosis-related CRGs in LUSC; hazard ratios (HRs) of the 18 CRGs, calculated using univariate Cox regression analysis. Panel C: Forest plot of alpha-2-macroglobulin (*A2M*) expression in the LUSC group vis-à-vis the control group; differences in *A2M* expression between the two groups were compared using a standardized mean difference (SMD) value. Panel D: Selection process for candidate genes for the CRGPS.

2.5. Establishing the machine learning based CRGPS

To construct a CRGPS with a steady and accurate manifestation, 10 machine learning based measurement methods were utilized, including CoxBoost, elastic network (Enet), the least absolute shrinking and selection operator (LASSO), generalized boosted regression modeling (GBM), random forest for survival (RFS), Ridge, stepwise Cox (forward), stepwise Cox (both), partial least squares regression for Cox (plsRcox), and survival support vector machine (survival-SVM). Ninety-two combinations were formed using the algorithms mentioned earlier. The GSE157010 cohort had the most samples among the microarray cohorts with prognosis information; thus, GSE157010 was chosen for the tenfold cross-validation and fitting of prognostic signatures. The other five microarray cohorts (GSE19188, GSE29013, GSE30219, GSE37745, and GSE50081) and the RNA-sequence cohort TCGA were defined as test cohorts for prognostic signatures. The cohort with the highest median C-index of all six test cohorts represented the best performance for predicting the prognosis of patients and was designated as the CRGPS.

2.6. Internal evaluation and external comparisons with the CRGPS

The C-index and the area under the curve (AUC) of the receiver operating characteristic (ROC) curves were both used to evaluate the accuracy of the CRGPS, certain CRGs, and the clinical parameters used in predicting the prognosis of LUSC patients. AUC was calculated using the *timeROC* [39] package for R. Gene prognosis signatures constructed by 19 prior LUSC-related studies (Supplementary material 2) were collected and utilized to compare the accuracy of these signatures and the CRGPS using the C-index and AUC.

2.7. Correlations between the CRGPS and immune infiltration levels, response to immunotherapy, and potential drugs for certain LUSC patients

The CIBERSORT [40] and MCP-counter [41] algorithms were used to calculate the immune cell filtration levels of the cohorts with prognosis information. Four immunotherapy cohorts—GSE103668 (breast cancer), GSE78220 (melanomas), GSE35640 (NSCLC), and GSE91061 (melanoma)—were collected and used to measure the ability of the CRGPS risk scores to predict the treatment response of cisplatin and bevacizumab, pembrolizumab, MAGE-A3, and nivolumab. The summary ROC curve was used to detect the accuracy of the CRGPS in predicting patient response to immunotherapy.

Genomics of Drug Sensitivity in Cancer (GDSC) [42,43] provides a large amount of experiment-based data on tumor cell drug sensitivity (shown through half maximal inhibitory concentration [IC50]) and tumor treatment genomics. The *oncoPredict* [44] package includes the drug IC50 data and the gene sequence results of the GDSC cell lines. Patient response to specific drugs can be predicted based on gene expression data using the *oncoPredict* package for the calculation. The lower the IC50 value, the higher the sensitivity of the patient to a specific drug. Using the *oncoPredict* package for R, drugs that may benefit LUSC patients with high CRGPS risk scores were predicted based on drug therapy data from GDSC. Gene expression data on the GSE157010 cohort were used to calculate the differences in drug responsiveness between patients with high and low CRGPS risk scores.

2.8. Statistical analysis

Except for the summary ROC curve, which was produced in Stata (version 15.0), the rest of the processing and visualization of the data in this study were performed in R (version 4.1.0). Wilcoxon rank-sum tests were performed on two comparison groups: LUSC group vis-à-vis control group and high-risk group vis-à-vis low-risk group. Spearman correlation analysis was used to measure the correlation between the two continuous variables. Kaplan–Meier analysis was performed and visualized using the *survival* package in R. The difference in survival rates was compared using log-rank testing. A *p*-value of <0.05 indicated statistical significance.

3. Results

3.1. Identifying CRGs, PRGs, DEGs, and CRGPS candidates

In this study, 3069 samples were used to select the CRGPS candidates. The sample counts for the 40 cohorts included in this research are presented in Fig. 2A. Seven of these cohorts (GSE157010, GSE19188, GSE29013, GSE30219, GSE37745, GSE50081, and TCGA) contained prognosis information, which was used for the construction, evaluation, and comparison of the CRGPS. An in-house microarray was performed on three LUSC tissue samples and three normal lung tissue samples (Supplementary Material 1).

To construct the CRGPS, a series of CRG, PRG, DEG, and CRGPS candidates were identified successively. A total of 683 CRGs were collected from the KEGG and MSigDB databases. The CRGs in the KEGG signaling pathways *hsa04610* (complement and coagulation cascades) and *hsa04611* (platelet activation) were collected and supplemented with those from the MSigDB gene sets: *GOBP COAG-ULATION, HALLMARK COAGULATION, HP ABNORMALITY OF COAGULATION,* and *WP COMPLEMENT AND COAGULATION CASCADEG*.

Univariate Cox regression analyses revealed that 348 of 683 CRGs are closely linked with the prognosis of LUSC patients (p < 0.05). Therefore, these 348 CRGs were considered PRGs; some PRG examples are presented in Fig. 2B. Furthermore, 146 of the 348 PRGs were identified as DEGs in LUSC based on their SMD values. For instance, the SMD of alpha-2-macroglobulin (*A2M*) was -2.42 (95% CI: -2.96 to -1.88, Fig. 2C), indicating a declining trend in *A2M* expression in LUSC. Finally, 144 of the 146 DEGs included in the seven cohorts (e.g., GSE157010) were selected as candidates for constructing the CRGPS (Fig. 2D).

0.66	0.63	0.63	0.57	0.52	0.63	0.54	Lasso + StepCox [forward]
0.66	0.63	0.63	0.57	0.52	0.63	0.54	Lasso + plsRcox
0.67	0.59	0.58	0.58	0.52	0.59	0.54	StepCox [forward] + Enet [a=0.9]
0.67	0.59	0.58	0.58	0.54	0.59	0.54	StepCox [forward] + Enet [a=0.8]
0.67	0.59	0.58	0.58	0.54	0.58	0.54	StepCox [forward] + Enet [α=0.7] StepCox [forward] + Enet [α=0.6]
0.67	0.59	0.58	0.58	0.53	0.58	0.54	StepCox [forward] + Enet [α =0.4]
0.67	0.58	0.58	0.58	0.54	0.57	0.54	StepCox [forward] + Enet [a=0.5]
0.67	0.59	0.58	0.58	0.54	0.57	0.54	StepCox [forward] + Enet [d=0.3] StepCox [forward] + Enet [d=0.2]
0.63	0.54	0.57	0.63	0.55	0.58	0.57	survival-SVM
0.63	0.54	0.57	0.63	0.55	0.58	0.57	StepCox [forward] + survival-SVM
0.63	0.50	0.57	0.63	0.52	0.58	0.53	RFS + survival-SVM
0.67	0.6	0.6	0.59	0.54	0.54	0.55	RFS + Ridge
0.67	0.58	0.56	0.58	0.53	0.6	0.55	StepCox [both] + Ridge
0.66	0.58	0.62	0.6	0.56	0.52	0.55	Ridge
0.65	0.56	0.62	0.58	0.52	0.57	0.54	Lasso
0.62	0.56	0.6	0.62	0.51	0.57	0.55	Enet [a=0]
0.66	0.6	0.52	0.58	0.53	0.6	0.55	StepCox [both] + plsRcox
0.66	0.6	0.51	0.58	0.52	0.61	0.55	StepCox [both]
0.65	0.56	0.58	0.61	0.52	0.51	0.56	StepCox [forward] + Ridge
0.67	0.59	0.52	0.58	0.53	0.61	0.55	StepCox [both] + Enet [α=0.9]
0.67	0.59	0.52	0.58	0.53	0.61	0.55	StepCox [both] + Enet [α =0.8] StepCox [both] + Enet [α =0.7]
0.67	0.59	0.52	0.58	0.53	0.61	0.55	StepCox [both] + Enet [α =0.6]
0.67	0.59	0.52	0.58	0.53	0.61	0.55	StepCox [both] + Enet [α=0.5]
0.73	0.49	0.56	0.58	0.56	0.62	0.54	StepCox [forward] + GBM StepCox [forward] + Enet [α=0.1]
0.64	0.59	0.65	0.6	0.52	0.52	0.54	StepCox [both] + survival-SVM
0.67	0.59	0.53	0.58	0.53	0.61	0.55	StepCox [both] + Lasso StepCox [both] + Enet [g=0.4]
0.67	0.59	0.52	0.58	0.53	0.61	0.55	StepCox [both] + Enet [a=0.3]
0.67	0.59	0.53	0.58	0.53	0.61	0.55	StepCox [both] + Enet [α=0.2] StepCox [both] + Enet [α=0.1]
0.67	0.59	0.54	0.58	0.53	0.61	0.55	StepCox [both] + CoxBoost
0.74	0.5	0.56	0.61	0.56	0.61	0.54	RFS + GBM GBM
0.65	0.55	0.6	0.59	0.53	0.53	0.55	RFS + Enet [α=0.1]
0.64	0.55	0.56	0.6	0.52	0.5	0.55	CoxBoost + Enet [α =0.9]
0.64	0.55	0.55	0.6	0.52	0.5	0.55	CoxBoost + Enet [α =0.5]
0.64	0.55	0.55	0.6	0.53	0.5	0.55	CoxBoost + Enet [α =0.4]
0.64	0.54	0.56	0.6	0.52	0.5	0.55	CoxBoost + Enet [α =0.8]
0.64	0.54	0.56	0.6	0.52	0.5	0.55	CoxBoost + Enet [α=0.7]
0.64	0.55	0.52	0.61	0.54	0.57	0.51	CoxBoost + Enet [a=0.2]
0.64	0.54	0.56	0.61	0.52	0.51	0.55	CoxBoost + Enet [a=0.1]
0.64	0.54	0.56	0.6	0.52	0.5	0.55	CoxBoost + StepCox [forward] CoxBoost + Ridge
0.64	0.54	0.56	0.6	0.52	0.5	0.55	CoxBoost + plsRcox
0.65	0.54	0.58	0.6	0.49	0.48	0.54	Lasso + survival-SVM CoxBoost + StepCox [both]
0.64	0.54	0.56	0.6	0.52	0.5	0.55	CoxBoost + Lasso
0.64	0.53	0.6	0.62	0.49	0.48	0.55	CoxBoost + survival-SVM CoxBoost + GBM
0.86	0.54	0.54	0.6	0.54	0.56	0.52	StepCox [forward] + RFS
0.64	0.52	0.6	0.59	0.53	0.49	0.55	Enet [a=0.1] StenCox [forward] + plsRcox
0.69	0.59	0.54	0.53	0.51	0.6	0.53	plsRcox
0.63	0.51	0.6	0.56	0.53	0.5	0.54	Enet [a=0.9]
0.63	0.53	0.58	0.58	0.53	0.5	0.54	Enet [α =0.3]
0.63	0.52	0.59	0.58	0.53	0.5	0.54	Enet [a=0.2]
0.69	0.59	0.54	0.53	0.49	0.6	0.53	StepCox [forward]
0.69	0.59	0.54	0.53	0.49	0.6	0.53	RFS + StepCox[forward]
0.63	0.51	0.6	0.56	0.53	0.5	0.54	Enet [α =0.6]
0.63	0.51	0.6	0.56	0.53	0.51	0.54	Enet [a=0.5]
0.69	0.59	0.54	0.53	0.53	0.55	0.53	RFS + plsRcox
0.63	0.51	0.6	0.57	0.52	0.51	0.54	RFS + Enet [α=0.2]
0.63	0.51	0.6	0.57	0.53	0.51	0.54	RFS + Lasso
0.62	0.5	0.61	0.56	0.52	0.49	0.54	RFS + Enet [α=0.4]
0.62	0.51	0.64	0.56	0.52	0.49	0.54	RFS + Enet [α=0.3]
0.6	0.51	0.64	0.56	0.49	0.49	0.54	RFS + Enet [α=0.8]
0.61	0.51	0.64	0.56	0.49	0.5	0.54	StepCox [forward] + Lasso RFS + Enet [α=0.7]
0.6	0.51	0.64	0.56	0.49	0.49	0.54	RFS + Enet [α=0.6]
0.6	0.51	0.64	0.56	0.49	0.49	0.54	κrs + Enet [α=0.5] Lasso + RFS
0.6	0.51	0.64	0.56	0.49	0.49	0.54	StepCox [forward] + CoxBoost
0.6	0.51	0.64	0.56	0.49	0.49	0.54	RFS + CoxBoost Lasso + GBM



Cohort

Cindex 0.9 0.8 0.7 0.6 0.5

Fig. 3. Construction and composition of the CRGPS. C-indices of 92 machine learning algorithms in seven cohorts.



Fig. 4. Protein levels of genes of the CRGPS LUSC samples vis-à-vis normal lung tissue samples. Panel A: Protein level expression of the CRGPS component genes under the microscope; the four values in each panel (e.g., *1470, not detected, negative,* and *none*) represent the patient identity document, staining, intensity, and quantity—these images were obtained from the Human Protein Atlas (version 22.0). Panel B: Levels of proteins encoded by the CRGPS in LUSC tissues vis-à-vis control tissues, based on the Human Protein Atlas data. Panels C–D: Levels of proteins encoded by the CRGPS in LUSC vis-à-vis control tissues for the (C) PDC000219 and (D) PDC000234 cohort data from the Proteomic Data Commons. *P*-values are based on the Wilcoxon rank-sum test; NSp > 0.05, ***p < 0.001.

3.2. Establishing the CRGPS using integrated algorithms

Tenfold cross-validation and 92 integrated algorithms were applied to the GSE157010 training cohort. Six independent test cohorts were also utilized to measure the prediction ability of the CRGPS using the C-index. As illustrated in Fig. 3, a large number of



Fig. 5. Performance evaluation of the CRGPS based on the cohorts included in this study. Panel A: Clinical features of the GSE157010 training cohort. Panel B: C-indices of the CRGPS and clinical features of LUSC patients; **p < 0.01, ***p < 0.001. Panel C: C-indices of the CRGPS and its constitutive genes. Panel D: Area under the curve (AUC) for the CRGPS in seven cohorts. Panel E: One-, three-, and five-year AUC of the CRGPS and clinical features in the GSE157010 cohort. Panels F–H: (F) One-year, (G) three-year, and (H) five-year AUC of the CRGPS and its constitutive genes for the seven cohorts.

G.-S. Li et al.

prognostic signatures exhibited high accuracy in predicting prognoses with the GSE157010 training cohort (e.g., the *StepCox[forward]* + *RFS* signature). When evaluated using the six test cohorts, several integrated algorithms performed well at predicting the prognosis of LUSC patients (e.g., the *lasso* + *plsRcox* signature), while the *lasso* + *StepCox[forward]* signature demonstrated the highest median C-index (0.600) across the six test cohorts and was thus designated as the CRGPS.



Fig. 6. Performance comparison of the CRGPS and published signatures. Panel A: C-indices of the CRGPS and 19 other signatures based on the GSE157010 training cohort; p < 0.05, p < 0.01, p < 0.001, p < 0.001. Panel B: AUC of the CRGPS and 19 other signatures based on the GSE157010 training cohort. Panel C: Median C-indices of the CRGPS and 19 other signatures based on the seven cohorts included in this study. Panel D: Median AUC of the CRGPS and 19 other signatures based on the seven cohorts included in this study.

3.3. Overview of the commutative CRGs of the CRGPS

The CRGPS is composed of nine genes: adenylate cyclase 7 (*ADCY7*), diacylglycerol kinase alpha (*DGKA*), nuclear factor erythroid 2 (*NFE2*), platelet and endothelial cell adhesion molecule 1 (*PECAM1*), penta-EF-hand domain containing 1 (*PEF1*), serine protease 23 (*PRSS23*), prostaglandin I2 receptor (*PTGIR*), serpin family A member 1 (*SERPINA1*), and TIMP metallopeptidase inhibitor 3 (*TIMP3*). The coefficients of the nine commutative CRGs of the CRGPS are provided in Supplementary Material 3A. A differential expression analysis of the mRNA levels revealed that *DGKA* was more upregulated in LUSC tissues than in normal lung tissues; this was also verified by the results of the in-house microarray (Supplementary Material 3B). The remaining eight CRGs (e.g., *ADCY7*) were downregulated in LUSC (the 95% confidence intervals did not include 0, Supplementary Material 4), and a similar trend can be seen in the in-house microarray results for *NFE2*, *PECAM1*, and *SERPINA1* (Supplementary Material 3B).

Upon microscopic examination, it was observed that LUSC tissues had higher DGKA protein levels and lower NFE2 and SERPINA1 protein levels when compared to normal lung tissues (Fig. 4A), which is consistent with the expression of these genes at the mRNA level. A similar trend was also observed in the results of the Wilcoxon rank-sum tests (Fig. 4B), although no statistically significant differences were detected—which may be due to the limited sample size. Furthermore, based on more samples from the Proteomic Data Commons, ADCY7 and SERPINA1 were downregulated in LUSC tissues to a greater degree than in normal lung tissues (p < 0.05; Fig. 4C and D).



Fig. 7. Prognostic value of the CRGPS in LUSC. Panel A: Kaplan–Meier curves and log-rank tests of the CRGPS for the seven cohorts. Panel B: Cox regression analysis of the CRGPS and its constitutive genes for the GSE157010 cohort. Panel D: Multivariate Cox regression analysis of the CRGPS and its constitutive genes for the GSE157010 cohort; *p < 0.05, ***p < 0.001.



Fig. 8. Immune landscape analysis of the CRGPS and its potential immunotherapy value. Panels A–B: Immune cell infiltration distribution for the high-risk group vis-à-vis the low-risk group based on the (A) GSE19188 and (B) GSE29013 cohorts. Panel C: Different infiltration levels of several immune cells for the high-risk group vis-à-vis the low-risk group. Panel D: Correlations between the CRGPS and the infiltration levels of several immune cells. Panels E–F: Immune cell infiltration distribution for the high-risk group based on the (E) GSE19188

and (F) GSE29013 cohorts. Panel G: Differences in NK cell infiltration levels for the high-risk group vis-à-vis the low-risk group. Panel H: Correlations between the CRGPS and NK cell infiltration levels. Panel I: Predictive power of the CRGPS risk scores for treatment response to immunotherapy. Panel J: Treatment response to immunotherapy for the high-risk group vis-à-vis the low-risk group for four cohorts. *p < 0.05, ***p < 0.001.

Furthermore, this study investigates the expression levels of the nine constituent genes of the CRGPS using single-cell RNAsequencing data. These genes were found to be expressed not only in malignant LUSC cells but also in other cell types (Supplementary Material 5). Specifically, *ADCY7* was predominantly expressed in macrophages, natural killer (NK) cells, CD4⁺ central memory T cells (CD4⁺ Tcm), CD4⁺ effector memory T cells (CD4⁺ Tem), CD8⁺ central memory T cells (CD8⁺ Tcm), CD8⁺ effector memory T cells (CD8⁺ Tem), and CD8⁺ effector T cells. *DGKA* was predominantly expressed in CD4⁺ Tcm, CD4⁺ Tem, CD8⁺ Tcm, CD8⁺ effector T cells, fibroblasts, memory B cells, and NK cells. *NFE2* was predominantly expressed in monocytes and dendritic cells, while *PECAM1* was predominantly expressed in monocytes, plasma cells, and endothelial cells. *PEF1* was predominantly expressed in endothelial cells, fibroblasts, macrophages, NK cells, and endothelial cells. *PRSS23* was predominantly expressed in endothelial cells, fibroblasts, and NK cells. *PTGIR* was predominantly expressed in CD4⁺ Tcm, CD8⁺ naïve T cells, and endothelial cells. *SERPINA1* was predominantly expressed in macrophages and monocytes. Finally, *TIMP3* was expressed primarily in endothelial cells and fibroblasts (Supplementary Material 5).

3.4. CRGPS: internal assessment and external comparisons

To assess the stability of CRGPS, internal evaluations and external comparisons were performed using the C-index and AUC metrics. The alluvial diagram in Fig. 5A shows the proportion of five kinds of clinical features: age (>65 or \leq 65 years old), gender, T stages (T1–T3), and status (alive or dead). The C-indices for age, gender, T stages, and the nine genes of the CRGPS were compared with the CRGPS, and the results indicate that the CRGPS exhibits optimal performance (p < 0.05, Fig. 5B–C). The AUC values of the CRGPS were not less than 0.6 for most of the cohorts in terms of the one-, three-, and five-year survival rates of LUSC patients (Fig. 5D), indicating the strong predictive power of the CRGPS in predicting the prognoses of individuals with LUSC. The CRGPS also performed better at predicting the prognoses of LUSC patients (Fig. 5E) and the nine commutative CRGs of the CRGPS (Fig. 5F–H) for all the one-, three-, and five-year survival rates, which is consistent with the results of the evaluation based on C-indices.

Prior studies have mentioned multiple gene prognostic signatures for LUSC, and external comparisons between these signatures and the CRGPS are made in this study. In the GSE157010 training cohort, the C-index of the CRGPS was significantly higher than that of most signatures (except for one) from the 19 previous studies (p < 0.05, Fig. 6A). Furthermore, for the GSE157010 cohort, the one-, three-, and five-year AUC values of the CRGPS were all larger than the 19 signatures (Fig. 6B). To comprehensively verify these findings further, more comparisons of the CRGPS and the other 19 signatures were made based on the median of all six test cohorts. It can be seen in Fig. 6C that the CRGPS has the highest median C-index compared to the 19 previous studies. In the AUC measure, the CRGPS also had optimal AUC values compared to the prior 19 signatures for at least one of the one-, three-, and five-year survival rates (Fig. 6D).

3.5. Underlying prognosis value of the CRGPS

Using Kaplan–Meier curves, log-rank tests, and Cox regression analyses, the relationship between the CRGPS and the prognoses of LUSC patients was also explored. Individuals with LUSC in the high-risk group—based on the CRGPS risk scores—had a poor prognosis in six cohorts, i.e., GSE157010, GSE19188, GSE29013, GSE30219, GSE50081, and TCGA (p < 0.05, Fig. 7A). The CRGPS represents a prognosis risk factor for LUSC patients, as its HR of 1.37 is greater than one in the combined analysis of the seven cohorts (p < 0.05, Fig. 7B). Furthermore, based on the results of the univariate and multivariate Cox regression analyses (Fig. 7C–D), the CRGPS was found to be an independent risk factor for LUSC patients, similar to the T stage (HR > 1, p < 0.05). One of the test cohorts (i.e., the TCGA cohort), which has more clinical parameters than the training cohort, also supports the finding of the CRGPS as an independent risk factor for LUSC patients (p < 0.05, Supplementary Material 6).

3.6. Correlations between the CRGPS and immune infiltration levels

Immune cells play a crucial role in the antitumor mechanism. In this study, we used two algorithms to explore the correlation between the CRGPS and the infiltration levels of specific immune cells in LUSC for seven cohorts. For the CIBERSORT algorithm (examples for GSE19188 and GSE29013 can be seen in Fig. 8A–B), lower infiltration levels of naïve B cells, active NK cells, and follicular helper T cells were observed in the high-risk groups than in the low-risk groups, and such results were supported by no less than two cohorts (p < 0.05, Fig. 8C). Furthermore, there were negative correlations between the risk scores of the CRGPS and the infiltration levels of the naïve B cells, active NK cells, and follicular helper T cells (p < 0.05, Fig. 8D). Of these three immune cell types, NK cells are crucial immune cells that directly contribute to the death of cancer cells. Therefore, we attempted to verify the relationship between the CRGPS and NK cell infiltration levels. Using the MCP-counter algorithm (examples for GSE19188 and GSE29013 can be seen in Fig. 8E–F). A higher CRGPS risk score was detected with lower NK cell infiltration levels, and the CRGPS risk score was negatively associated with NK cell infiltration levels (p < 0.05, Fig. 8G–H).

3.7. Potential immunotherapy value of the CRGPS

Immunotherapy is one of the most essential and promising treatments for LUSC; thus, this study explores the potential value of CRGPS in predicting immunotherapy. Via the summary ROC curve in Fig. 8I, the CRGPS makes it feasible to discern whether patients will respond to immunotherapy (AUC = 0.70 [95% CI: 0.66-0.74]). Furthermore, it can be seen in Fig. 8J that more cancer patients who did not respond to immunotherapy were detected in the high-risk groups than in the low-risk groups. Thus, the CRGPS may be a reliable indicator for predicting patient response to immunotherapy.

3.8. Drugs that may potentially benefit LUSC patients with high CRGPS scores

Compared to LUSC patients with low CRGPS scores, those with high CRGPS scores had a more pessimistic prognosis; thus, it is meaningful to explore drugs that may benefit the latter. The drug response scores for each LUSC patient were calculated using the *oncoPredict* package. The results show that LUSC patients in the high-risk group were more sensitive to five drugs: Staurosporine, Dasatinib, Trametinib, ERK 2440, and BMS.754807 (p < 0.05, Fig. 9).

4. Discussion

Despite advances in medical care, the mortality rate of advanced-stage LUSC patients remains disheartening [7-11]. Therefore, the identification of novel, reliable markers for predicting the prognosis of LUSC patients may facilitate clinical decision-making considerably.

This study utilizes transcriptome data from a substantial number of samples to construct a reliable CRGPS and investigates its correlations to the prognosis and immune environment of LUSC patients, as well as therapy responses to certain drugs. Leveraging large samples and multi-center data, we initially identified CRGPS candidate genes and then used them to establish the CRGPS. Internal tests and external comparisons demonstrate that the CRGPS is not only superior to traditional clinical parameters (T stage, age, and gender) and its commutative genes but also outperforms 19 preexisting prognostic signatures for LUSC. The CRGPS score is positively correlated with a poor prognosis in patients with LUSC, indicating its suitability as a prognostic marker for this disease. Furthermore, the CRGPS score is observed to be inversely correlated with the degree of NK cell infiltration in individuals with LUSC—NK cells are a critical antitumorigenic cell type. In some tumors, patients with lower CRGPS scores are more likely to experience enhanced immunotherapy effects, indicating that the CRGPS may have the potential to predict immunotherapy efficacy. LUSC patients with high CRGPS scores are predicted to be sensitive to several drugs. Collectively, these findings indicate that the CRGPS may serve as a reliable predictor of the prognosis of patients with LUSC.

As a frequent complication associated with tumors, abnormalities in the coagulation system may significantly impact the development of certain types of cancers. Activation of the coagulation system can induce inflammatory responses, such as the recruitment and activation of polymorphonuclear cells and macrophages [14], and the cytokines generated in the process can induce immunosuppression and modulate the tumor microenvironment to engender tumor progression [45]. Hypercoagulable states have been linked to an elevated risk of venous thrombosis in patients with cancers located in the pancreas, breasts, and lungs [46]. Cancer-associated venous thrombosis is recognized as an unfavorable prognostic factor in tumor patients and is a major contributor to nontumor mortality in this population [46]. These findings emphasize the essential role of coagulation in the tumor microenvironment and cancer immunity, and they indicate that CRGs may affect cancer prognosis. Nonetheless, only a few studies have investigated the importance of CRGs in LUSC. To this end, this study identified 144 prominent CRGs (expressed differently in LUSC and correlated with patient prognosis) from a combination of datasets and numerous samples and used them to construct prognostic signatures. To the best of our knowledge, no previous study has adopted CRGs to generate LUSC prognostic signatures, indicating the novelty of this study.

The CRGPS has been established as a reliable marker for the prognosis of LUSC. In this study, a series of prognostic signatures for LUSC were constructed from a CRG perspective using 92 machine learning combination algorithms. Tenfold cross-validation was applied to the expression data of 144 CRGs to generate 92 prognostic signatures, which were then verified in six independent cohorts. The results show that the CRGPS achieved the highest median C-index across the six cohorts. Gender, age, and T stage may be potential prognosis predictors for certain tumor types [47,48], but the CRGPS developed in this study is more accurate at predicting the prognoses of LUSC patients. Furthermore, most of the recently proposed LUSC predictive signatures are based on a single algorithm, whereas our CRGPS is distinguished as a signature derived using 92 integrated algorithms. Based on the training cohort, the CRGPS



Fig. 9. Potentially beneficial drugs for LUSC patients with high CRGPS scores. Treatment responses of patients with high- and low-risk CRGPS scores. *p < 0.05, ***p < 0.001.

demonstrated superior accuracy at predicting the prognoses of patients with LUSC compared to 19 recently published signatures, as evidenced by greater C-index and AUC values. Notably, among the median AUC in all the test cohorts, the signatures constructed by Li et al. [49] (3-year AUC = 0.63), Huang et al. [50] (3-year AUC = 0.64, 5-year AUC = 0.64), and Miao et al. [51] (5-year AUC = 0.61) showed slightly better accuracies at three-year and/or five-year survival predictions than our CRGPS (3-year AUC = 0.62, 5-year AUC = 0.60). Nevertheless, the C-index values for these signatures, as with the remaining 16 signatures, are lower than that of our CRGPS across all six test cohorts. Thus, the CRGPS can reasonably predict the prognoses of LUSC patients. In addition, it can be seen from these results that multiple measures (e.g., C-index, and AUC) and multiple datasets should be employed to assess the credibility of prognostic signatures. Furthermore, the CRGPS is an independent prognosis risk factor for LUSC patients based on multiple cohorts and several statistical analyses, indicating that it is a reliable predictor of prognoses for patients with LUSC.

The potential of the CRGPS to predict patient response to immunotherapy is also demonstrated in this study. A previous study by Saidak et al. [52] has shown that the coagulation process is linked to the tumor immune microenvironment, thereby demonstrating a direct correlation between coagulation function and the immune microenvironment. The CRGPS, comprising CRGs—which play an important role in coagulation—was found to be correlated with the infiltration levels of several immune cells (i.e., B cells, follicular helper T cells, and NK cells) in this study. Notably, several algorithms revealed that LUSC patients with high CRGPS risk scores exhibited lower NK cell infiltration levels than LUSC patient with low CRGPS risk scores, with the CRGPS risk score being inversely correlated to NK cell infiltration levels. NK cells are among the most critical immune cells in terms of antitumor activity, and their decreased infiltration tends to indicate a weakening of antitumor abilities in humans [53,54]. This negative correlation between CRGPS and NK cell infiltration indicates that the CRGPS may serve as a risk factor for LUSC patients in terms of predicting patient response to immunotherapy.

Patients with high and low CRGPS risk scores for LUSC displayed different levels of immunoinfiltration (at least, as demonstrated by several immune cell types mentioned earlier), indicating that the CRGPS may be used to predict immune response in LUSC patients. In this study, this hypothesis was initially validated in analyses of breast cancer, melanomas, and NSCLC—a lower proportion of positive response to immunotherapy was observed in the group of tumor patients with high CRGPS scores. However, more direct evidence is still required to verify whether the CRGPS risk score can directly predict immune response in LUSC patients. Taking into account that the prognoses for LUSC patients with high CRGPS scores are significantly worse than those for low scores, this study also explores several drugs that are potentially well suited to patients with high CRGPS scores. The drug sensitivity data indicates that LUSC patients in the high-risk group were more sensitive to five drugs (Staurosporine, Dasatinib, Trametinib, ERK_2440, and BMS.754807), and administration of these drugs may help to achieve more precise clinical treatment of patients with high CRGPS scores.

This study has some limitations. Primarily, the 683 CRGs used to construct the CRGPS developed in this study were derived from coagulation-associated mechanisms recorded in public databases. However, the direct association between these CRGs and coagulation functions requires further validation. It is not clear whether CRGs affect the progression of LUSC by influencing other mechanisms, such as the organization of the tumor extracellular matrix. We did not collect sufficient proteomics data to validate the differential protein expression levels of the CRGs in LUSC vis-à-vis normal lung tissues. Due to the lack of data, we were unable to compare the accuracy of the CRGPS against that of certain clinicopathological parameters (e.g., pathological grade) at predicting the prognoses of LUSC patients. In the future, larger cohort samples are required to validate the link between the CRGPS and the prognoses of LUSC patients. In addition, further in vivo and in vitro experiments are also needed to verify the correlation between the CRGPS and immune cell (e.g., NK cells) infiltration levels and to explore drugs well suited to LUSC patients with different CRGPS characteristics.

5. Conclusions

In this study, multiple machine learning algorithms were integrated to construct a robust CRGPS that can predict the prognoses of LUSC patients. The CRGPS may be an independent prognostic risk factor for LUSC patients and is correlated with some critical immune cell infiltration levels. The CRGPS may also predict the immunotherapy response of certain tumors. Overall, the CRGPS may serve as a tool that facilitates the clinical management of LUSC.

Data availability

Publich data used in this study can be acquired from the ArrayExpress (https://www.ebi.ac.uk/biostudies/arrayexpress), Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/), The Cancer Genome Atlas (TCGA) (https://xenabrowser.net/), the Proteomic Data Commons (https://pdc.cancer.gov/pdc/), the Human Protein Atlas (https://www.proteinatlas.org/), and The Cancer Single-cell Expression Map (https://ngdc.cncb.ac.cn/cancerscem/index). The in-house microarray data used in this study can be obtained from the corresponding author for reasonable reasons.

Funding statement

This work was supported by Guangxi Zhuang Autonomous Region Medical Health Appropriate Technology Development and Application Promotion Project [S2020031], Guangxi Medical High-level Key Talents Training "139" Program (2020), and Guangxi Medical University Key Textbook Construction Project [GXMUZDJC2223].

Human and animal rights

This study was approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University (No. 2021[KY-E-246]). All procedures were performed strictly in compliance with the Declaration of Helsinki 1964 or equivalent ethical principles. For in-house samples, signed informed consent was obtained from corresponding patients.

CRediT authorship contribution statement

Guo-Sheng Li: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rong-Quan He:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. **Zhi-Guang Huang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Hong Huang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Jun Huang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Jun Liu:** Conceptualization, Data curation, Software, Validation, Visualization, Data curation, Data curation, Data curation, Methodology, Writing – original draft. **Jun Liu:** Conceptualization, Data curation, Software, Supervision, Validation, Writing – original draft. **Jun Liu:** Conceptualization, Software, Supervision, Validation, Writing – original draft. **Huang:** Data curation, Software, Supervision, Validation, Writing – original draft. **Hua-Fu Zhou:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. **Jin-Liang Kong:** Conceptualization, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing. **Gang Chen:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Gang Chen reports financial support was provided by Health Commission of Guangxi Zhuang Autonomous Region. Gang Chen reports financial support was provided by Health Department of Guangxi Zhuang Autonomous Region. Other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful for the technical support from the Guangxi Key Laboratory of Medical Pathology in computational pathology and clinical pathology. The authors also thank the contributors of the TCGA (https://tcga-data.nci.nih.gov/), ArrayExpress (https://www.ebi.ac.uk/biostudies/arrayexpress), GEO (http://www.ncbi.nlm.nih.gov/geo/), the Proteomic Data Commons (https://pdc.cancer.gov/pdc/), The Human Protein Atlas (https://www.proteinatlas.org/), and The Cancer Single-cell Expression Map (https://ngdc.cncb.ac.cn/cancerscem/index) for sharing their data on open access.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27595.

References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, CA A Cancer J. Clin. 72 (2022) 7–33.
- [2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249.
- [3] Y. Pan, H. Han, H. Hu, H. Wang, Y. Song, Y. Hao, et al., KMT2D Deficiency Drives Lung Squamous Cell Carcinoma and Hypersensitivity to RTK-RAS Inhibition, Cancer cell, 2022.
- [4] C. Chen, D. Tang, C. Gu, B. Wang, Y. Yao, R. Wang, et al., Characterization of the immune microenvironmental landscape of lung squamous cell carcinoma with immune cell infiltration, Dis. Markers 2022 (2022) 2361507.
- [5] L. Ji, N. Moghal, X. Zou, Y. Fang, S. Hu, Y. Wang, et al., The NRF2 antagonist ML385 inhibits PI3K-mTOR signaling and growth of lung squamous cell carcinoma cells, Cancer Med 12 (2022) 5688–5702.
- [6] E.S. Santos, E. Rodriguez, Treatment considerations for patients with advanced squamous cell carcinoma of the lung, Clin, Lung Cancer 23 (2022) 457-466.
- [7] S.C.M. Lau, Y. Pan, V. Velcheti, K.K. Wong, Squamous Cell Lung Cancer: Current Landscape and Future Therapeutic Options, Cancer cell, 2022.
- [8] R.S. Herbst, D. Morgensztern, C. Boshoff, The biology and management of non-small cell lung cancer, Nature 553 (2018) 446–454.
- [9] Y. Liu, Y. Feng, T. Hou, A. Lizaso, F. Xu, P. Xing, et al., Investigation on the potential of circulating tumor DNA methylation patterns as prognostic biomarkers for lung squamous cell carcinoma, Transl. Lung Cancer Res. 9 (2020) 2356–2366.
- [10] L. Yang, S. Wei, J. Zhang, Q. Hu, W. Hu, M. Cao, et al., Construction of a predictive model for immunotherapy efficacy in lung squamous cell carcinoma based on the degree of tumor-infiltrating immune cells and molecular typing, J. Transl. Med. 20 (2022) 364.
- [11] X. Wang, Z. Huang, L. Li, G. Wang, L. Dong, Q. Li, et al., DNA damage repair gene signature model for predicting prognosis and chemotherapy outcomes in lung squamous cell carcinoma, BMC Cancer 22 (2022) 866.
- [12] H.K. Chew, T. Wun, D.J. Harvey, H. Zhou, R.H. White, Incidence of venous thromboembolism and the impact on survival in breast cancer patients, J. Clin. Oncol. : official journal of the American Society of Clinical Oncology 25 (2007) 70–76.

- [13] H.F. Dvorak, Tumors: wounds that do not heal-A historical perspective with a focus on the fundamental roles of increased vascular permeability and clotting, Semin. Thromb. Hemost. 45 (2019) 576–592.
- [14] L.C. Burzynski, M. Humphry, K. Pyrillou, K.A. Wiggins, J.N.E. Chan, N. Figg, et al., The coagulation and immune systems are directly linked through the activation of interleukin-1alpha by thrombin, Immunity 50 (2019) 1033–10342 e6.
- [15] O. Repetto, V. De Re, Coagulation and fibrinolysis in gastric cancer, Ann. N. Y. Acad. Sci. 1404 (2017) 27-48.
- [16] I. Lal, K. Dittus, C.E. Holmes, Platelets, coagulation and fibrinolysis in breast cancer progression, Breast Cancer Res. 15 (2013) 207.
- [17] Y. Fu, Y. Liu, Y. Jin, H. Jiang, [Value of coagulation and fibrinolysis biomarker in lung cancer patients with thromboembolism], Zhongguo Fei Ai Za Zhi 21 (2018) 583–587.
- [18] T. Ueno, M. Toi, M. Koike, S. Nakamura, T. Tominaga, Tissue factor expression in breast cancer tissues: its correlation with prognosis and plasma concentration, Br. J. Cancer 83 (2000) 164–170.
- [19] G.C. Connolly, R. Chen, O. Hyrien, P. Mantry, A. Bozorgzadeh, P. Abt, et al., Incidence, risk factors and consequences of portal vein and systemic thromboses in hepatocellular carcinoma, Thromb. Res. 122 (2008) 299–306.
- [20] Q. He, J. Yang, Y. Jin, Immune infiltration and clinical significance analyses of the coagulation-related genes in hepatocellular carcinoma, Briefings Bioinf. 23 (2022).
- [21] E. Koncina, S. Haan, S. Rauh, E. Letellier, Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges, Cancers 12 (2020).
 [22] E. Clough, T. Barrett, The gene expression Omnibus database, Methods Mol. Biol. 1418 (2016) 93–110.
- [22] Z. Wang, M.A. Jensen, J.C. Zenklusen, A practical guide to the cancer Genome Atlas T(GA), Methods Mol. Biol. 1418 (2016) 111–141.
- [24] G.T. Consortium, The genotype-tissue expression (GTEx) project, Nat. Genet. 45 (2013) 580–585.
- [25] A.N. Tran, A.M. Dussaq, T. Kennell Jr., C.D. Willey, A.B. Hjelmeland, HPAanalyze: an R package that facilitates the retrieval and analysis of the Human Protein Atlas data. BMC Bioinf, 20 (2019) 463.
- [26] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, et al., Proteomics. Tissue-based map of the human proteome, Science 347 (2015) 1260419.
- [27] Y.J. Chen, T.I. Roumeliotis, Y.H. Chang, C.T. Chen, C.L. Han, M.H. Lin, et al., Proteogenomics of non-smoking lung cancer in east Asia delineates molecular signatures of pathogenesis and progression, Cell 182 (2020) 226–244 e17.
- [28] S. Satpathy, K. Krug, P.M. Jean Beltran, S.R. Savage, F. Petralia, C. Kumar-Sinha, et al., A proteogenomic portrait of lung squamous cell carcinoma, Cell 184 (2021) 4348–4371 e40.
- [29] J. Zeng, Y. Zhang, Y. Shang, J. Mai, S. Shi, M. Lu, et al., CancerSCEM: a database of single-cell expression map across various human cancers, Nucleic Acids Res. 50 (2022) D1147–D1155.
- [30] D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, et al., Phenotype molding of stromal cells in the lung tumor microenvironment, Nat. Med. 24 (2018) 1277–1289.
- [31] B.S. Carvalho, R.A. Irizarry, A framework for oligonucleotide microarray preprocessing, Bioinformatics 26 (2010) 2363–2367.
- [32] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015) e47.
- [33] J.T. Leek, J.D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis, PLoS Genet. 3 (2007) 1724–1735.
- [34] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, Nucleic Acids Res. 45 (2017) D353–D361.
- [35] M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: integrating viruses and cellular organisms, Nucleic Acids Res. 49 (2021) D545–D551.
- [36] M. Kanehisa, Toward understanding the origin and evolution of cellular organisms, Protein Sci. 28 (2019) 1947–1951.
- [37] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
- [38] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell systems 1 (2015) 417–425.
- [39] P. Blanche, J.F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, Stat. Med. 32 (2013) 5381–5397.
- [40] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, et al., Determining cell type abundance and expression from bulk tissues with digital cytometry, Nat. Biotechnol. 37 (2019) 773–782.
- [41] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (2016) 218.
- [42] F. Iorio, T.A. Knijnenburg, D.J. Vis, G.R. Bignell, M.P. Menden, M. Schubert, et al., A landscape of pharmacogenomic interactions in cancer, Cell 166 (2016) 740–754.
- [43] W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (2013) D955–D961.
- [44] D. Maeser, R.F. Gruener, R.S. Huang, oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data, Briefings Bioinf. (2021) 22.
- [45] A.A. Bhat, S. Nisar, M. Singh, B. Ashraf, T. Masoodi, C.P. Prasad, et al., Cytokine- and chemokine-induced inflammatory colorectal tumor microenvironment: emerging avenue for targeted therapy, Cancer Commun. 42 (2022) 689–715.
- [46] M. Tinholt, P.M. Sandset, N. Iversen, Polymorphisms of the coagulation system and risk of cancer, Thromb. Res. 140 (Suppl 1) (2016) S49–S54.
- [47] L. Wu, T. Zou, D. Shi, H. Cheng, M. Shahbaz, M. Umar, et al., Age in combination with gender is a valuable parameter in differential diagnosis of solid pseudopapillary tumors and pancreatic neuroendocrine neoplasm, BMC Endocr. Disord. 22 (2022) 255.
- [48] Z. Sun, W. Tao, X. Guo, C. Jing, M. Zhang, Z. Wang, et al., Construction of a lactate-related prognostic signature for predicting prognosis, tumor microenvironment, and immune response in kidney renal clear cell carcinoma, Front. Immunol. 13 (2022) 818984.
- [49] W. Li, X. Li, L.N. Gao, C.G. You, Integrated analysis of the functions and prognostic values of RNA binding proteins in lung squamous cell carcinoma, Front. Genet. 11 (2020) 185.
- [50] G. Huang, J. Zhang, L. Gong, Y. Huang, D. Liu, A glycolysis-based three-gene signature predicts survival in patients with lung squamous cell carcinoma, BMC Cancer 21 (2021) 626.
- [51] T.W. Miao, D.Q. Yang, F.Y. Chen, Q. Zhu, X. Chen, A ferroptosis-related gene signature for overall survival prediction and immune infiltration in lung squamous cell carcinoma, Biosci. Rep. (2022) 42.
- [52] Z. Saidak, S. Soudet, M. Lottin, V. Salle, M.A. Sevestre, F. Clatot, et al., A pan-cancer analysis of the human tumor coagulome and its link to the tumor immune microenvironment. Cancer immunology, immunotherapy, CII 70 (2021) 923–933.
- [53] J. Chu, F. Gao, M. Yan, S. Zhao, Z. Yan, B. Shi, et al., Natural killer cells: a promising immunotherapy for cancer, J. Transl. Med. 20 (2022) 240.
- [54] E. Russo, M. Laffranchi, L. Tomaipitinca, A. Del Prete, A. Santoni, S. Sozzani, et al., NK cell anti-tumor surveillance in a myeloid cell-shaped environment, Front. Immunol. 12 (2021) 787116.