

ORIGINAL ARTICLE OPEN ACCESS

Identifying Lipid Metabolism-Related Therapeutic Targets and Diagnostic Markers for Lung Adenocarcinoma by Mendelian Randomization and Machine Learning Analysis

Su Wei^{1,2,3}  | Zhou Guangyao^{2,3,4} | Tian Xiangdong^{1,2,3} | Guo Feng^{1,2,3} | Zhang Lianmin^{2,3,4}  | Zhang Zhenfa^{2,3,4} 

¹Department of Endoscopy, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin, China | ²Key Laboratory of Cancer Prevention and Therapy, Tianjin, China | ³Tianjin's Clinical Research Center for Cancer, Tianjin, China | ⁴Department of Lung Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Tianjin, China

Correspondence: Zhang Zhenfa (zhangzhenfa@tmu.edu.cn)

Received: 20 November 2024 | **Revised:** 6 February 2025 | **Accepted:** 8 February 2025

Funding: This study was supported by National Natural Science Foundation of China, 82273119.

Keywords: lipid metabolism | lung adenocarcinoma | machine learning | Mendelian randomization

ABSTRACT

Background: Lipid metabolic disorders are emerging as a recognized influencing factors of lung adenocarcinoma (LUAD). This study aims to investigate the influence of lipid metabolism-related genes (LMRGs) on the diagnosis and treatment of LUAD and to identify significant biomarkers.

Methods: DESeq2 and robust rank aggregation (RRA) analyses were employed to determine the differential expression of LMRGs from TCGA-LUAD and five GEO datasets. Mendelian randomization (MR) was conducted utilizing protein quantitative trait loci (pQTLs) in the deCODE, prot-a, and UKB-PPP Study to estimate causal relationships between plasma proteins and LUAD within the ieu-a-984, ieu-a-965, and FinnGen R10 cohorts as potential drug targets of LUAD. Subsequently, an optimal machine learning model for diagnosing LUAD was established by comparing four models: support vector machine, random forest (RF), glmBoost, and eXtreme Gradient Boosting. Finally, the diagnostic performance of five plasma proteins was validated through nomogram analysis, calibration curve assessment, decision curve analysis (DCA), independent internal and external datasets.

Result: A total of five biomarkers were identified from 1034 LMRGs via MR and differential expression analysis. TNFRSF21 exhibited a positive association with LUAD risk; conversely, BCHE, FABP4, LPL, and PLBD1 demonstrated negative correlations with this risk. The RF machine learning model was determined to be the optimal model for diagnosing LUAD using these five plasma proteins. Ultimately, nomogram construction, calibration curve analysis, DCA, as well as independent internal and external dataset validation confirmed that these biomarkers exhibit excellent diagnostic performance.

Conclusions: BCHE, FABP4, LPL, PLBD1, and TNFRSF21 represent potential novel reliable diagnostic markers as well as therapeutic targets for LUAD.

Su Wei and Zhou Guangyao have contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Thoracic Cancer* published by John Wiley & Sons Australia, Ltd.

1 | Introduction

Lung cancer is a highly prevalent and fatal malignant tumor worldwide [1], accounting for approximately 20% of global cancer-related deaths [2, 3]. Among the primary lung cancers, lung adenocarcinoma (LUAD) represents the most common histological subtype [4]. The advent of multislice spiral CT technology and the widespread utilization of low-dose spiral CT screening have led to an increased detection rate of LUAD. However, these methods suffer from limited specificity and false-positive results [5]. Considering the limitations of existing screening methods, there is an urgent need for more precise, efficient, and convenient screening approaches to further improve diagnostic efficiency. Consequently, it becomes imperative to identify reliable diagnostic marker genes.

Lipid metabolism is closely associated with the occurrence and progression of cancer [6, 7]. Tumor cells exhibit enhanced proliferative capacity and require a greater energy supply compared to normal cells [8, 9]. There is evidence that suggests a correlation between abnormal lipid metabolism and various cancers, including breast cancer, postmenopausal tumors, colorectal cancer, endometrial cancer, pancreatic cancer, and biliary tract cancer [10–12]. Lipid metabolism-related genes (LMRGs) have potential as biomarkers for tumor diagnosis since dysregulation of lipid metabolism plays a significant role in the pathogenesis and progression of malignancies [13–16]. Investigating the underlying mechanisms of lipid metabolism in LUAD has the potential to enhance therapeutic strategies and improve patient survival rates. Abnormal lipid metabolism is linked to lung cancer development and progression [17]. However, the role of LMRGs in LUAD remains poorly understood. Further research is required to comprehend the diagnostic and treatment value of these genes in LUAD.

Mendelian randomization (MR) serves as a valuable tool for investigating causal relationships between exposures and outcomes by leveraging genetic variation as an instrumental variable (IV) [18, 19]. By utilizing genetic differences as substitutes for exposure, MR compares health outcomes among individuals with different genetic variations to estimate the causal effects of risk factors on health while minimizing bias caused by other factors [20–22]. This approach offers a novel way to explore the causal link between illnesses and environmental variables, thereby enhancing our understanding of disease pathophysiology [23–25].

Machine learning represents a pivotal artificial intelligence technique extensively employed for the prediction of cancer diagnostic markers and holds significant promise for forecasting cancer prognosis [26, 27]. However, the development of reliable cancer outcome prediction models for routine clinical use remains a formidable challenge. In this study, we identified differentially expressed lipid metabolism-related genes (DE-LMRGs) in LUAD patients from the TCGA and GEO databases. Subsequently, we performed MR analysis on these DE-LMRGs to identify five genes associated with LUAD. Finally, through comparative analysis of various machine learning methods, we validated BCHE, FABP4, LPL, PLBD1, and TNFRSF21 as potential therapeutic targets and diagnostic markers for LUAD.

2 | Materials and Methods

2.1 | LMRGs Availability

LMRGs were selected through the Gene Set Enrichment Analysis (GSEA) database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the REACTOME database [28]. A total of 1034 LMRGs were extracted from 1 fatty acid metabolism hallmark gene set from the GSEA database, 12 LMRG sets were extracted from the KEGG database, and 22 LMRG sets from the REACTOME database (Table S1-1) [28]. The scheme of this study is illustrated in Figure 1.

2.2 | Data Collection and Processing

We retrieved a total of 541 cases of LUAD and 59 normal cases from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) for subsequent differential analysis. Additionally, we obtained six LUAD cohorts (GSE30219, GSE31210, GSE32863, GSE43458, GSE75037, and GSE40419) from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Among the five cohorts (GSE30219, GSE31210, GSE32863, GSE43458, and GSE75037), there were 532 LUAD tissues and 205 normal cases used to perform robust rank aggregation (RRA) analysis and receiver operating characteristic (ROC) analysis for diagnostic biomarkers. The cohort GSE40419, consisting of 87 LUAD tissues and 77 normal cases, served as an external validation cohort for diagnostic markers of LUAD. Corresponding clinical information and platform-probe annotations were also downloaded. Detailed information about these datasets can be found in Table S2.

The LUAD GWAS data were obtained from three studies, namely FinnGen R10, TRICL, and ILCCO [29]. The FinnGen R10 study provided 1590 cases of LUAD along with 314,193 controls (<https://r10.finnngen.fi/>). The TRICL (ieu-a-984) study comprised a total of 11,245 cases of LUAD and 54,619 controls (<https://gwas.mrcieu.ac.uk/datasets/>). Additionally, the ILCCO (ieu-a-965) study included 3442 cases of LUAD and 14,894 controls (<https://gwas.mrcieu.ac.uk/datasets/>).

2.3 | Differential Expression Genes (DEGs) Analysis for LMRGs

The counts format of gene expression data was extracted by setting class “TCGA-LUAD” and gene type “protein coding” using “TCGAbiolinks” package. Firstly, the TCGA-LUAD dataset were executed DEGs analysis by DESeq2 package. Subsequently, the LMRGs were extracted from TCGA-LUAD dataset to executed DEGs analysis. The R program “ggplot2” and “ComplexHeatmap” were used to display volcano plots and heatmap of DEGs. The cutoff values were adjusted p value < 0.05 and $|\log_2 \text{FC}| > 1$.

2.4 | RRA Analysis for LMRGs

We utilized the RRA approach to integrate the aforementioned five microarray datasets, aiming to mitigate bias and errors.

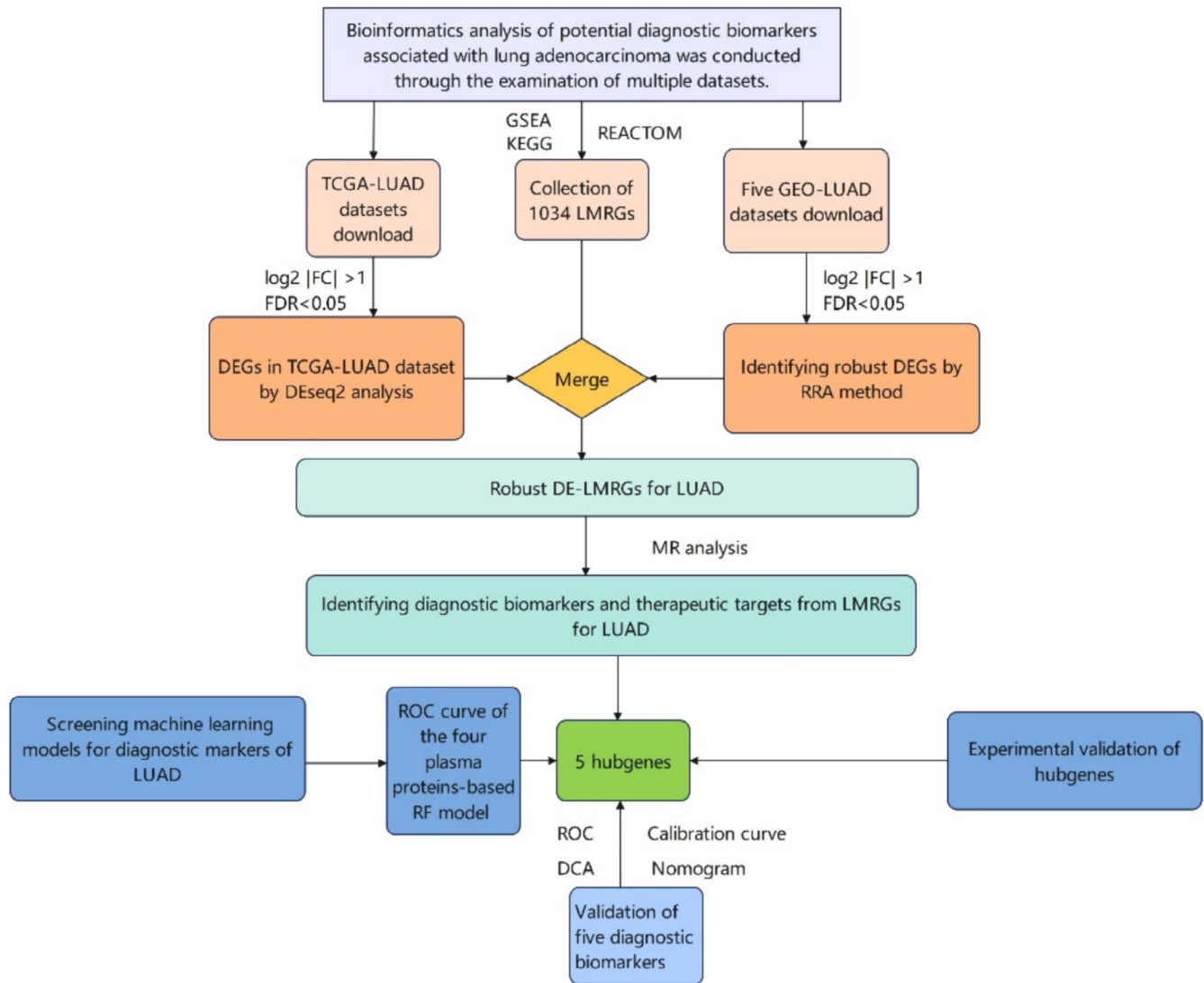


FIGURE 1 | Study flow chart.

Initially, the “limma” package was utilized in each dataset to categorize upregulated and downregulated genes. Subsequently, robust DEGs were filtered using the “RobustRankAggreg” package on these categorized genes. A heatmap displaying the top 30 robust DEGs was generated using the “pheatmap” package. The cutoff criteria for selection were $|\log_2 FC| > 1$ and $FDR < 0.05$.

2.5 | MR Analysis

We selected single nucleotide polymorphisms (SNPs) that were strongly associated with plasma proteins at a significance level of $p < 5e-08$ as IVs. These SNPs should exhibit low linkage disequilibrium ($R^2 < 0.001$) and are located within a genetic distance of less than 10,000 kb. SNPs with weak associations or low phenotypic variance (F -test value < 10) were excluded from the analysis. The outcome data used in this study were obtained from the GWAS summary dataset available at IEU and FinnGen R10 database. Specifically, we utilized the GWAS IDs ieu-a-984, ieu-a-965, and finngen_R10_C3_NSCLC_ADENO_EXALLC for our analyses. After removing SNPs

with $p < 5 \times 10^{-5}$ and plasma proteins without IVs in the LUAD dataset, we employed five different methods (MR Egger, Wald ratio, weighted median, weighted mode methods, and inverse variance weighted [IVW]) using the “TwoSampleMR” package to determine causal relationships between exposure factors and outcomes. The p value was used with a significant threshold of < 0.05 . All five methods yielded beta values that were either greater than or less than zero, indicating positive results. Heterogeneity was assessed using Cochran’s Q test with a significance threshold of $p > 0.05$, while pleiotropy was evaluated using the Egger intercept test, also with a significance threshold of $p > 0.05$.

2.6 | Acquisition of Intersection Genes

Using the “VennDiagram” package, we identified 1 gene that intersected between a total of 36 genes with $OR > 1$ and 16 up-regulated LMRGs. Additionally, by analyzing another set of 38 genes with $OR < 1$ and 36 downregulated LMRGs using the same package, we found a total of 4 intersected genes.

2.7 | Constructing Machine Learning Models for LUAD

The “caret” package was utilized for the development of machine learning methods, including random forest (RF), support vector machine (SVM), generalized linear model (GLM), and eXtreme Gradient Boosting (XGB). These four machine learning methods were employed to construct prediction models with default parameters using the identified five proteins. A total of 737 LUAD samples from GSE30219, GSE31210, GSE32863, GSE43458, and GSE75037 datasets were randomly divided into a training cohort (70%, $N=516$) and a test cohort (30%, $N=221$). The “caret” package facilitated robust parameter tuning through grid search and cross-validation techniques to identify optimal model parameter combinations and accurately estimate model performance. Residual distributions were analyzed and visualized using the “DALEX” package to assess feature importance in the four machine learning models. Performance evaluation and visualization of the four models were conducted using the “pROC” software package. By comparing area under the curve (AUC) values obtained through five-fold cross-validation among these models, we can determine the optimal choice based on the highest AUC.

2.8 | Internal and External Dataset Validation Analysis for Diagnostic Performance of Plasma Proteins

The optimal machine learning model was selected to predict the diagnostic performance of the identified five plasma proteins. ROC curve analysis was independently conducted on five datasets (GSE30219, GSE30210, GSE32863, GSE43458, and GSE75037) to validate the diagnostic value of the model. Additionally, an external dataset (GSE40419) was employed for further validation of the predictive ability of the model. Higher AUC values indicate superior diagnostic performance of the identified plasma proteins.

2.9 | Development and Evaluation of a LMRGS Clinicopathologic Nomogram

Utilizing the “rms” R package, we developed a nomogram model to individualize risk assessment and prognosis prediction for patients with LUAD, thereby optimizing treatment strategies. Each predictive variable was assigned a score based on its respective model weight or coefficient, reflecting its influence on the target variable. The cumulative score, known as the “total score,” represents the combined impact of all predictor variables. Additionally, deviation correction analyses were employed to evaluate LUAD risk using identified diagnostic markers in selected machine learning models. Furthermore, decision curve analysis (DCA) and sensitivity analysis were conducted to explore these predictive performances and enable the utilization of diagnostic markers for predicting treatment outcomes in patients.

2.10 | RNA Isolation and Quantitative RT-PCR Assay

Total RNA was isolated from LUAD tissues using the SPARKeasy Improved Tissue/Cell RNA Kit (SparkJade, AC0202). The

complementary DNA (cDNA) was synthesized as per the manufacturer's instructions, utilizing the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). qRT-PCR was performed with the SYBR Green PCR kit (Takara Bio, Otsu, Japan) on a Step One Real-Time PCR system (Thermo Fisher Scientific). The relative gene expression levels were quantified by employing the $2^{-\Delta\Delta CT}$ method. Primer sequences:

- *TNFRSF21-F* ATTGGCACATACCGCCATGTT
- *TNFRSF21-R* GGCTTGTGTGGTACAATGCTC
- *BCHE-F* GTCAGAGGGATGAACTTGACAG
- *BCHE-R* TGAATCGAAGTCTACCAAGAGGT
- *FABP4-F* ACTGGGCCAGGAATTTGACG
- *FABP4-R* CTCGTGGAAGTGACGCCTT
- *LPL-F* TCATTCCCGGAGTAGCAGAGT
- *LPL-R* GGCCACAAGTTTTGGCACC
- *PLBD1-F* CAAGATAAGTGACCCGGAAAA
- *PLBD1-R* TGTGCCATCACATAGCCTGTA

3 | Results

3.1 | Identification of Robust DE-LMRGs for LUAD

To identify differentially expressed LMRGs (DE-LMRGs) between LUAD and normal tissues, we initially extracted 1034 LMRGs from GSEA, KEGG, and REACTOME datasets (Figure S1; Table S1-2). Subsequently, DESeq2 analysis identified 3543 upregulated genes and 1824 downregulated genes (Figure 2A; Table S3), including 185 upregulated LMRGs and 84 downregulated LMRGs (Figure 2B; Table S4), between 541 LUAD cases and 59 normal cases from TCGA database. Furthermore, RRA analysis revealed 353 upregulated genes and 693 downregulated genes (Figure 2C; Table S5), including 18 upregulated LMRGs and 42 downregulated LMRGs (Figure 2D; Table S6), between 532 LUAD tissues and 205 normal tissues in five integrated GEO datasets. Finally, a total of 16 upregulated genes and 36 downregulated genes were identified through the intersection of LMRGs, DEGs from TCGA and GEO datasets (Figure 2E,F; Table S7).

3.2 | Identification of Potential Diagnostic Biomarkers and Therapeutic Targets From LMRGs for LUAD

Given their genetic susceptibility of individuals to diseases, plasma proteins expression quantitative trait loci (pQTLs) have emerged as promising targets for drug development and diagnostic biomarkers [30]. Utilizing LMRGs and screening criteria for IVs, we identified 189 pQTLs (UKB-PPP), 129 pQTLs (prot-a), and 327 pQTLs (deCODE) associated with LMRGs (Tables S8–S10). MR analysis between these pQTLs and three outcomes of LUAD (ieu-a-965, ieu-a-984, and Finnngen R10) revealed a total of 74 significant union-proteins across the UKB-PPP study (24 proteins), prot-a (17 proteins), and deCODE study (48 proteins) with a statistical significance threshold of p value

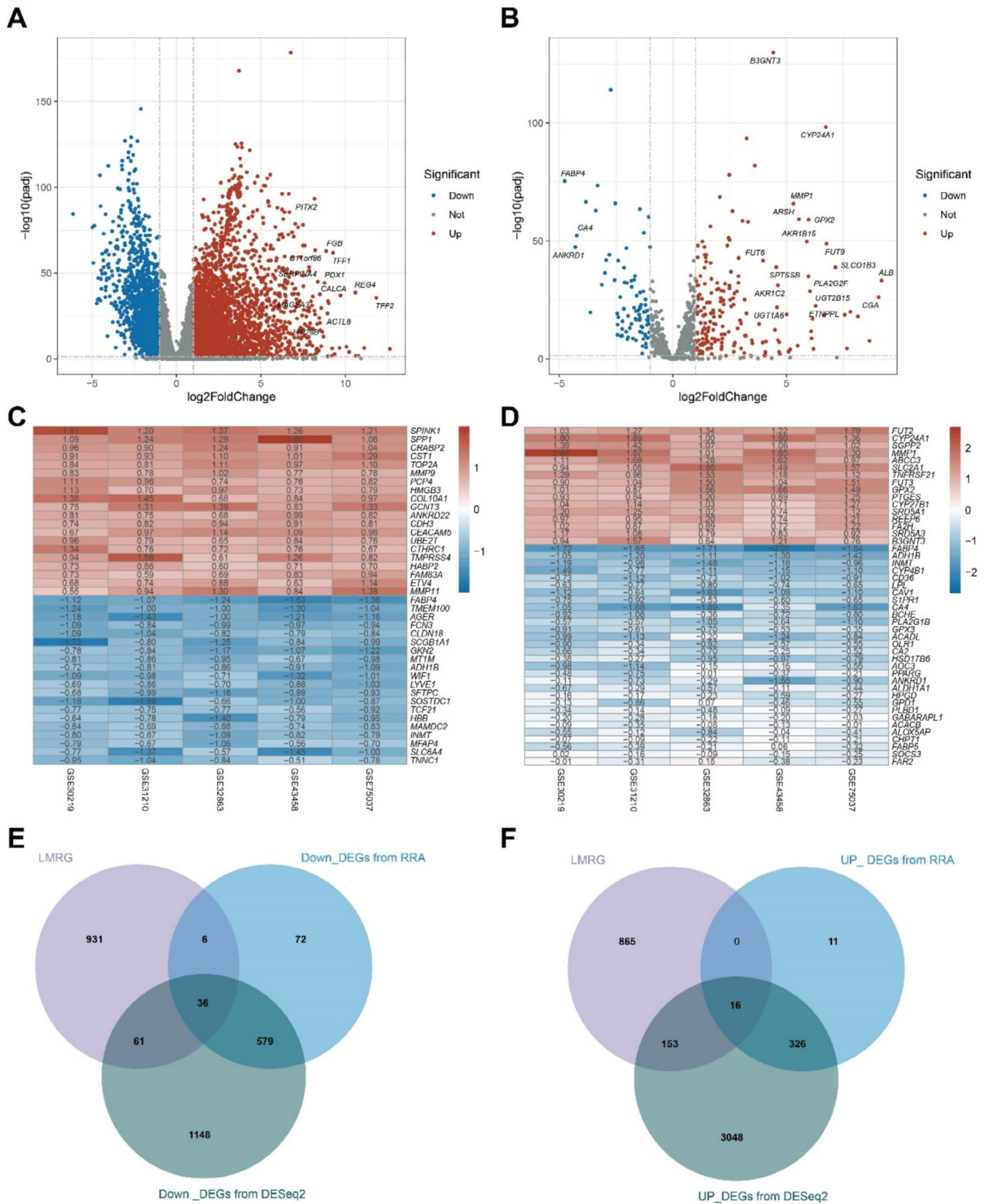


FIGURE 2 | Identification of Robust DE-LMRGs for LUAD. (A) Volcano plot of differentially expressed genes in TCGA-LUAD dataset by DESeq2 analysis. (B) Volcano plot of differentially expressed LMRGs in TCGA-LUAD dataset by DESeq2 analysis. (C) Heatmap of differentially expressed genes in GSE30219, GSE30210, GSE32863, GSE43458, and GSE75037 GEO dataset by RRA analysis. (D) Heatmap of differentially expressed LMRGs in GSE30219, GSE30210, GSE32863, GSE43458, and GSE75037 GEO dataset by RRA analysis. (E) Venn diagram for identifying down-regulated DE-LMRGs. (F) Venn diagram for identifying up-regulated DE-LMRGs.

<0.05 (Table S11). The Venn diagram and volcano plot demonstrated that one risk-associated LMRG gene *TNFRSF21*, along with four protective LMRG genes *BCHE*, *FABP4*, *LPL*, and *PLBD1*, were implicated in LUAD development (Figure 3A,B). MR analysis using the IVW method showed a significant positive causal association between *TNFRSF21* and LUAD, with an odds ratio (OR) of OR = 1.218 (95% CI: 1.083–1.369) and $p < 0.049$ (Figure 3C). *BCHE*, *FABP4*, *LPL*, and *PLBD1* exhibited significant negative associations with LUAD accompanied by $p < 0.05$

(Figure 3C). The OR values for *FABP4* and *PLBD1* were 0.682 (95% CI: 0.495–0.938) and 0.738 (95% CI: 0.565–0.963), respectively. The OR values for *BCHE* in both deCODE and UKB-PPP databases were found to be significantly associated with LUAD, with ORs of 0.928 (95% CI: 0.868–0.993) and 0.923 (95% CI: 0.868–0.982), respectively. Additionally, the OR value for *LPL* in the UKB-PPP database was observed as significantly associated with LUAD in both FinnGen R10 dataset and ieu-a-965 dataset, yielding ORs of 0.717 (95% CI: 0.561–0.915) and

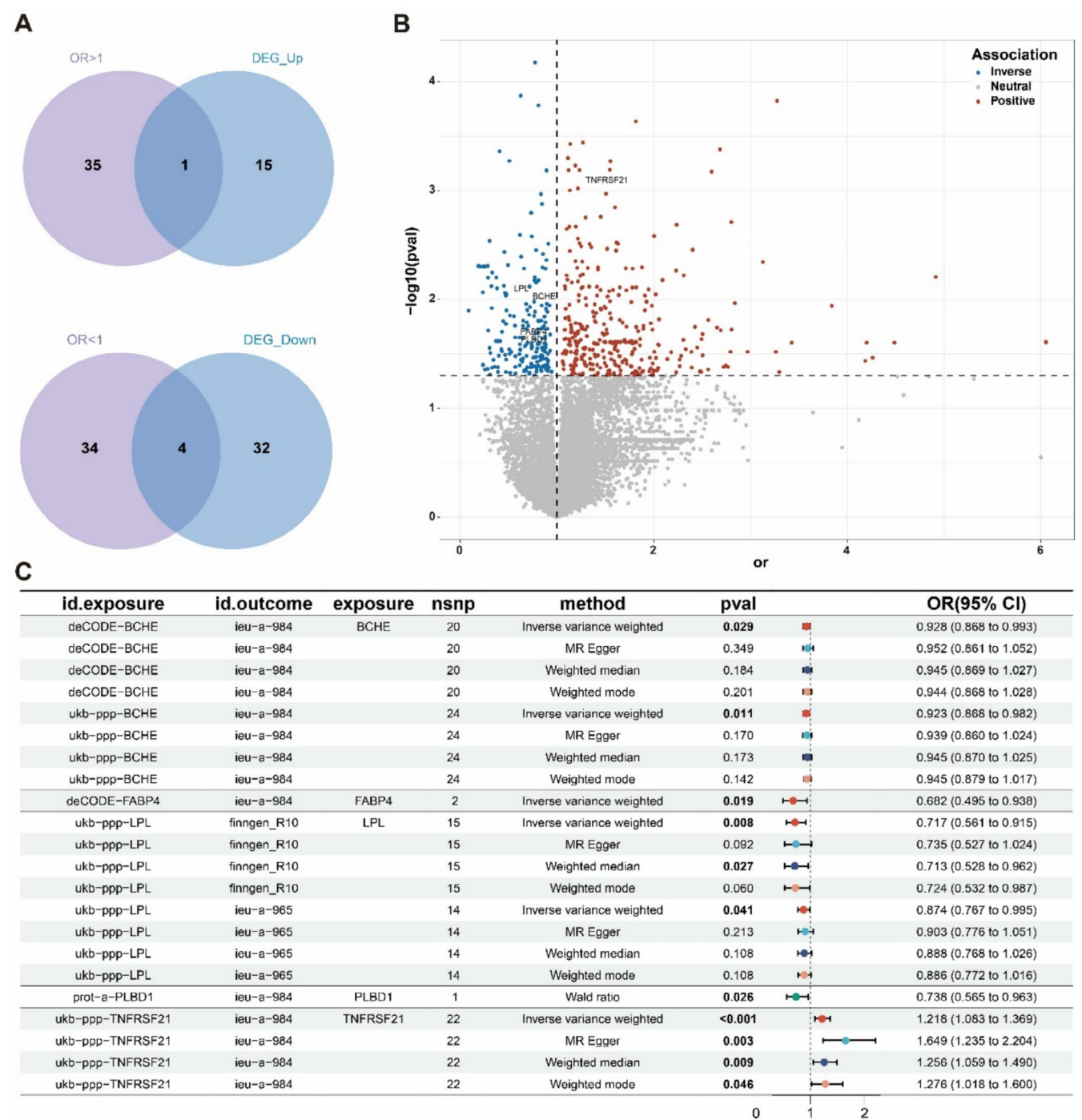


FIGURE 3 | Identification of potential diagnostic biomarkers and therapeutic targets from LMRGs for LUAD. (A) Venn diagram illustrating the potential drug targets of LUAD between DE-LMRGs and the MR-identified proteins. (B) Volcano plots revealing the association of proteins based on OR in the deCODE and UKB-PPP studies. (C) OR, 95% confidence intervals (CI), and p value for the effect of plasma proteins on LUAD estimated by the IVW, MR Egger, weighted median, weighted mode, or Wald ratio approaches.

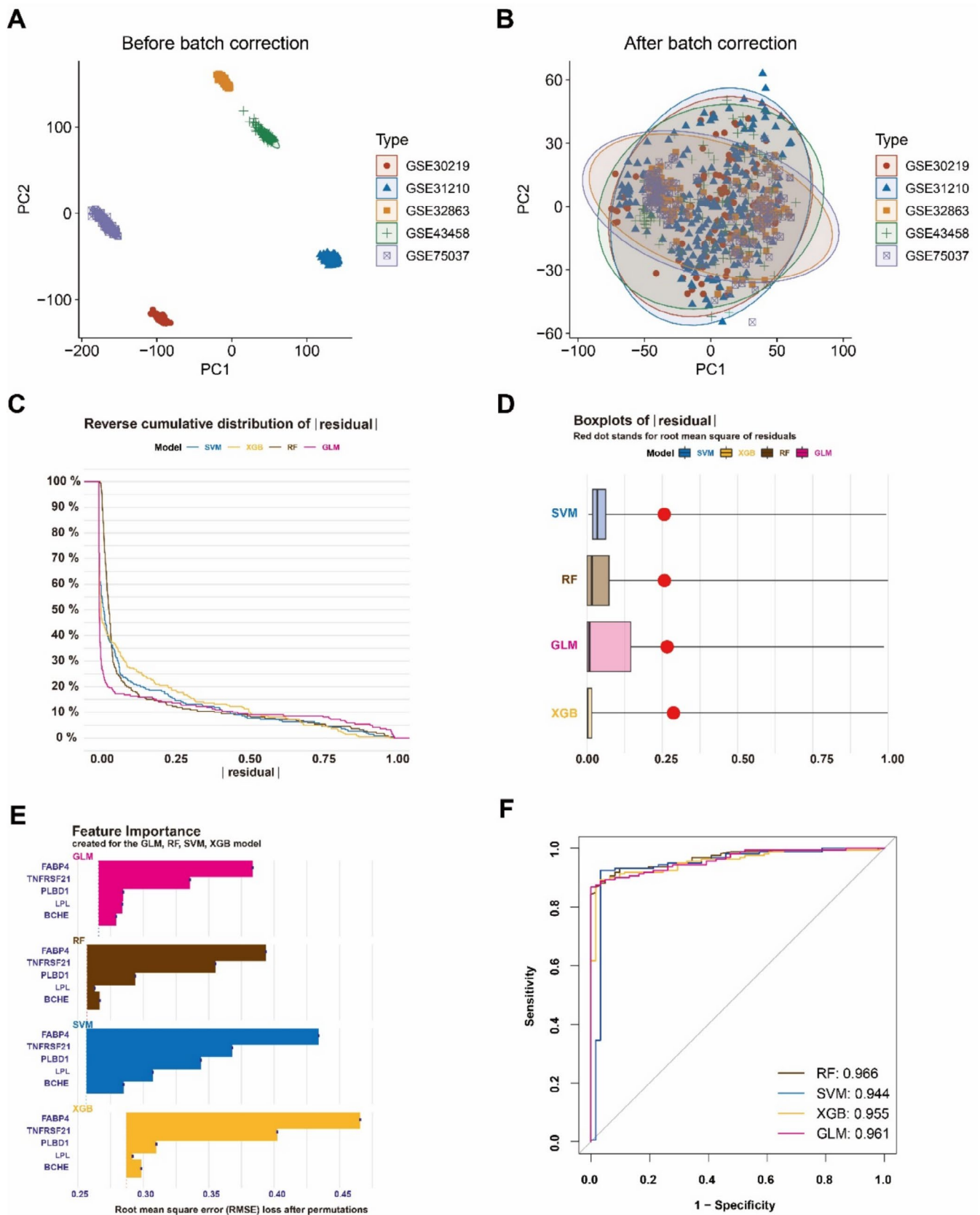


FIGURE 4 | Screening machine learning models for diagnostic markers of LUAD. (A) PCA analysis for five GEO datasets before batch correction. (B) PCA analysis for five GEO datasets after batch correction. (C) Cumulative residual distribution for SVM, RF, GLM, and XGB machine learning models. (D) Residuals boxplots for each machine learning model. (E) The important features of five plasma proteins in each machine learning model. (F) ROC curves for four machine learning models using five-fold cross-validation on the test cohort.

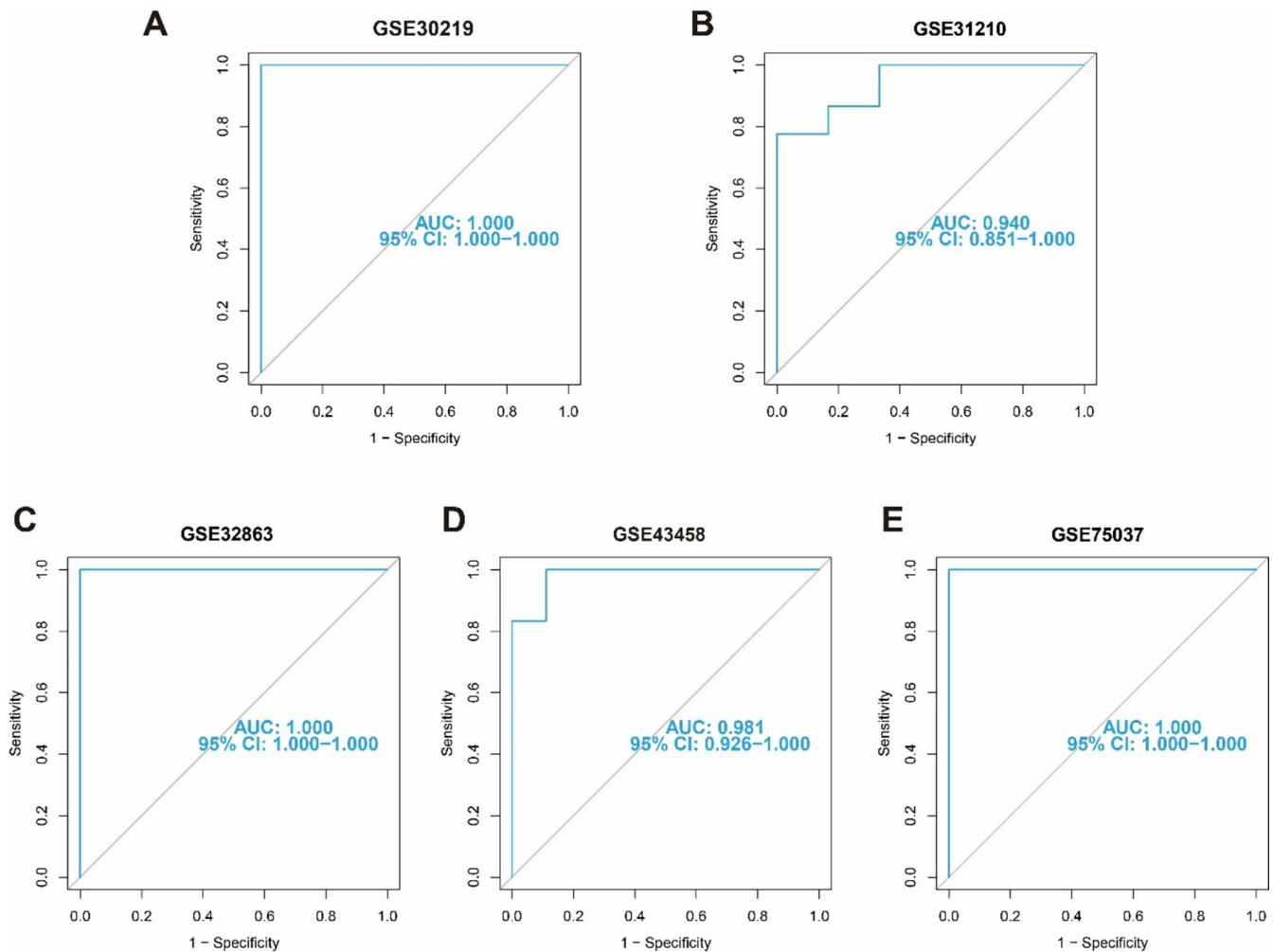


FIGURE 5 | ROC curve of the four plasma protein-based RF model. ROC curve of the four plasma proteins-based RF model in GSE30219 (A), GSE31210 (B), GSE32863 (C), GSE43458 (D), and GSE75037 (E) cohorts.

0.874 (95% CI: 0.767–0.995), respectively. These results suggest that BCHE, FABP4, LPL, PLBD1, and TNFRSF21 were potential diagnostic biomarkers and therapeutic targets for LUAD.

3.3 | Construction of Machine Learning Models

Since the persistent suitability of proteins as diagnostic markers, we employed four machine learning models (RF, SVM, GLM, and XGB) to evaluate the diagnostic performance of the aforementioned candidate genes. Firstly, in order to obtain a large sample dataset for accurate machine learning analysis, we standardized and merged the expression values of each gene from GSE32863, GSE30219, GSE31210, GSE43458, and GSE75037 datasets to create a comprehensive sample pool consisting of 532 LUAD samples and 205 normal tissues. Principal component analysis (PCA) also demonstrated successful elimination of batch effects across these five datasets (Figure 4A,B). Subsequently, the “DALEX” package was utilized to visualize test set residuals across each model. Notably, RF machine learning models exhibited relatively lower residuals (Figure 4C,D). The five plasma proteins in each model were ranked using root mean square error (RMSE), with FABP4 displaying the highest feature importance followed by TNFRSF21, PLBD1, LPL, and BCHE (Figure 4E). Finally, we evaluated the

performance of each algorithm on the large dataset using five-fold cross-validation and calculated the AUC of the ROC curve. The RF machine learning model demonstrated the highest AUC (AUC-RF=0.966, AUC-SVM=0.944, AUC-XGB=0.955, AUC-GLM=0.961; Figure 4F).

3.4 | Evaluation of the Prediction Model

To further assess the predictive efficacy of the RF model, we validated our five-gene prediction model across five independent datasets. The ROC curve analysis demonstrates that the RF predictive model based on the five-gene signature exhibits commendable performance. Specifically, the AUC values for datasets GSE30219, GSE30210, GSE32863, GSE43458, and GSE75037 are 1.0, 0.940, 1.0, 0.981, and 1.0, respectively (Figure 5A–E). These findings underscore the robustness of our predictive model.

Moreover, a nomogram was employed to estimate the scores of FABP4, TNFRSF21, PLBD1, LPL, and BCHE in order to facilitate intuitive prediction of the risk of onset in a specific patient population (Figure 6A). The calibration curve demonstrates that the RF model exhibits high predictive accuracy with minimal discrepancy between the actual and predicted risks associated with LUAD

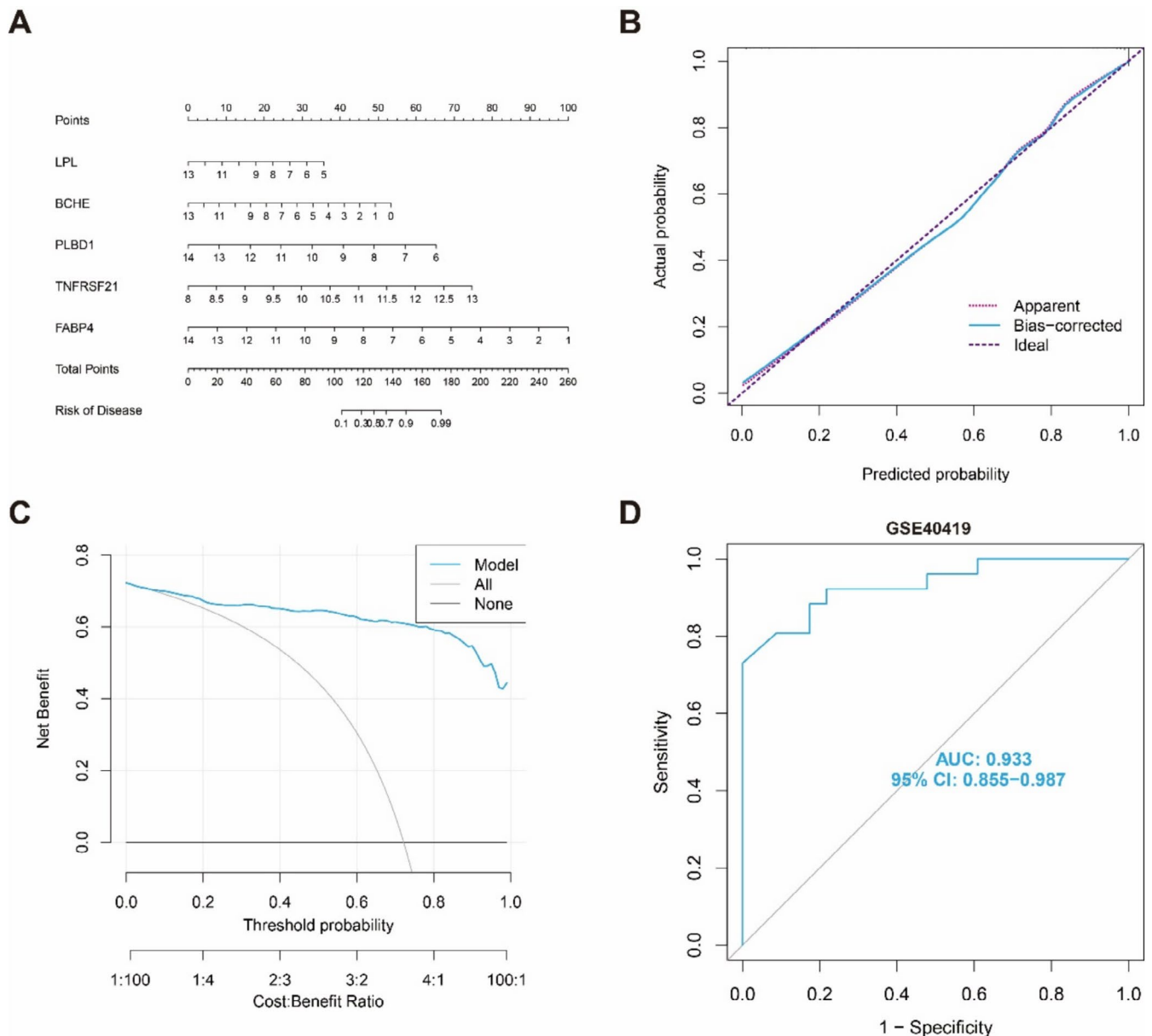


FIGURE 6 | Validation of five diagnostic biomarkers. (A) Nomogram of the five plasma proteins for predicting the risk of LUAD. (B, C) Calibration curve (B) and DCA (C) to evaluate the nomogram's predictive efficiency. (D) ROC analysis verifying the five proteins-based RF model in external validation cohorts (GSE40419).

patients (Figure 6B). Furthermore, DCA analysis reveals that our nomogram possesses exceptional accuracy and has potential utility in guiding clinical decision-making (Figure 6C). Finally, we validated our five-gene prediction model using an external validation dataset and observed excellent performance of the five-gene signature with an AUC value of 0.933 in the GSE40419 dataset (Figure 6D), indicating that our diagnostic biomarkers are equally effective at distinguishing LUAD patients from normal individuals.

3.5 | The Experiment of Genes Involved in the Risk Signature

In order to validate the reliability of our findings, we conducted additional biological experiments. For this study, we acquired seven pairs of samples comprising both normal and tumor tissues

from patients diagnosed with LUAD. Remarkably, these samples exhibited consistent expression patterns. As anticipated, the expression levels of *BCHE*, *FABP4*, *LPL*, and *PLBD1* were found to be significantly downregulated in all seven pairs of LUAD tissue samples (Figure 7A–D). Furthermore, *TNFRSF21* demonstrated a notable upregulation across all seven pairs of LUAD tissue samples (Figure 7E).

4 | Discussion

Effective diagnostic biomarkers and drug targets are crucial for the clinical diagnosis and treatment strategy of LUAD [31, 32]. Disordered lipid metabolism is a malignant characteristic of LUAD that contributes to its carcinogenesis and progression [12, 33]. In this study, we identified 66 upregulated LMRGs and

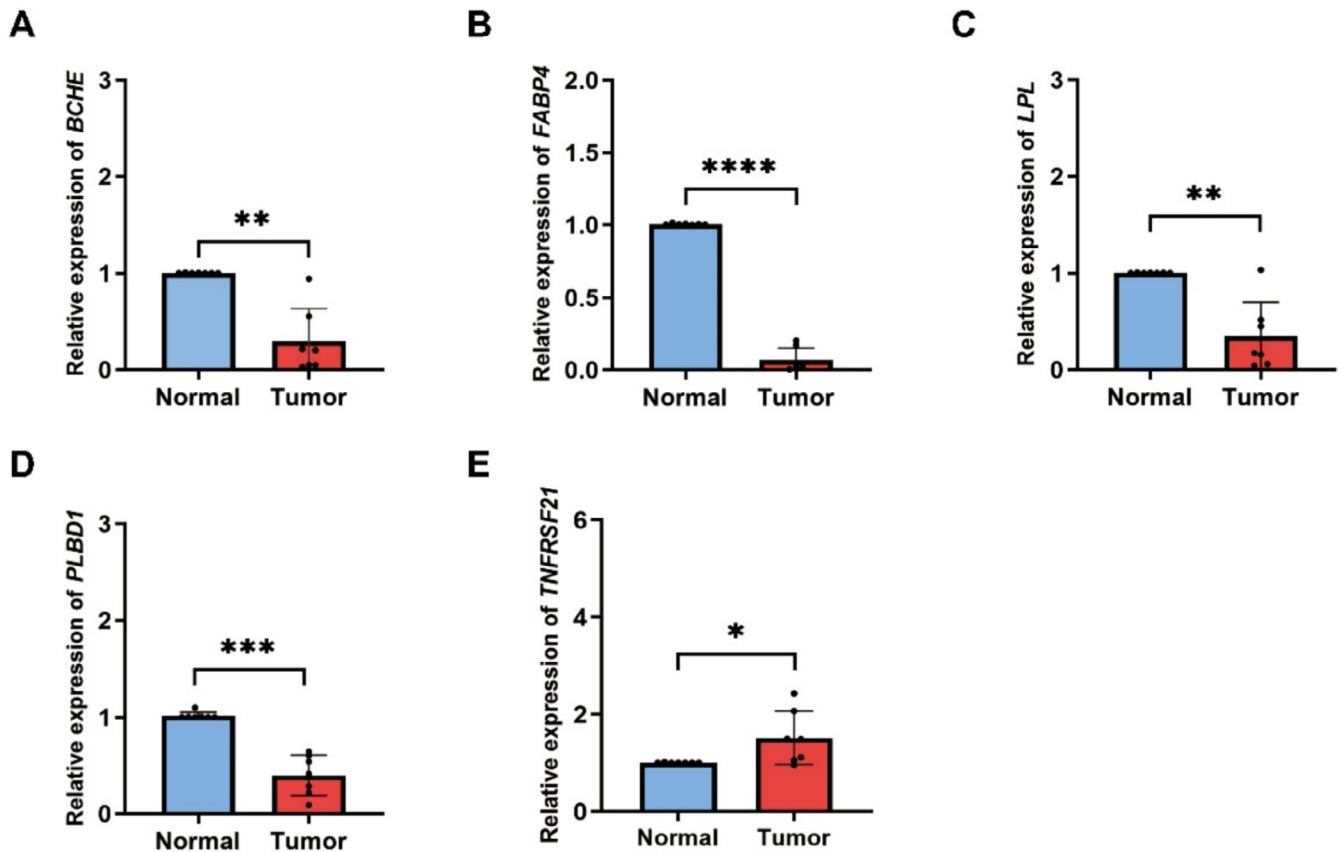


FIGURE 7 | The experimental validation of hubgenes. The expression of *BCHE* (A), *FABP4* (B), *LPL* (C), *PLBD1* (D) and *TNFRSF21* (E) in normal tissue and tumor tissue of LUAD patients. A *t*-test was used to compare the expression of genes between normal and tumor. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

36 downregulated LMRGs from TCGA and GEO datasets using DESeq2 and RRA analysis after selecting 1034 LMRGs from GSEA, KEGG, and REACTOME datasets. We then estimated the causal relationship between plasma proteins and LUAD in ieu-a-965, ieu-a-984, and FinnGen R10 cohort by analyzing pQTLs of these 102 LMRGs in deCODE, prot-a, UKB-PPP Study through MR analysis. *BCHE*, *FABP4*, *LPL*, and *PLBD1* were significantly negatively associated with LUAD accompanied by $OR > 1$ and $p < 0.05$ while *TNFRSF21* was significantly positively causally associated with LUAD with $OR < 1$ and $p < 0.049$. These five plasma proteins were considered as potential therapeutic targets for treating LUAD. Finally, the exceptional diagnostic performance of these five LMRGs was demonstrated through machine learning and validation using internal or external datasets, with AUC values ranging from 0.933 to 1.000. In conclusion, our findings highlight *BCHE*, *FABP4*, *LPL*, *PLBD1*, and *TNFRSF21* as novel and reliable diagnostic markers and potential therapeutic targets for LUAD.

As a member of the cholinesterase family, butyrylcholinesterase (*BCHE*) is an enzyme that catalyzes the hydrolysis of acylcholine. It belongs to the alpha-glycoprotein family and is present in both the nervous system and liver [34]. *BCHE* has been implicated in tumor cell choline metabolism, potentially affecting their energy supply and signal transduction. Previous studies have demonstrated associations between *BCHE* expression levels and cell adhesion, differentiation, apoptosis, as well as tumorigenesis [35, 36]. Furthermore, *BCHE* expression levels may vary according to tumor stage [37]. Notably, high

expression of *BCHE* has been reported in breast cancer [36], while minimal expression has been observed in colorectal cancer [38]. In this study, we observed low levels of *BCHE* expression specifically in LUAD.

Fatty acids in adipose tissue (*FABP4*), a member of the fatty acid-binding protein superfamily [39], play a pivotal role in regulating intracellular transport and metabolism of fatty acids. This regulation is crucial for lipid metabolism and signal transduction, thereby governing cell growth, differentiation, and immune response. Notably, *FABP4* exhibits low expression levels in endometrial cancer [40]. Furthermore, the expression of *FABP4* regulates the development of obesity-associated breast cancer and pancreatic cancer [41, 42]. Our study revealed that *FABP4* was a protective factor for LUAD patients.

Low expressions of lipoprotein lipase (*LPL*) have been correlated with worse overall survival in LUAD patients [12, 43]. This association may be attributed to the involvement of *LPL* in lipid metabolism, thereby contributing to the maintenance of tumor microenvironment homeostasis and suppression of tumor growth and metastasis. Our study corroborates these findings through the utilization of MR and machine learning analysis.

TNFRSF21, also known as Death Receptor 6 (*DR6*) [44], belongs to the tumor necrosis factor receptor superfamily [45]. It exhibits significant upregulation in certain lung cancer cell lines and may be associated with cancer progression, invasion,

and metastasis [46]. TNFRSF21 plays a role in apoptosis induction, immune system regulation, and inhibiting immune evasion and metastasis in lung cancer. In our research, we observed elevated expression of TNFRSF21 in LUAD and identified it as a potential risk factor.

The phospholipase B domain containing 1 (PLBD1) was initially identified in neutrophils [47]. It was predicted to possess phospholipase activity and to be involved in the catabolic process of phospholipids. Vanhaverbeke et al. discovered that PLBD1 and QSOX1 were upregulated, serving as novel independent markers for left ventricular dysfunction following myocardial infarction [48]. In contrast to our findings, PLBD1 was downregulated and considered a protective factor for LUAD. Furthermore, we identified that together with BCHE, FABP4, LPL, and TNFRSF21, it could serve as promising diagnostic markers for LUAD.

Cao et al. identified a panel of 11 LMRGs, namely *LDHA*, *SEC14L3*, *LPL*, *ACSS3*, *GAPDH*, *ESYT3*, *SEC14L4*, *CYP17A1*, *CFTR*, *SLC16A1*, and *CIDEA*, that exhibited high accuracy in predicting overall survival (OS) for LUADs. The constructed risk model using LASSO regression yielded an AUC of 0.722 for 5-year OS [49]. Similarly, another study demonstrated the prognostic value of a nine-gene LMRGs signature consisting of *CYP4B1*, *KLF4*, *DPEP2*, *PTGDS*, *CYP27A1*, *ACSS3*, *HSD17B13*, *HPGDS*, and *FA2H* with an AUC of 0.646 for 5-year OS based on time-dependent ROC curves [17]. Furthermore, the emergence of plasma proteins as highly promising diagnostic markers due to their minimally invasive nature and ease of detection is widely acknowledged. By integrating MR analysis with DEGs and machine learning techniques in our study, five plasma proteins were identified as potential therapeutic targets and diagnostic markers with excellent performance in internal and external validation databases (AUC values ranging from 0.933 to 1.000). Compared with previous results, the five plasma protein signature we proposed demonstrates superior diagnostic accuracy for LUAD.

5 | Conclusion

Overall, we have identified five biomarkers associated with LMRGs as potential drug targets for LUAD. These five proteins exhibited exceptional diagnostic performance, with AUC values ranging from 0.933 to 1.000. Therefore, BCHE, FABP4, LPL, PLBD1, and TNFRSF21 represent novel and reliable diagnostic markers and therapeutic targets for LUAD that hold promise for effective clinical applications.

Author Contributions

S.W. and Z.Z. were involved in the conception and design of the study; S.W., Z.G., and T.X. completed the data analysis; S.W. and G.F. performed the biological validation experiment; S.W. finished writing the manuscript; T.X. and Z.L. contributed to manuscript revisions. Z.Z. supervised the overall study. All authors reviewed and approved the final manuscript.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (82273119) and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-010A).

Ethics Statement

The studies involving human participants were reviewed and approved by the Tianjin Medical University Cancer Institute and Hospital Ethical Board, which gave the study its approval. Batch number: bc2023152.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets supporting the conclusions of this article are available on the GEO website (<https://www.ncbi.nlm.nih.gov/geo/>) and the Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). The original contributions presented in the study are included in the article and [Supporting Information](#).

References

1. H. Sung, J. Ferlay, R. L. Siegel, et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians* 71, no. 3 (May 2021): 209–249.
2. M. Zhu, Q. Zeng, T. Fan, et al., "Clinical Significance and Immunometabolism Landscapes of a Novel Recurrence-Associated Lipid Metabolism Signature in Early-Stage Lung Adenocarcinoma: A Comprehensive Analysis," *Frontiers in Immunology* 13 (2022): 783495.
3. Y. Liu and J. Lu, "Mechanism and Clinical Application of Thymosin in the Treatment of Lung Cancer," *Frontiers in Immunology* 14 (2023): 1237978.
4. B. D. Hutchinson, G. S. Shroff, M. T. Truong, and J. P. Ko, "Spectrum of Lung Adenocarcinoma," *Seminars in Ultrasound, CT, and MR* Jun 2019;40(3):255–264. 40, no. 3 (2019): 255–264, <https://doi.org/10.1053/j.sult.2018.11.009>.
5. C. Zhang, W. Leng, C. Sun, et al., "Urine Proteome Profiling Predicts Lung Cancer From Control Cases and Other Tumors," *eBioMedicine* 30 (April 2018): 120–128.
6. N. M. Anderson, P. Mucka, J. G. Kern, and H. Feng, "The Emerging Role and Targetability of the TCA Cycle in Cancer Metabolism," *Protein & Cell* 9, no. 2 (2018): 216–237.
7. N. Mabrouk, B. Lecoer, A. Bettaieb, C. Paul, and F. Vegran, "Impact of Lipid Metabolism on Antitumor Immune Response," *Cancers (Basel)* 14, no. 7 (April 2022): 1850.
8. S. McGuirk, Y. Audet-Delage, and J. St-Pierre, "Metabolic Fitness and Plasticity in Cancer Progression," *Trends Cancer* 6, no. 1 (2020): 49–61.
9. X. Bian, R. Liu, Y. Meng, D. Xing, D. Xu, and Z. Lu, "Lipid Metabolism and Cancer," *Journal of Experimental Medicine* 218, no. 1 (January 2021): e20201606.
10. D. Delmas, C. Rebe, O. Micheau, et al., "Redistribution of CD95, DR4 and DR5 in Rafts Accounts for the Synergistic Toxicity of Resveratrol and Death Receptor Ligands in Colon Carcinoma Cells," *Oncogene* 23, no. 55 (November 2004): 8979–8986.
11. F. Guillaumond, G. Bidaut, M. Ouaisi, et al., "Cholesterol Uptake Disruption, in Association With Chemotherapy, Is a Promising Combined Metabolic Therapy for Pancreatic Adenocarcinoma," *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 8 (February 2015): 2473–2478.
12. J. Li, Q. Li, Z. Su, et al., "Lipid Metabolism Gene-Wide Profile and Survival Signature of Lung Adenocarcinoma," *Lipids in Health and Disease* 19, no. 1 (October 2020): 222.
13. X. Shi, Y. Zheng, L. Jiang, et al., "EWS-FLI1 Regulates and Cooperates With Core Regulatory Circuitry in Ewing Sarcoma," *Nucleic Acids Research* 48, no. 20 (November 2020): 11434–11451.

14. C. Cheng, F. Geng, X. Cheng, and D. Guo, "Lipid Metabolism Reprogramming and Its Potential Targets in Cancer," *Cancer Communications (London)* 38, no. 1 (May 2018): 27.
15. D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell* 144, no. 5 (March 2011): 646–674.
16. C. Ding, Z. Shan, M. Li, H. Chen, X. Li, and Z. Jin, "Characterization of the Fatty Acid Metabolism in Colorectal Cancer to Guide Clinical Therapy," *Molecular Therapy - Oncolytics* 20 (March 2021): 532–544.
17. F. Shao, H. Mao, T. Luo, Q. Li, L. Xu, and Y. Xie, "HPGDS Is a Novel Prognostic Marker Associated With Lipid Metabolism and Aggressiveness in Lung Adenocarcinoma," *Frontiers in Oncology* 12 (2022): 894485.
18. H. Yang, D. Liu, C. Zhao, et al., "Mendelian Randomization Integrating GWAS and eQTL Data Revealed Genes Pleiotropically Associated With Major Depressive Disorder," *Translational Psychiatry* 11, no. 1 (April 2021): 225.
19. Y. Wu, Z. Wang, Y. Yang, et al., "Exploration of Potential Novel Drug Targets and Biomarkers for Small Cell Lung Cancer by Plasma Proteome Screening," *Frontiers in Pharmacology* 14 (2023): 1266782.
20. Z. Huang, H. Zheng, H. Wang, H. Ning, A. Che, and C. Cai, "Identification of Potential Therapeutic Targets for Breast Cancer Using Mendelian Randomization Analysis and Drug Target Prediction," *Environmental Toxicology* (April 2024).
21. J. Xin, D. Gu, S. Chen, et al., "SUMMER: A Mendelian Randomization Interactive Server to Systematically Evaluate the Causal Effects of Risk Factors and Circulating Biomarkers on Pan-Cancer Survival," *Nucleic Acids Research* 51, no. D1 (January 2023): D1160–D1167.
22. W. Hu and Y. Xu, "Transcriptomics in Idiopathic Pulmonary Fibrosis Unveiled: A New Perspective From Differentially Expressed Genes to Therapeutic Targets," *Frontiers in Immunology* 15 (2024): 1375171.
23. S. Burgess, C. N. Foley, E. Allara, J. R. Staley, and J. M. M. Howson, "A Robust and Efficient Method for Mendelian Randomization With Hundreds of Genetic Variants," *Nature Communications* 11, no. 1 (January 2020): 376.
24. G. Davey Smith and G. Hemani, "Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies," *Human Molecular Genetics* 23, no. R1 (September 2014): R89–R98.
25. F. Wang, D. Liu, Y. Zhuang, et al., "Mendelian Randomization Analysis Identified Genes Potentially Pleiotropically Associated With Periodontitis," *Saudi Journal of Biological Sciences* 28, no. 7 (July 2021): 4089–4095.
26. Y. Lai, C. Lin, X. Lin, L. Wu, Y. Zhao, and F. Lin, "Identification and Immunological Characterization of Cuproptosis-Related Molecular Clusters in Alzheimer's Disease," *Frontiers in Aging Neuroscience* 14 (2022): 932676.
27. B. Guvenc Paltun, S. Kaski, and H. Mamitsuka, "Machine Learning Approaches for Drug Combination Therapies," *Briefings in Bioinformatics* 22, no. 6 (November 2021): bbab293.
28. L. Shen, H. Huang, J. Li, et al., "Exploration of Prognosis and Immunometabolism Landscapes in ER+ Breast Cancer Based on a Novel Lipid Metabolism-Related Signature," *Frontiers in Immunology* 14 (2023): 1199465.
29. Y. Wang, J. D. McKay, T. Rafnar, et al., "Rare Variants of Large Effect in BRCA2 and CHEK2 Affect Risk of Lung Cancer," *Nature Genetics* 46, no. 7 (2014): 736–741.
30. J. Sun, J. Zhao, F. Jiang, et al., "Identification of Novel Protein Biomarkers and Drug Targets for Colorectal Cancer by Integrating Human Plasma Proteome With Genome," *Genome Medicine* 15, no. 1 (September 2023): 75.
31. A. Passaro, M. Al Bakir, E. G. Hamilton, et al., "Cancer Biomarkers: Emerging Trends and Clinical Implications for Personalized Treatment," *Cell* 187, no. 7 (March 2024): 1617–1635.
32. X. P. Duan, B. D. Qin, X. D. Jiao, K. Liu, Z. Wang, and Y. S. Zang, "New Clinical Trial Design in Precision Medicine: Discovery, Development and Direction," *Signal Transduction and Targeted Therapy* 9, no. 1 (March 2024): 57.
33. Y. Chen, Y. Zhou, R. Ren, Y. Chen, J. Lei, and Y. Li, "Harnessing Lipid Metabolism Modulation for Improved Immunotherapy Outcomes in Lung Adenocarcinoma," *Journal for Immunotherapy of Cancer* 12, no. 7 (July 2024): e008811.
34. J. Liu, T. Tian, X. Liu, and Z. Cui, "BCHE as a Prognostic Biomarker in Endometrial Cancer and Its Correlation With Immunity," *Journal of Immunology Research* 2022 (2022): 6051092.
35. I. Wessler, C. J. Kirkpatrick, and K. Racke, "Non-Neuronal Acetylcholine, a Locally Acting Molecule, Widely Distributed in Biological Systems: Expression and Function in Humans," *Pharmacology & Therapeutics* 77, no. 1 (January 1998): 59–79, [https://doi.org/10.1016/s0163-7258\(97\)00085-5](https://doi.org/10.1016/s0163-7258(97)00085-5).
36. R. Kumar, S. Razab, K. Prabhu, S. Ray, and B. Prakash, "Serum Butyrylcholinesterase and Zinc in Breast Cancer," *Journal of Cancer Research and Therapeutics* 13, no. 2 (2017): 367–370.
37. Y. Gu, M. J. Chow, A. Kapoor, et al., "Biphasic Alteration of Butyrylcholinesterase (BChE) During Prostate Cancer Development," *Translational Oncology* 11, no. 4 (August 2018): 1012–1022.
38. M. F. Montenegro, F. Ruiz-Espejo, F. J. Campoy, et al., "Cholinesterases Are Down-Expressed in Human Colorectal Carcinoma," *Cellular and Molecular Life Sciences* 63, no. 18 (September 2006): 2175–2182.
39. J. R. Mauney, T. Nguyen, K. Gillen, C. Kirker-Head, J. M. Gimble, and D. L. Kaplan, "Engineering Adipose-Like Tissue In Vitro and In Vivo Utilizing Human Bone Marrow and Adipose-Derived Mesenchymal Stem Cells With Silk Fibroin 3D Scaffolds," *Biomaterials* 28, no. 35 (2007): 5280–5290.
40. Z. Wu, J. H. Jeong, C. Ren, et al., "Fatty Acid-Binding Protein 4 (FABP4) Suppresses Proliferation and Migration of Endometrial Cancer Cells via PI3K/Akt Pathway," *Oncotargets and Therapy* 14 (2021): 3929–3942.
41. J. Zeng, E. R. Sauter, and B. Li, "FABP4: A New Player in Obesity-Associated Breast Cancer," *Trends in Molecular Medicine* 26, no. 5 (2020): 437–440.
42. J. Yang, S. Liu, Y. Li, et al., "FABP4 in Macrophages Facilitates Obesity-Associated Pancreatic Cancer Progression via the NLRP3/IL-1 β Axis," *Cancer Letters* 575 (October 2023): 216403.
43. M. I. Sigurdsson, N. Jamshidi, E. Steingrimsdottir, I. Thiele, and B. O. Palsson, "A Detailed Genome-Wide Reconstruction of Mouse Metabolism Based on Human Recon 1," *BMC Systems Biology* 4 (October 2010): 140.
44. S. McNeal, P. Bitterman, J. M. Bahr, et al., "Association of Immunosuppression With DR6 Expression During the Development and Progression of Spontaneous Ovarian Cancer in Laying Hen Model," *Journal of Immunology Research* 2016 (2016): 6729379.
45. S. Clay, J. Alladina, N. P. Smith, et al., "Gene-Based Association Study of Rare Variants in Children of Diverse Ancestries Implicates TNFRSF21 in the Development of Allergic Asthma," *Journal of Allergy and Clinical Immunology* 153, no. 3 (March 2024): 809–820.
46. G. Pan, J. H. Bauer, V. Haridas, et al., "Identification and Functional Characterization of DR6, a Novel Death Domain-Containing TNF Receptor," *FEBS Letters* 431, no. 3 (July 1998): 351–356.
47. S. Xu, L. Zhao, A. Larsson, and P. Venge, "The Identification of a Phospholipase B Precursor in Human Neutrophils," *FEBS Journal* 276, no. 1 (2009): 175–186.
48. M. Vanhaverbeke, M. Vausort, D. Veltman, et al., "Peripheral Blood RNA Levels of QSOX1 and PLBD1 Are New Independent Predictors of Left Ventricular Dysfunction After Acute Myocardial Infarction," *Circulation: Genomic and Precision Medicine* 12, no. 12 (December 2019): e002656.

49. X. Cao, B. Wu, Y. Hou, and J. Chen, “Lipid Metabolism-Related Gene Signatures for Predicting the Prognosis of Lung Adenocarcinoma,” *Translational Cancer Research* 12, no. 8 (August 2023): 2099–2114.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.