frontiers | Research Topics

Protein modifications in epigenetic dysfunctional diseases: Mechanisms and potential therapeutic strategies

Edited by
Bin Liu, Wei Liu and Lan Zhu

Published in
Frontiers in Genetics
Frontiers in Cell and Developmental Biology

ISSN 1664-8714
ISBN 978-2-8325-2558-6
DOI 10.3389/978-2-8325-2558-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Protein modifications in epigenetic dysfunctional diseases: Mechanisms and potential therapeutic strategies

Topic editors

Bin Liu — Jiangsu Ocean Universiity, China

Wei Liu — Arizona State University, United States

Lan Zhu — Medical College of Wisconsin, United States

Citation

Liu, B., Liu, W., Zhu, L., eds. (2023). *Protein modifications in epigenetic dysfunctional diseases: Mechanisms and potential therapeutic strategies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2558-6

Table of contents

frontiers | Frontiers in Cell and Developmental Biology

TYPE Editorial
PUBLISHED 15 May 2023
DOI 10.3389/fcell.2023.1216637

OPEN ACCESS

EDITED AND REVIEWED BY
Michael E. Symonds,
University of Nottingham,
United Kingdom

*CORRESPONDENCE
Wei Liu,
w.liu@asu.edu
Bin Liu,
liubin@jou.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 04 May 2023
ACCEPTED 09 May 2023
PUBLISHED 15 May 2023

CITATION
Ji J, Jing A, Geng T, Ma X, Liu W and Liu B (2023), Editorial: Protein modifications in epigenetic dysfunctional diseases: mechanisms and potential therapeutic strategies.
Front. Cell Dev. Biol. 11:1216637.
doi: 10.3389/fcell.2023.1216637

Editorial: Protein modifications in epigenetic dysfunctional diseases: mechanisms and potential therapeutic strategies

Jing Ji[1†], Aixin Jing[1†], Ting Geng[1†], Xinhui Ma[1†], Wei Liu[2*] and Bin Liu[1*]

[1]Key Laboratory of Marine Pharmaceutical Compound Screening, College of Pharmacy, Jiangsu Ocean University, Lianyungang, China, [2]School of Molecular Sciences and Biodesign Center for Applied Structural Discovery, Arizona State University, Tempe, AZ, United States

KEYWORDS

DNA methylation, post-translational modifications (PMTs), epigenetic regulation, epigenetic mechanism, epigenetic-related disorders

Editorial on the Research Topic
Protein modifications in epigenetic dysfunctional diseases: mechanisms and potential therapeutic strategies

Introduction

Epigenetic regulation refers to the chemical modifications that affect the chromatin state and regulate gene expression without altering the DNA sequence. It consists of various molecular mechanisms, including DNA methylation, histone modifications, chromosomal remodeling, and non-coding RNA regulation. Abnormal epigenetic regulation is closely associated with the occurrence and development of many diseases (Tsankova et al., 2007; Orioli and Dellambra, 2018; Lu et al., 2020), with DNA methylation and post-translational modifications (PTMs) being important mechanisms. DNA methylation is a chemical modification in which a methyl group (CH3) is covalently attached to cytosine nucleotides in DNA. This modification can silence genes and affect gene expression. Abnormal DNA methylation in many diseases leads to changes in gene expression levels, affecting cell growth, differentiation, and apoptosis, ultimately promoting disease development (Greenberg and Bourc'his, 2019). Post-translational modifications refer to the process by which the structure and function of a protein molecule are changed through chemical modifications after protein synthesis. These modifications include phosphorylation, ubiquitination, and SUMOylation, among others. Abnormal post-translational modifications in many diseases lead to changes in protein structure and function, affecting cell metabolism, signaling, and apoptosis, among other processes, ultimately promoting disease development (Vucic et al., 2011). Therefore, in-depth research on the role of abnormal epigenetic regulation in diseases can help to deepen our understanding of the mechanisms underlying disease occurrence and development, providing new ideas and methods for the prevention and treatment of related diseases.

This Research Topic aims to uncover new insights into the roles of DNA methylation and PTMs involved in multiple epigenetic-related disorders, including tumors, ischemic stroke, atherosclerosis, and coronary heart disease (CHD). Such discoveries hold the potential to elucidate the underlying mechanisms driving the onset, progression, and therapeutic opportunities for these diseases.

DNA methylation in epigenetic-related disorders

DNA methylation is a widely studied field in cancer research due to its significant role in the development and progression of the disease. DNA methylation can promote or inhibit cancer development by altering gene expression levels. Specific subtypes of cancer have demonstrated distinct DNA methylation patterns, which can serve as biomarkers for diagnosis or predicting treatment response. Zeng et al. conducted a comprehensive review of studies investigating the relationship between DNA methylation and the etiology, diagnosis, treatment, and prognosis of gastric cancer. Huang et al. created a gene signature and regulatory network related to enhancers, utilizing transcriptome and methylation data for predicting the survival of patients with lung adenocarcinoma.

CHD, characterized by myocardial ischemia and myocardial infarction, arises from inadequate blood supply to the coronary arteries. Abnormal DNA methylation has been shown to affect the expression of genes involved in vascular function regulation, inflammatory factors, and blood clotting (Xia et al., 2021). These aberrant DNA methylation changes may exacerbate the development and progression of CHD. Zhao et al. proposed the use of blood-based F2RL3 methylation as a potential biomarker for CHD, particularly in older individuals or those with a history of myocardial infarction. Combining F2RL3 methylation with conventional risk factors could be an effective approach for early-stage evaluation of CHD.

Ischemic stroke occurs when the brain receives insufficient blood supply, resulting in tissue damage. Research indicates a close relationship between abnormal DNA methylation patterns and the occurrence and prognosis of stroke. Altered DNA methylation patterns can influence the expression of stroke-related genes, such as neuroprotective factors and inflammatory factors, thereby affecting brain tissue damage and repair processes (Choi et al., 2022). Sun et al. presented findings of genome-wide alterations in DNA methylation in patients with ischemic stroke compared to controls. They identified 462 functional differentially methylated positions (DMPs) corresponding to 373 annotated genes. Notably, hypomethylated sites were eight times more abundant than hypermethylated sites in ischemic stroke cases, highlighting the predominant role of hypomethylation in this condition.

Atherosclerosis, a chronic inflammatory disease characterized by the buildup of lipid-rich plaques in arterial walls. Aberrant DNA methylation patterns have been observed in various cell types implicated in atherosclerosis, including endothelial cells, smooth muscle cells, and immune cells (Dai et al., 2022). Xu et al. conducted an analysis of GEO data obtained from 15 atherosclerotic and paired healthy tissues. They systematically screened the entire genome and identified a total of 110,695 differentially methylated sites (DMPs) and 918 differentially methylated regions (DMRs). Additionally, they discovered six genes exhibiting significant methylation differences in the CpG islands of the promoter regions, encompassing 49 DMPs. Notably, they also observed a notable increase in monocyte infiltration levels within atherosclerotic (AS) tissues. These findings provide valuable insights into potential DNA methylation-related biomarkers and shed light on the involvement of monocytes in early-stage atherosclerosis.

PTMs in epigenetic-related disorders

As an important form of PTMs, SUMOylation modifications play a critical role in various biological processes, including the development and progression of cancer (Han et al., 2018). Chen et al. investigated the SUMOylation patterns in lung adenocarcinoma by employing unsupervised consensus clustering based on the expression of SUMOylation regulatory genes. Their findings demonstrated that these SUMOylation patterns effectively differentiate the tumor microenvironment characteristics of lung adenocarcinoma, particularly the status of immune cell infiltration. Additionally, they developed a SUMOylation score that enables the assessment of the relationship between SUMOylation and immune cell crosstalk. This score holds significant prognostic value and can be utilized to predict the response to immunotherapy and chemotherapy in patients with lung adenocarcinoma.

Conclusion

The present Research Topic collected some interesting papers and revealed better understanding the mechanisms of multiple epigenetic-related disorders. We hope further researches about this Research Topic will be continued, and will contribute to the clinical transformation.

Author contributions

JJ, AJ, XM, TG wrote the first manuscript, JJ, WL, BL reviewed and revised the manuscript and wrote the final manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the National Natural Science Foundation of China (82273167 and 82104174), Jiangsu Province Basic Research Program Natural Science Foundation (Outstanding

Youth Fund Project, BK20220063), the Key Program of Basic Science (Natural Science) of Jiangsu Province (22KJA350001), "Huaguo Mountain Talent Plan" of Lianyungang City (Innovative Talents BL), Qing Lan Project of Jiangsu Universities (Outstanding young backbone teachers, JJ), Jiangsu Key Laboratory of Marine Pharmaceutical Compound Screening (HY202103), Priority Academic Program Development of Jiangsu Higher Education Institutions.

Acknowledgments

We extend our deepest gratitude to all the authors and reviewers who contributed to this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Choi, D. H., Choi, I. A., and Lee, J. (2022). The role of DNA methylation in stroke recovery. *Int. J. Mol. Sci.* 23 (18), 10373. doi:10.3390/ijms231810373

Dai, Y., Chen, D., and Xu, T. (2022). DNA methylation aberrant in atherosclerosis. *Front. Pharmacol.* 13, 815977. doi:10.3389/fphar.2022.815977

Greenberg, M. V., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 20 (10), 590–607. doi:10.1038/s41580-019-0159-6

Han, Z. J., Feng, Y. H., Gu, B. H., Li, Y. M., and Chen, H. (2018). The post-translational modification, SUMOylation, and cancer (Review). *Int. J. Oncol.* 52 (4), 1081–1094. doi:10.3892/ijo.2018.4280

Lu, Y., Chan, Y. T., Tan, H. Y., Li, S., Wang, N., and Feng, Y. (2020). Epigenetic regulation in human cancer: The potential role of epi-drug in cancer therapy. *Mol. cancer* 19, 79–16. doi:10.1186/s12943-020-01197-3

Orioli, D., and Dellambra, E. (2018). Epigenetic regulation of skin cells in natural aging and premature aging diseases. *Cells* 7 (12), 268. doi:10.3390/cells7120268

Tsankova, N., Renthal, W., Kumar, A., and Nestler, E. J. (2007). Epigenetic regulation in psychiatric disorders. *Nat. Rev. Neurosci.* 8 (5), 355–367. doi:10.1038/nrn2132

Vucic, D., Dixit, V. M., and Wertz, I. E. (2011). Ubiquitylation in apoptosis: A post-translational modification at the edge of life and death. *Nat. Rev. Mol. Cell Biol.* 12 (7), 439–452. doi:10.1038/nrm3143

Xia, Y., Brewer, A., and Bell, J. T. (2021). DNA methylation signatures of incident coronary heart disease: Findings from epigenome-wide association studies. *Clin. Epigenetics* 13 (1), 186–216. doi:10.1186/s13148-021-01175-6

frontiers in Genetics

REVIEW
published: 04 March 2022
doi: 10.3389/fgene.2022.823905

DNA Methylation: An Important Biomarker and Therapeutic Target for Gastric Cancer

*Yunqing Zeng[1], Huimin Rong[2], Jianwei Xu[3], Ruyue Cao[1], Shuhua Li[1], Yanjing Gao[1], Baoquan Cheng[1] and Tao Zhou[4]**

[1]Department of Gastroenterology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [2]Department of Reconstructive Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [3]Department of Pancreatic Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China, [4]Department of Geriatric Medicine, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China

Gastric cancer (GC) is a very common malignancy with a poor prognosis, and its occurrence and development are closely related to epigenetic modifications. Methylation of DNA before or during gastric cancer is an interesting research topic. This article reviews the studies on DNA methylation related to the cause, diagnosis, treatment, and prognosis of gastric cancer and aims to find cancer biomarkers to solve major human health problems.

Keywords: DNA methylation, gastric cancer, diagnosis, therapy, prognosis

OPEN ACCESS

Edited by:
Trygve Tollefsbol, University of Alabama at Birmingham, United States

Reviewed by:
Apiwat Mutirangura, Chulalongkorn University, Thailand
Fazlur Rahman Talukdar, International Agency For Research On Cancer (IARC), France

****Correspondence:***
Tao Zhou
zhoutao@sdu.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *28 November 2021*
Accepted: *07 February 2022*
Published: *04 March 2022*

Citation:
Zeng Y, Rong H, Xu J, Cao R, Li S, Gao Y, Cheng B and Zhou T (2022) DNA Methylation: An Important Biomarker and Therapeutic Target for Gastric Cancer. Front. Genet. 13:823905. doi: 10.3389/fgene.2022.823905

1 INTRODUCTION

Gastric cancer (GC) is a major unresolved clinical problem, with over a million new cases globally in 2020 (Sexton et al., 2020). It is the fourth most common cancer in men (Sexton et al., 2020), and the third most frequent cause of cancer-related deaths worldwide (Global, 2019). Most patients are diagnosed at an advanced stage and have a poor prognosis (Digklia and Wagner, 2016). Recently, multiple studies showed that epigenetic dysregulations, including DNA methylation, histone post-translational modifications, chromatin remodeling and non-coding RNAs, play a vital role in the oncogenesis and progression of GC. Among the epigenetic modifications mentioned, DNA methylation is the earliest known and well-investigated epigenetic change (Skvortsova et al., 2019).

The types of aberrant DNA methylation in human cancers include global DNA hypomethylation and local hypermethylation of genes. Genome-wide hypomethylation mainly occurs in repetitive elements that are normally hypermethylated to maintain genomic integrity and stability. Long Interspersed Nucleotide Element 1 (LINE-1), Alu repetitive elements and human endogenous retroviruses (HERVs) are the major constituents of interspersed repetitive sequences (IRS) (Chansangpetch et al., 2018). Regional hypermethylation of genes occurs in CpG (5′-cytosine-phosphate-guanine-3′) islands, which are normally unmethylated and cause silencing of tumor-suppressor genes, cell cycle regulator genes, and DNA repair genes (Eyvazi et al., 2019; Moore et al., 2013; Shao et al., 2018; Zeng et al., 2017). CpG islands (CGI) are frequently found in mammalian promoters (Field et al., 2018). Three types of DNA methyltransferases including DNMT1, DNMT3A, and DNMT3B, are responsible for DNA methylation. DNMT3A and DNMT3B are primarily ab initio methyltransferases, while DNMT1 maintains the methylation of symmetrically methylated CpGs during DNA duplication (Ebrahimi et al., 2020; Zeng et al., 2017).

Although many studies explored the prospect of DNA methylation as a biomarker with the aim of decreasing GC deaths, the methylation levels in those studies were mainly detected from the tissues by invasive methods. There are still many aberrantly methylated genes whose roles in GC have not

been fully investigated, especially those detected by non-invasive methods. Novel non-invasive biomarkers are necessary for early detection, the prediction of prognosis and recurrence, and the evaluation of treatment efficacy. In this review, from the perspective of clinical practicality, we briefly described DNA methylation associated with pathogens of GC. Then, we highlighted the value of aberrant DNA methylation as a non-invasive biomarker for GC management.

2 DNA METHYLATION ASSOCIATED WITH PATHOGENESIS

Pathogens invade host cells and cause epigenetic changes, such as DNA methylation, making it a safer environment for the pathogen. This allows the infection to persist and promotes the development of GC (Fattahi et al., 2018). The most important pathogens associated with gastric carcinogenesis are *Helicobacter pylori* (*H. pylori*, HP) and the Epstein-Barr virus (EBV).

2.1 Infection Mediated by *Helicobacter pylori*

Infection by *H. pylori* induces hypermethylation in the promoter regions of many DNA repair genes and tumor suppressor genes (MHL1, RUNX 3, APC, and PTEN), thus silencing the genes and facilitating carcinogenesis (Muhammad et al., 2019). Kosumi et al. found that LINE-1 hypomethylation of non-cancerous gastric mucosae in gastric cancer patients was significantly correlated with *H. pylori* infection (p = 0.037) and prospectively confirmed the similar result in non-gastric cancer patients (p = 0.010) (Kosumi et al., 2015). Yoshida et al. found that compared to the gastric mucosae of *H. pylori*-negative healthy volunteers, the Alu methylation level was significantly lower in the gastric mucosae of *H. pylori*-positive healthy volunteers and *H. pylori*-positive gastric cancer patients (Yoshida et al., 2011).

When chronic inflammation, triggered by *H. pylori* infection in Mongolian gerbils, was repressed by cyclosporin A, aberrant DNA methylation was substantially suppressed; however, the abundance of *H. pylori* in the gastric mucosa was not reduced. Therefore, it was concluded that the inflammation, rather than *H. pylori*, was responsible for inducing abnormal DNA methylation (Niwa et al., 2013). However, recurrent inflammation caused by alcohol or saturated NaCl did not induce abnormal DNA methylation (Hur et al., 2011). *Helicobacter pylori* infection activates the secretion of IL-1β and TNF-α and the production of reactive oxygen species. Together, they induce DNA methyltransferase 1 (DNMT1) and cause aberrant DNA methylation in gastric epithelial cells (Kim, 2019). Aggregation of aberrantly methylated DNA in the gastric mucosa might favor cancerogenesis (Maeda et al., 2017).

Although eliminating *H. pylori* significantly decreases methylation of tumor suppressor genes, DNA methylation does not return to the same level as that in individuals who are never infected by *H. pylori,* and the higher levels of methylated DNA in the previously infected individuals have adverse effects on the gastric mucosa in the long term (Nakajima et al., 2010). Therefore, individuals with ongoing presence of aberrant DNA methylation would face a higher risk of GC even after the eradication of *H. pylori* (Shin et al., 2012). From this perspective, determination of DNA methylation in *H. pylori*-negative subjects, including subjects whose *H. pylori* has been eliminated, can also act as a helpful diagnostic biomarker for assessing the risk of gastric cancer (Tahara and Arisawa, 2015).

2.2 EBV Infection

Based on The Cancer Genome Atlas (TCGA) project, gastric cancer was classified into four molecular subtypes: Epstein-Barr virus, microsatellite instability, genomically stable, and chromosomal instability (Sohn et al., 2017). Epstein-Barr virus-associated gastric carcinoma (EBVaGC), which comprises nearly 10% of gastric carcinomas (Fukayama and Ushiku, 2011), shows the most extreme DNA hypermethylation in all human malignancies (Usui et al., 2021). Matsusaka et al. divided GC into three epigenotypes according to DNA methylation patterns: EBV−/low methylation, EBV−/high methylation, and EBV+/high methylation (Matsusaka et al., 2011). CXXC4, TIMP2, and PLXND1 genes are methylated only in the EBV + tumors. COL9A2, EYA1, and ZNF365 genes are methylated both in EBV+ and EBV−/high tumors. AMPH, SORCS3, and AJAP1 genes are methylated in all gastric cancers (Matsusaka et al., 2011). Many studies have shown that promoter hypermethylation plays a crucial role in the carcinogenesis of EBVaGC (Kaneda et al., 2012; Matsusaka et al., 2014; Okada et al., 2013; Yau et al., 2014; Zong and Seto, 2014). EBVaGC exhibited promoter hypermethylation in multiple genes (e.g., p15, p16, p73, hMLH1, MGMT, GSTP1, CDH1, TIMP1, TIMP3, DAPK, bcl-2, APC, PTEN, and RASSF1A) associated with regulation of cell proliferation (Shinozaki-Ushiku et al., 2015). In contrast, the relationship between EBV infection and demethylation in interspersed DNA repeats is unclear.

Unlike HP infection, EBV infection-induced hypermethylation is considered to be caused not by the intermediate pathways related to inflammation induced by infection but by the pathogen itself (Kaneda et al., 2012). The molecular mechanism of aberrant DNA methylation triggered by EBV infection remains unclear. The proposed mechanism includes the upregulation of the expression of DNMT1 and DNMT3b and the downregulation of the activity of TET2 demethylase by EBV (Hino et al., 2009; Namba-Fukuyo et al., 2016; Zhao et al., 2013).

3 DIAGNOSTIC BIOMARKERS

Exploring the role of DNA methylation in early diagnosis of gastric cancer is important for reducing mortality. Most investigations of DNA methylation are based on the evaluation of disparities in methylation levels between the tumor and the adjacent tissues. Many studies have found that repetitive element hypomethylation and site-specific CpG island

TABLE 1 | Aberrant DNA methylation as diagnostic biomarkers in body fluids and stool of GC patients.

Study	Source	Hypermethylated gene (sensitivity; specificity)	Methods	References
Watanabe, 2009	Gastric wash	MINT25 (90.0%; 95.8%) RORA (60.0%; 85.4%) GDNF (65.0%; 89.6%) ADAM23 (70.0%; 83.3%) PRDM5 (65.0%; 93.7%) MLF1 (60.0%; 85.4%)	Pyrosequencing	Watanabe et al. (2009)
Yamamoto, 2016	Gastric juice-derived exosomes	BARHL2 (90.0%; 100%)	Pyrosequencing	Yamamoto et al. (2016)
Zhang, 2014	Whole blood	SPG20 (48.8%; 100%)	MSP	Zhang et al. (2014)
Liu, 2015	Serum	SFRP1 (30.95%; 93.2%)	MSP	Liu et al. (2015)
Zheng, 2011a	Serum	KCNA4 (67.4%; 97.4%) CYP26B1 (73.9%; 93.4%)	Q-MSP	Zheng et al. (2011a)
Lee, 2002	Serum	p15 (55.6%; 100%)	MSP	Lee et al. (2002)
Kanyama, 2003	Serum	p16 (26%; 100%)	MSP	Kanyama et al. (2003)
Abbaszadegan, 2008	Serum	p16 (26.9%; 100%)	MSP	Abbaszadegan et al. (2008)
Wang, 2008	Serum	RASSF1A (34%; 100%)	MSP	Wang et al. (2008)
Lu, 2012	Serum	RUNX3 (94.1%; 100%)	Q-MSP	Lu et al. (2012)
Chen, 2009	Serum	HSULF1 (55%; 81%)	MSP	Chen et al. (2009)
Hibi, 2011	Serum	TFPI2 (10%; 100%)	Q-MSP	Hibi et al. (2011)
Leung, 2005	Serum	TIMP3 (17%; 100%) APC (17%; 100%) E-cadherin (13%; 100%) hMLH1 (41%; 92%)	Q-MSP	Leung et al. (2005)
Lee, 2002	Serum	E-cadherin (57.4%; 100%) DAPK (48.1%; 100%) GSTP1 (14.8%; 100%)	MSP	Lee et al. (2002)
Xue, 2016	Serum	RASSF10 (81.71%; 89.5%)	BSP	Xue et al. (2016)
Wang, 2015	Serum	FLNC (67.1%; 93.0%) THBS1 (63.4%; 94.2%) UCHL1 (56.1%; 89.5%) DLEC1 (80.5%; 93.0%)	Q-MSP	Wang et al. (2015)
Balgkouranidou, 2013	Serum	SOX17 (58.9%; 100%)	MSP	Balgkouranidou et al. (2013)
Guo, 2011	Plasma	HLTF (20.8%; 100%)	MSP	Guo et al. (2011)
Bernal, 2008	Plasma	Reprimo (95.3%; 90.3%)	MSP	Bernal et al. (2008)
Alarcon, 2020	Plasma	RPRML (56.0%; 88.0%)	MethyLight assay	Alarcon et al. (2020)
Miao, 2020	Plasma	SFRP2 (60.9%; 86.0%)	Q-MSP	Miao et al. (2020)
Cheung, 2012	Plasma	RNF180 (56%; 100%)	Q-MSP	Cheung et al. (2012)
Ng, 2011	Plasma	SLC19A3 (85%; 85%)	Q-MSP	Ng et al. (2011)
Pimson, 2016	Plasma	PCDH10 (94.1%; 97.03%)	MSP	Pimson et al. (2016)
Guo, 2010	Plasma	IRX1 (73.3%; 90%)	MSP	Guo et al. (2010)
Lee, 2013	Plasma	SEPT9 (17.7%; 90.6%)	PCR	Lee et al. (2013)
Chen, 2015b	Plasma	ZIC1 (60.6%; 100%)	MSP	Chen et al. (2015b)
Guo, 2021	Stool	SDC2 (Train set: 40.9%, test set: 40.9%; Train set: 93.3%; test set: 91.7%) TERT (Train set: 36.4%, test set: 34.1%; Train set: 90.0%; test set: 91.7%) RASSF2 (Train set: 31.8%; Train set: 93.3%) SFRP2 (Train set: 22.7%; Train set: 90.0%)	PCR	Guo et al. (2021)
Liu, 2017	Stool	TERT (54.3%; 90%)	PCR	Liu et al. (2017)

Abbreviations: MSP, methylation-specific PCR; q-MSP, Quantitative methylation-specific PCR; BSP, bisulfite sequencing PCR.

promoter hypermethylation are associated with increased risk of GC (Hu X et al., 2021; Min et al., 2017; Saito et al., 2012). In such cases, surgery is required to access the affected tissue, which is a major limitation for clinical application. Growing evidence has shown that methylation-related alterations in cancer patients arise systematically and might be measured in surrogate tissues (Yuasa, 2010). Developing non-invasive detection techniques is quite important for GC patients. Therefore, increasing researchers are devoted to exploring the clinical significance of aberrant DNA methylation detected in body fluids (including peripheral blood, gastric juice, etc.) and stool. Methylation of tumor suppressor genes in peripheral blood has been studied most extensively. In contrast, non-invasive tests related to genome-wide hypomethylation are performed less frequently. Aberrant methylation of multiple genes in body fluids and stool could be a valuable non-invasive biomarker for the early screening and diagnosis of gastric cancer (Qu et al., 2013; Yamamoto et al., 2020) (**Table 1**) (**Figure 1**).

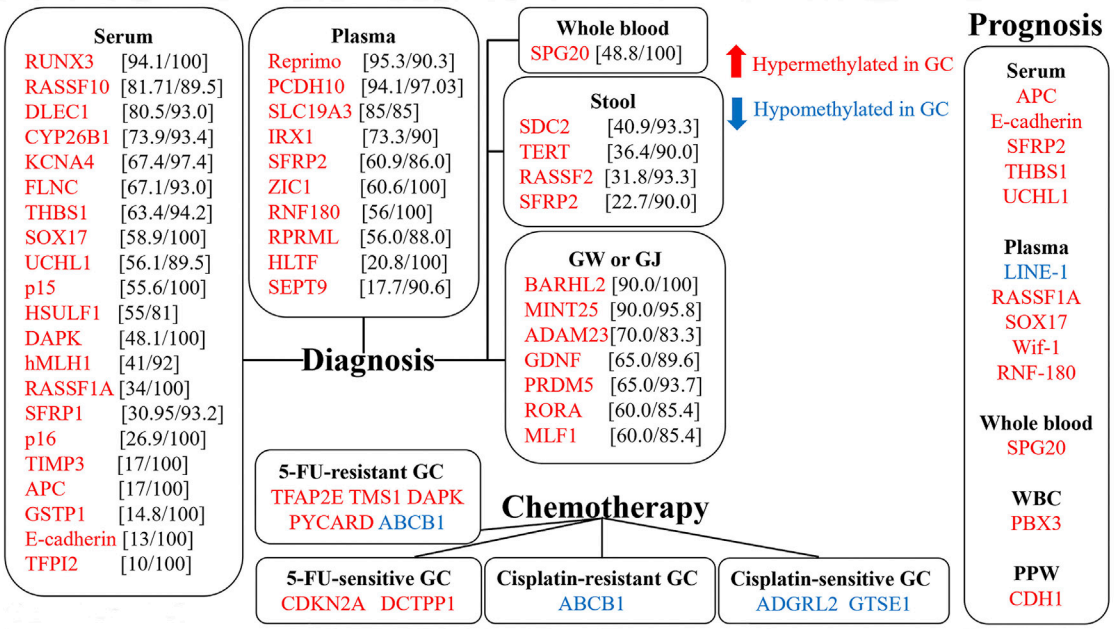


FIGURE 1 | Biomarkers for diagnosis, therapy, and prognosis of gastric cancer. Hypermethylated (in red) and Hypomethylated (in blue) genes are shown. Sensitivity and specificity are shown in square brackets. GW: gastric washes. GJ: gastric juice. WBC: white blood cell. PPW: preoperative peripheral washes.

3.1 Biomarkers in Peripheral Blood

Various blood biomarker-based tests can be used for the early diagnosis of cancer (Huang et al., 2021). Tumor cells can release DNA into peripheral blood, causing abundant circulating DNA levels in the blood of cancer patients to be several-fold higher than that of individuals without cancer (Tahara and Arisawa, 2015). Increasing evidence indicates that the detection of methylated DNA in peripheral blood is more well-developed and stable than detection of gene mutation (Qi et al., 2016). Testing DNA methylation in the blood as a risk marker for carcinoma is of special significance, since the non-invasive and convenient collection of peripheral blood DNA is easily accepted by patients.

3.1.1 Circulating Cell-free DNA Methylation

Circulating cell-free DNA (cfDNA), derived from both normal and tumor cells, is present in the blood. In particular, the cfDNA that is derived from tumors and possesses tumor-specific mutations is called circulating tumor DNA (ctDNA) (Pessoa et al., 2020). Numerous studies have investigated the feasibility of measuring serum or plasma DNA methylation to detect methylation of tumor-derived circulating DNA as a latent diagnostic biomarker for gastric cancer (Tahara and Arisawa, 2015). Methylation of p16, CDH1, RARβ, Reprimo, Rassf1A, hMLH1, RUNX3, APC, E-cadherin, SFRP1, KCNA4, p15, SFRP2, HSULF1, PCDH10, TFPI2, TIMP3, CYP26B1, DAPK, DLEC1, FLNC, THBS1, UCHL1, GSTP1, HLTF, RPRML, SLC19A3, RNF180, etc. are markedly higher in the DNA from peripheral blood of GC subjects than in that of control subjects (Wang et al., 2015; Abbaszadegan et al., 2008; Alarcón et al., 2020; Balgkouranidou et al., 2013; Bernal et al., 2008; Chen et al., 2015a; Chen et al., 2009; Cheng et al., 2007; Cheung et al., 2012; Guo et al., 2010, 2011,; Hibi et al., 2011; Huang et al., 2021; Ikoma et al., 2006; Kanyama et al., 2003; Lee et al., 2002; Leung et al., 2005; Liu et al., 2015; Lu et al., 2012; Miao et al., 2020; Ng et al., 2011; Pimson et al., 2016; Sakakura et al., 2009; Tahara et al., 2013; Wang et al., 2008; Xue et al., 2016; Zhang et al., 2010, 2014; Chen et al., 2015b; Zheng et al., 2011b). Lin et al. evaluated the methylation state of three genes (ZIC1, HOXD10, and RUNX3) from the blood samples of GC patients using methylation-specific polymerase chain reaction. They discovered that the Odds ratios of ZIC1, HOXD10, and RUNX3 methylation for predicting GC were 4.285 (95%CI: 2.435–7.542), 3.133 (95%CI: 1.700–5.775), and 2.674 (95%CI: 1.441–4.960), respectively. The joint detection sensitivity of these three genes was 91.6%. Therefore, the combined detection of multiple gene promoter hypermethylation exhibited a cooperative effect compared to a single biomarker used to predict GC (Lin et al., 2017).

What is noteworthy is that even for the same methylated gene, there are significant discrepancies between the results of different studies, which might be attributed to differences in the sample size, detection methods, and study regions (Huang et al., 2021). To understand this heterogeneity and evaluate the accuracy of DNA methylation markers in the blood for identifying gastric cancer patients, Hu et al. conducted a meta-analysis with 32 studies, containing 69 analyses of blood DNA methylation tests that were conducted to evaluate GC. The 32 studies included 2,098 GC patients and 2,098 control subjects. The blood test based on DNA methylation had an overall sensitivity of 57% and specificity of 97% for gastric cancer. Plasma-based tests showed a sensitivity of 71% and specificity of 89%. Serum-based tests showed a sensitivity of 50% and specificity of 98%. The

sensitivity of using multiple methylation genes was 76% and specificity was 85%. These results suggested that the blood-based DNA methylation test has high specificity but moderate sensitivity for the detection of gastric cancer. The determination of various methylation genes or the use of plasma samples might improve the sensitivity of the diagnosis (Hu et al., 2017).

3.1.2 DNA Methylation in Peripheral Blood Leukocytes

Unlike tumor DNA, leukocyte DNA can be obtained non-invasively and relatively inexpensively (Tahara et al., 2018). Studies which determine whether selected tumor suppressor genes and genome-wide repetitive sequence methylation in peripheral blood leukocytes of subjects with gastric cancer and healthy controls are different are rapidly emerging. Hypermethylation of KIBRA, DLEC1, FAT4, WT1, H19, MALAT1, APC, ACIN1, BCL2, CD44, TNFRSF10C and RARB promoters in peripheral blood leukocytes was found to be statistically significant in GC patients (Dauksa et al., 2014; Hu D. et al., 2021; Sun et al., 2018; Xie et al., 2020; Zhang et al., 2018; Zhou et al., 2019). To date, only a few studies have examined GC risk associated with Alu and LINE-1 methylation in peripheral blood leukocytes and the results are variable. Dauksa et al. found that the mean methylation level in Alu and LINE-1 repeats of GC patients was slightly lower than the mean level in the controls (Dauksa et al., 2014). Hou et al. demonstrated that GC risk increased with a decrease in the methylation of Alu or LINE-1, although the trends were not statistically significant (Hou et al., 2010). Gao et al. found that Alu methylation in blood leukocyte DNA was inversely associated with GC risk, but LINE-1 methylation levels were not correlated with GC risk (Gao et al., 2012). Barchitta et al. also showed that the LINE-1 methylation levels were significantly different in tissue samples but not in blood samples (Barchitta et al., 2014). These results suggested that studies with more individuals must be performed to determine the clinical applicability of leukocyte DNA methylation to detect gastric cancer non-invasively.

3.1.3 DNA Methylation in Whole Blood

Several studies evaluated the association of aberrant DNA methylation with the risk of gastric cancer. SOCS3, SPG20, and SFRP1 promoter hypermethylation in whole blood significantly increased GC risk (Han et al., 2020; Tahara et al., 2013; Zhang et al., 2014).

3.2 Biomarkers in Gastric Washes

Since a large number of mucosal cells could be extracted from gastric juice (GJ)), DNA biomarkers in gastric juice might be used to detect gastric cancer. However, DNA is easily degraded in an acidic environment; thus, gastric wash (GW) is used as an alternative source for determining aberrant DNA methylation (Yamamoto et al., 2020). Unfortunately, only several early studies showed that the methylation levels of MINT25, RORA, GDNF, ADAM23, PRDM5, CDH1, and MLF1 in gastric washes of GC patients were significantly higher than those of control subjects (Muretto et al., 2008; Watanabe et al., 2009). Among them, MINT25 methylation has the optimal sensitivity (90.0%) and specificity (95.8%), and thus, can distinguish GC from non-GC and be a potential biomarker for screening GC. Yoshiyuki et al. investigated the relationship of the methylation levels between biopsy and gastric washes. The methylation levels of all six genes were tightly associated by statistical analysis (MINT25: $p = 0.001$; RORA: $p = 0.03$; PRDM5: $p < 0.001$; MLF1: $p < 0.001$; ADAM23 $p < 0.001$; GDNF: $p < 0.001$). These results indicated that DNA from gastric washes can be used as a suitable substitute for DNA from biopsied tissues to determine the accumulation of DNA methylation in GC patients (Watanabe et al., 2009).

Hypermethylation of BARHL2 was detected in gastric wash-derived and gastric juice-derived exosomal DNA in early-stage GC patients before endoscopic treatment, whereas methylation levels considerably decreased with a curative endoscopic therapy. These results indicated that BARHL2 methylation might contribute to the detection of residual cancer after endoscopic resection and the potential prediction of tumor relapse (Yamamoto et al., 2020). Some disturbing factors such as aging, HP infection, and chronic inflammation can also induce aberrant DNA methylation. BARHL2 methylation is not affected by those factors. Therefore, GW or GJ exoDNA-based methylation analysis of BARH2 is expected to be an accurate biomarker for detecting early and advanced gastric cancer (Yamamoto et al., 2016).

3.3 Biomarkers in Stool

Guo et al. (Guo et al., 2021) evaluated the feasibility of gene methylation in feces for screening gastric cancer. All GC patients and normal controls were divided into training sets and test sets. The sensitivity and specificity of a single marker for gastric cancer detection in the training set for SDC2 were 40.9 and 93.3%, for TERT were 36.4 and 90.0%, for RASSF2 were 31.8 and 93.3%, for SFRP2 were 22.7 and 90.0%, and for Hb were 27.3 and 90.0%. The sensitivity and specificity of the three markers for methylation of SDC2, TERT, and Hb in the test set for gastric cancer detection were 40.9 and 91.7%, 34.1 and 91.7%, and 25.0 and 86.7%, respectively. The results showed that the methylation of SDC2, SFRP2, TERT, and RASSF2 has a certain sensitivity and high specificity in GC screening, which is a potential fecal biomarker of gastric cancer. Another study (Liu et al., 2017) also suggested the feasibility of stool TERT promoter methylation analyses for the non-invasive screening of gastric cancer.

4 THERAPEUTIC TARGET

DNA methylation is also critical for the adjuvant treatment of gastric cancer. Chemotherapy is one of the major methods for treating GC. The main problem with chemotherapy is drug resistance, which is primarily related to DNA methylation. Correcting aberrant methylation patterns can improve chemotherapy response and patient survival (Housman et al., 2014). Animal studies have also shown that direct repression of aberrant DNA methylation can inhibit gastric carcinogenesis (Maeda et al., 2017). Therefore, DNMT inhibitors (DNMTi) are being actively investigated as novel cancer treatments. Additionally, adjuvant radiotherapy of GC has been debated

TABLE 2 | Aberrantly methylated genes as prognostic biomarkers in GC patients.

Study	Source	Aberrantly methylated gene	Prognosis	Methods	References
Balgkouranidou, 2015	Serum	APC	Poorer OS (HR = 4.6, 95% CI = 1.1–20.3, p = 0.046)	MSP	Balgkouranidou et al. (2015)
Ikoma, 2006	Serum	E-cadherin	Poorer 3-year survival rate (p < 0.05)	MSP	Ikoma et al. (2006)
Yan, 2021	Serum	SFRP2	Shorter PFS (HR = 13.05; 95% CI = 3.05–55.95) and OS (HR = 7.80; 95% CI = 1.81–33.60) (stage III); Shorter PFS (HR = 2.74; 95% CI = 1.58–4.78) and OS (HR = 3.14; 95% CI = 1.67–5.92) (stage IV)	PCR	Yan et al. (2021)
Hu, 2021	Serum	THBS1	Worse DFS (HR = 3.838; 95% CI = 1.691–8.710; p = 0.001)	Q-MSP	Hu X et al. (2021)
Wang, 2015	Serum	UCHL1	Worse OS (p = 0.033)	Q-MSP	Wang et al. (2015)
		THBS1	Worse OS (p = 0.0483)		
Ko, 2021	Plasma	LINE-1	Worse OS (p = 0.006)	qPCR	Ko et al. (2021)
Pimson, 2016	Plasma	RASSF1A	Lower OS (HR = 2.33, 95% CI = 1.14–4.85, p = 0.002)	MSP	Pimson et al. (2016)
Karamitrousis, 2021	Plasma	SOX17	Lower PFS and OS (p < 0.001)	MSP	Karamitrousis et al. (2021)
		Wif-1	Lower PFS and OS (p = 0.001)		
		RASSF1A	Lower PFS and OS (p = 0.004)		
Cheung, 2012	Plasma	RNF-180	Poorer OS (HR = 2.13; 95% CI = 1.11–4.08; p = 0.02)	Q-MSP	Cheung et al. (2012)
Zhang, 2014	Whole blood	SPG20	Shorter OS (p = 0.037)	MSP	Zhang et al. (2014)
Xie, 2020	Peripheral blood leukocytes	PBX3	Poorer cum survival (HR = 1.678, 95% CI = 1.046–2.693) (elderly group); Poorer cum survival (HR = 2.058, 95% CI = 1.024–4.137) (female group)	MS-HRM	Xie et al. (2020)
Yu, 2012	PPW	CDH1	Worse DFS (p < 0.000)	MSP	Yu et al. (2012)

Abbreviations: CI, confidence interval; HR, hazard ratio; OR, odds ratio; MS-HRM, methylation-sensitive high-resolution melting; PFI, progression-free interval; PFS, progression free survival; PPW, preoperative peritoneal washes.

over the past few decades. However, it has been suggested that hypermethylation and inactivation of certain genes associated with cell cycle regulation, DNA repair, apoptosis, and signal transduction can lead to radiotherapy resistance in GC cells (Zhang et al., 2006). In the future, it may be possible to improve radiotherapy response by altering DNA methylation patterns to benefit GC patients.

4.1 Chemotherapy

Studies have shown that DNA methylation in gastric cancer cells is related to the sensitivity of chemotherapy and resistance of anticancer agents such as 5-FU and cisplatin. The biomarkers used to identify resistance or sensitivity to chemotherapeutic drugs can be investigated. Hypermethylated TFAP2E, TMS1, PYCARD (ASC/TMS1), and DAPK might be appropriate biomarkers for 5-FU-resistant gastric cancer. Hypermethylated CDKN2A (p16INK4a) and DCTPP1 might be useful biomarkers for 5-FU-sensitive gastric cancer. Hypomethylated ADGRL2 (LPHN2) and GTSE1 are potential biomarkers of cisplatin-sensitive gastric cancer. Regardless of the type of drug, the hypomethylated ATP-binding cassette gene B1 (ABCB1) could be an effective biomarker for chemotherapy-resistant gastric cancer (Choi et al., 2017) (**Figure 1**). This is because ABCB1 hypermethylation silences genes that encode cellular factors necessary for cancer cell resistance to the chemotherapeutic drugs 5-FU and cisplatin (Shitara et al., 2010).

4.2 DNMT Inhibitors

DNMT inhibitors are either nucleoside analogs or non-nucleoside analogs (Erdmann et al., 2015). Azacytidine and decitabine are nucleoside analogs of cytosine that cannot accept a methyl donor at the 5′ position of the pyrimidine ring and depletes cellular DNMT1 (Zeng et al., 2017). Decitabine is integrated into DNA instead of cytidine during duplication, and azacitidine can be incorporated directly into RNA, inhibiting protein synthesis, which causes a substantial reduction in DNMT activity (Navada et al., 2014). Azacitidine was found to suppress the proliferation of GC cell lines and alter DNA methylation (Chen & Wang et al., 2015). Decitabine treatment can cause growth inhibition and reduction in DNMT3A and DNMT3B levels, accompanied by demethylation of the P16 INK4A gene (Liu et al., 2013). Zebularine, another kind of nucleoside analog, is a novel DNMT inhibitor that reduces the expression of the DNMT protein and reactivates epigenetically silenced genes (Tan et al., 2013).

At present, neither azacytidine nor decitabine has been identified as monotherapy for gastric cancer in the clinical setting. This might be because DNMT inhibitors alone cannot reactivate gene expression (Si et al., 2010). However, growing evidence suggests that the combination of DNMT inhibitors and traditional chemotherapy can improve chemosensitization by restoring aberrant epigenetic changes. For example, the combined administration of decitabine and 5-FU showed that TFAP2E is reactively expressed in GC by demethylation and an increase in chemosensitization (Wu F. L. et al., 2015). Additionally, azacytidine upregulates DAPK2, DAPK3, RASSF1, and THBS1 genes which might synergize with chemotherapeutic agents (Tan et al., 2013; Wu F et al., 2015; Zhang et al., 2006). In a clinical trial (Phase 1) (Schneider et al.,

TABLE 3 | Aberrantly methylated genes that play an important role in both diagnosis and prognosis of GC.

	SFRP2
Detection	Sensitivity: 60.9%; Specificity: 86.0% (plasma) Miao et al. (2020)
	Sensitivity: 22.7%; Specificity: 90.0% (stool) Guo et al. (2021)
Prognosis	Shorter PFS (HR = 13.05; 95% CI = 3.05–55.95) and OS (HR = 7.80; 95% CI = 1.81–33.60) (stage III); Shorter PFS (HR = 2.74; 95% CI = 1.58–4.78) and OS (HR = 3.14; 95% CI = 1.67–5.92) (stage IV) (serum) Yan et al. (2021)
	THBS1
Detection	Sensitivity: 63.4%; Specificity: 94.2% (serum) Wang et al. (2015)
Prognosis	Worse DFS (HR = 3.838; 95% CI = 1.691–8.710; p = 0.001) (serum) Hu X. et al. (2021)
	Worse OS survival (p = 0.0483) (serum) Wang et al. (2015)
	UCHL1
Detection	Sensitivity: 56.1%; Specificity: 89.5% (serum) Wang et al. (2015)
Prognosis	Worse OS (p = 0.033) (serum) Wang et al. (2015)
	SOX17
Detection	Sensitivity: 58.9%; Specificity: 100% (serum) Balgkouranidou et al. (2013)
Prognosis	Lower PFS and OS (p < 0.001) (plasma) Karamitrousis et al. (2021)
	APC
Detection	Sensitivity: 17%; Specificity: 100% (serum) Leung et al. (2005)
Prognosis	Poorer OS (HR = 4.6, 95% CI = 1.1–20.3, p = 0.046) (serum) Balgkouranidou et al. (2015)
	E-cadherin
Detection	Sensitivity: 13%; Specificity: 100% (serum) Leung et al. (2005)
	Sensitivity: 57.4%; Specificity: 100% (serum) Lee et al. (2002)
Prognosis	Poorer 3-years survival rate (p < 0.05) (serum) Ikoma et al. (2006)
	RASSF1A
Detection	Sensitivity: 34%; Specificity: 100% (serum) Wang et al. (2008)
Prognosis	Lower OS (HR = 2.33, 95% CI = 1.14–4.85, p = 0.002) (plasma) Pimson et al. (2016)
	Lower PFS and OS (p = 0.004) (plasma) Karamitrousis et al. (2021)
	RNF-180
Detection	Sensitivity: 56%; Specificity: 100% (plasma) Cheung et al. (2012)
Prognosis	Poorer OS (HR = 2.13; 95% CI = 1.11–4.08; p = 0.02) (plasma) Cheung et al. (2012)
	SPG20
Detection	Sensitivity: 48.8%; Specificity: 100% (whole blood) Zhang et al. (2014)
Prognosis	Shorter OS (p = 0.037) (whole blood) Zhang et al. (2014)

2017), researchers investigated pre-treatment with 5-azacitidine as a demethylation reagent in late-stage gastric cancer. They used 5-azacytidine (V) before EOX (epirubicin, oxaliplatin, capecitabine) neoadjuvant chemotherapy in GC patients and the result showed hypomethylation of tumor-related loci such as HPP1, TIMP3, CDKN2A, ESR1, and MGMT. Most patients easily tolerate neoadjuvant VEOX therapy with significant clinical and epigenetic responses. More randomized studies are required to further determine whether the efficacy of this combination is better than chemotherapy alone.

Other types of non-nucleoside analogs such as procaine, hydrazone-gallate, genistein, miRNA-21, miRNA-335, miRNA-148a, and miRNA-155-5p have also been investigated, but further studies need to be performed before approval for clinical use (Erdmann et al., 2016; Li et al., 2014, 2018,; Pan et al., 2010; Xie et al., 2014; Zhang et al., 2015; Zuo et al., 2013).

4.3 Radiotherapy

Despite active anticancer treatment, the overall prognosis of advanced GC is not ideal. Hence, an effective biomarker is required for selecting suitable patients who might benefit from adjuvant radiotherapy. Unfortunately, such predictive indicators have not been determined.

An et al. (An et al., 2020) analyzed methylation maps of 397 gastric cancer samples downloaded from The Cancer Genome Atlas (TCGA) and established a new biomarker called promoter methylation burden of DNA repair genes (RPMB), which meant the ratio of methylated DNA repair genes to the number of all DNA repair genes in order to identify patients who were sensitive to radiotherapy. Subgroup analyses based on overall survival (OS) and disease-free survival (DFS) showed that most of the subgroups tended toward the high-RMPB group. High-RPMB patients receiving radiotherapy with both ≥ T2 tumor and positive lymph nodes showed longer DFS than the low -RPMB group (p = 0.010). High-RPMB patients receiving radiotherapy with both ≥ T2 tumor and positive lymph nodes survived low-RPMB patients in disease-free status (p = 0.010). Therefore, RPMB might be a promising biomarker to evaluate the indications for adjuvant radiotherapy in GC. Furthermore, treatment with 5-aza-CdR can positively affect radiotherapy sensitivity of gastric cancer cells by enhancing the expression of some genes such as p53, RASSF1, and DAPK (Qiu et al., 2009).

5 PROGNOSTIC BIOMARKERS

Aberrant DNA methylation in peripheral blood is also related to multiple prognostic results of gastric cancer. Therefore, it could be used as a prognostic biomarker of GC (**Figure 1**). Hypermethylation of most genes such as APC, E-cadherin, UCHL1, SPG20, RASSF1A, SFRP2, CDH1, THBS1, SOX17, Wif-1, RNF-180, MED12L, HMLH1, MGMT, FLNC, LOX, HOXD10, BNIP3, and PCDH10 was significantly associated with adverse prognosis in GC (Balgkouranidou et al., 2015; Cheung et al., 2012; Hu X et al., 2021; Ikoma et al., 2006; Karamitrousis et al., 2021; Necula et al., 2019; Pimson et al., 2016; Wang et al., 2015; Yan et al., 2021; Yu et al., 2012; Zhang et al., 2014) (**Table 2**). Additionally, Xie et al. found that PBX3 methylation in peripheral blood leukocytes was associated with poorer GC prognosis only in the elderly group (HR = 1.678, 95% CI = 1.046–2.693) and the female group (HR = 2.058, 95% CI = 1.024–4.137) (Xie et al., 2020).

Ko et al. evaluated the prognostic value of LINE-1 methylation level in cfDNA in gastric cancer patients undergoing radical surgery and chemotherapy. The overall survival (OS) of patients with low methylation levels before starting treatment was significantly worse than those with high methylation levels. But methylation level before surgery had no effect on recurrence-free survival (RFS) and OS (Ko et al., 2021). However, another study showed that the methylation status of LINE-1 in leukocyte DNA was not an independent prognostic factor of GC (Tahara et al., 2018).

The association between aberrant DNA methylation and the prognosis of GC needs to be evaluated in larger cohorts and diverse populations. Additionally, more intensive studies are required to determine the potential molecular biomarkers for predicting prognosis in GC patients.

6 CONCLUSION

Helicobacter pylori and EBV are the most important pathogens associated with gastric cancer, which can cause carcinogenesis by

inducing aberrant DNA methylation. DNA methylation has high clinical application value. Aberrant methylation of various genes in body fluids and feces can be used as a non-invasive method for early screening and diagnosis of gastric cancer. Specifically, Reprimo, RUNX3, PCDH10, BARHL2, and MINT25 hypermethylation have both high sensitivity and specificity, which indicates their value in the diagnosis of GC. However, the sensitivity of detecting other types of DNA methylation from peripheral blood and stool is not satisfactory. Using a combination of multiple genes can yield higher sensitivity. DNA methylation can also affect the response to chemoradiotherapy in gastric cancer patients. A combination of DNMT inhibitors and chemotherapy drugs seems to have a better therapeutic effect. Therefore, more DNMT inhibitors that have lower toxicity, an effective response, and a low price need to be developed. Furthermore, DNA methylation can predict a variety of prognostic results for GC patients, such as overall survival (OS) and disease-free survival (DFS). Aberrant methylation of APC, SFRP2, LINE-1, E-cadherin, SOX17, Wif-1, RASSF1A, RNF-180, UCHL1, and SPG20 in peripheral blood was significantly associated with shorter OS in GC. The methylation levels of SFRP2, SOX17, Wif-1, and RASSF1A in peripheral blood had an impact on progression-free survival (PFS). THBS1 methylation in the serum and CDH1 methylation in preoperative peripheral washes (PPW) were related to worse DFS. Additionally, aberrantly methylated genes such as SFRP2, THBS1, UCHL1, SOX17, APC, E-cadherin, RASSF1A, RNF-180, and SPG20 could play an important role in both diagnosis and prognosis of GC (**Table 3**). New tests with improved sensitivity, simplicity, standardization, and cost-effectiveness need to be developed, and new biomarkers need to be validated in larger prospective clinical studies.

AUTHOR CONTRIBUTIONS

YZ designed the study, searched the literature, and wrote the manuscript. HR designed the study and searched the literature. JX contributed to manuscript revision. RC and SL searched the literature. YG, BC, and TZ contributed to thesis supervision. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the Key Technology Research and Development Program of Shandong Province (2019GSF108254).

REFERENCES

Abbaszadegan, M. R., Moaven, O., Sima, H. R., Ghafarzadegan, K., A'Rabi, A., Forghani, M. N., et al. (2008). P16 Promoter Hypermethylation: A Useful Serum Marker for Early Detection of Gastric Cancer. *World J. Gastroenterol.* 14 (13), 2055–2060. doi:10.3748/wjg.14.2055

Alarcón, M. A., Olivares, W., Córdova-Delgado, M., Muñoz-Medel, M., de Mayo, T., Carrasco-Aviño, G., et al. (2020). The Reprimo-like Gene Is an Epigenetic-Mediated Tumor Suppressor and a Candidate Biomarker for the Non-invasive Detection of Gastric Cancer. *Int. J. Mol. Sci.* 21 (24). doi:10.3390/ijms21249472

An, N., Yu, Z., He, X. J., Zhao, Y. Y., Yu, L., Zhang, Y. C., et al. (2020). Promoter Methylation of DNA Repair Genes Predicts Disease-free Survival of Gastric Adenocarcinoma after Adjuvant Radiotherapy. *Mol. Ther. Oncolytics* 18, 109–117. doi:10.1016/j.omto.2020.06.006

Balgkouranidou, I., Karayiannakis, A., Matthaios, D., Bolanaki, H., Tripsianis, G., Tentes, A. A., et al. (2013). Assessment of SOX17 DNA Methylation in Cell Free DNA from Patients with Operable Gastric Cancer. Association with Prognostic Variables and Survival. *Clin. Chem. Lab. Med.* 51(7), 1505–1510. doi:10.1515/cclm-2012-0320

Balgkouranidou, I., Matthaios, D., Karayiannakis, A., Bolanaki, H., Michailidis, P., Xenidis, N., et al. (2015). Prognostic Role of APC and RASSF1A Promoter Methylation Status in Cell Free Circulating DNA of Operable Gastric Cancer Patients. *Mutat. Research/Fundamental Mol. Mech. Mutagenesis* 778, 46–51. doi:10.1016/j.mrfmmm.2015.05.002

Barchitta, M., Quattrocchi, A., Maugeri, A., Vinciguerra, M., and Agodi, A. (2014). LINE-1 Hypomethylation in Blood and Tissue Samples as an Epigenetic Marker for Cancer Risk: A Systematic Review and Meta-Analysis. [Journal Article; Meta-Analysis; Review; Systematic Review. *PLoS One* 9 (10), e109478. doi:10.1371/journal.pone.0109478

Bernal, C., Aguayo, F., Villarroel, C., Vargas, M., Díaz, I., Ossandon, F. J., et al. (2008). Reprimo as a Potential Biomarker for Early Detection in Gastric Cancer. [Journal Article; Research Support, Non-U.S. Gov't]. *Clin. Cancer Res.* 14 (19), 6264–6269. doi:10.1158/1078-0432.CCR-07-4522

Chansangpetch, S., Prombhul, S., Tantisevi, V., Sodsai, P., Manassakorn, A., Hirankarn, N., et al. (2018). DNA Methylation Status of the Interspersed Repetitive Sequences for LINE-1, Alu, HERV-E, and HERV-K in Trabeculectomy Specimens from Glaucoma Eyes. *J. Ophthalmol.* 2018, 9171536. doi:10.1155/2018/9171536

Chen, X., Lin, Z., Xue, M., Si, J., and Chen, S. (2015a). Zic1 Promoter Hypermethylation in Plasma DNA Is a Potential Biomarker for Gastric Cancer and Intraepithelial Neoplasia. *PLoS One* 10 (7), e133906. doi:10.1371/journal.pone.0133906

Chen, X., Wang, F., Li, J., He, X., Liu, X., and Ma, L. (2015b). The Effect of Two Nucleoside Antitumor Drugs on the Proliferation and DNA Methylation of Human Gastric Cancer Cells. *Oncol. Lett.* 10 (3), 1919–1923. doi:10.3892/ol.2015.3427

Chen, Z., Fan, J. Q., Li, J., Li, Q. S., Yan, Z., Jia, X. K., et al. (2009). Promoter Hypermethylation Correlates with the Hsulf-1 Silencing in Human Breast and Gastric cancerResearch Support, Non-U.S. Gov't]. *[Journal Articleint. J. Cancer* 124 (3), 739–744. doi:10.1002/ijc.23960

Cheng, Y. Y., Yu, J., Wong, Y. P., Man, E. P., To, K. F., Jin, V. X., et al. (2007). Frequent Epigenetic Inactivation of Secreted Frizzled-Related Protein 2 (SFRP2) by Promoter Methylation in Human Gastric cancerResearch Support, Non-U.S. Gov't]. *[Journal Articlebr J. Cancer* 97 (7), 895–901. doi:10.1038/sj.bjc.6603968

Cheung, K. F., Lam, C. N., Wu, K., Ng, E. K., Chong, W. W., Cheng, A. S., et al. (2012). Characterization of the Gene Structure, Functional Significance, and Clinical Application of RNF180, a Novel Gene in Gastric Cancer. *Cancer-am. Cancer Soc.* 118 (4), 947–959. doi:10.1002/cncr.26189

Choi, S. J., Jung, S. W., Huh, S., Chung, Y. S., Cho, H., and Kang, H. (2017). Alteration of DNA Methylation in Gastric Cancer with Chemotherapy. *J. Microbiol. Biotechnol.* 27 (8), 1367–1378. doi:10.4014/jmb.1704.04035

Dauksa, A., Gulbinas, A., Endzinas, Z., Oldenburg, J., and El-Maarri, O. (2014). DNA Methylation at Selected CpG Sites in Peripheral Blood Leukocytes Is Predictive of Gastric Cancer. *Anticancer Res.* 34 (10), 5381–5388.

Digklia, A., and Wagner, A. D. (2016). Advanced Gastric Cancer: Current Treatment Landscape and Future Perspectives. *[Journal Article; Review] world J. Gastroenterol.* 22 (8), 2403–2414. doi:10.3748/wjg.v22.i8.2403

Ebrahimi, V., Soleimanian, A., Ebrahimi, T., Azargun, R., Yazdani, P., Eyvazi, S., et al. (2020). Epigenetic Modifications in Gastric Cancer: Focus on DNA Methylation. *Gene* 742, 144577. doi:10.1016/j.gene.2020.144577

Erdmann, A., Halby, L., Fahy, J., and Arimondo, P. B. (2015). Targeting DNA Methylation with Small Molecules: What's Next? *J. Med. Chem.* 58 (6), 2569–2583. doi:10.1021/jm500843d

Erdmann, A., Menon, Y., Gros, C., Masson, V., Aussagues, Y., Ausseil, F., et al. (2016). Identification and Optimization of Hydrazone-Gallate Derivatives as Specific Inhibitors of DNA Methyltransferase 3A. *Future Med. Chem.* 8 (4), 373–380. doi:10.4155/fmc.15.192

Eyvazi, S., Khamaneh, A. M., Tarhriz, V., Bandehpour, M., Hejazi, M. S., Sadat, A. T. E., et al. (2020). CpG Islands Methylation Analysis of CDH11, EphA5, and HS3ST2 Genes in Gastric Adenocarcinoma Patients. *J. Gastrointest. Cancer.* 51 (2), 579–583. doi:10.1007/s12029-019-00290-1

Fattahi, S., Kosari-Monfared, M., Ghadami, E., Golpour, M., Khodadadi, P., Ghasemiyan, M., et al. (2018). Infection-associated epigenetic alterations in gastric cancer: New insight in cancer therapy. *J. Cell. Physiol.* 233 (12), 9261–9270. doi:10.1002/jcp.27030

Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T., and Adams, P. D. (2018). DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol. Cel.* 71 (6), 882–895. doi:10.1016/j.molcel.2018.08.008

Fukayama, M., and Ushiku, T. (2011). Epstein-Barr Virus-Associated Gastric Carcinoma. *Pathol. Res. Pract.* 207 (9), 529–537. doi:10.1016/j.prp.2011.07.004

Gao, Y., Baccarelli, A., Shu, X. O., Ji, B. T., Yu, K., Tarantini, L., et al. (2012). Blood Leukocyte Alu and LINE-1 Methylation and Gastric Cancer Risk in the Shanghai Women's Health Study[Journal Article; Research Support. *Br. J. Cancer* 106 (3), 585–591. doi:10.1038/bjc.2011.562

Global, B. O. D. C. (2019). Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global burden of Disease Study. *JAMA Oncol.* 5 (12), 1749–1768. doi:10.1001/jamaoncol.2019.2996

Guo, P., Zhou, X. J., Xu, L., Chen, H., Zhao, L., Sun, M. H., et al. (2021). [Application of Fecal DNA Methylation Biomarkers Detection in Gastric Cancer Screening. *Zhonghua Yi Xue Za Zhi* 101 (11), 808–812. doi:10.3760/cma.j.cn112137-20200916-02659

Guo, W., Dong, Z., Guo, Y., Chen, Z., Kuang, G., and Yang, Z. (2011). Aberrant Methylation of the CpG Island of HLTF Gene in Gastric Cardia Adenocarcinoma and dysplasiaResearch Support, Non-U.S. Gov't]. *[Journal Articleclin. Biochem.* 44 (10-11), 784–788. doi:10.1016/j.clinbiochem.2011.04.006

Guo, X., Liu, W., Pan, Y., Ni, P., Ji, J., Guo, L., et al. (2010). Homeobox Gene IRX1 Is a Tumor Suppressor Gene in Gastric Carcinoma. *Oncogene* 29 (27), 3908–3920. doi:10.1038/onc.2010.143

Han, Q., Zhou, H., Xie, W., Sun, T., Wei, R., Nie, C., et al. (2020). Association between the Methylation of the STAT1 and SOCS3 in Peripheral Blood and Gastric Cancer. *J. Gastroen. Hepatol.* 35 (8), 1347–1354. doi:10.1111/jgh.15021

Hibi, K., Goto, T., Shirahata, A., Saito, M., Kigawa, G., Nemoto, H., et al. (2011). Detection of TFPI2 Methylation in the Serum of Gastric Cancer Patients. [Comparative Study; Journal Article]. *Anticancer Res.* 31 (11), 3835–3838.

Hino, R., Uozaki, H., Murakami, N., Ushiku, T., Shinozaki, A., Ishikawa, S., et al. (2009). Activation of DNA Methyltransferase 1 by EBV Latent Membrane Protein 2A Leads to Promoter Hypermethylation of PTEN Gene in Gastric Carcinoma. *Cancer Res.* 69 (7), 2766–2774. doi:10.1158/0008-5472.CAN-08-3070

Hou, L., Wang, H., Sartori, S., Gawron, A., Lissowska, J., Bollati, V., et al. (2010). Blood Leukocyte DNA Hypomethylation and Gastric Cancer Risk in a High-Risk Polish Population. [Comparative Study; Journal Article; Randomized Controlled Trial; Research Support, N.I.H., Intramural]. *Int. J. Cancer* 127 (8), 1866–1874. doi:10.1002/ijc.25190

Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., et al. (2014). Drug Resistance in Cancer: An Overview. *[Journal Article; Review] Cancers (Basel).* 6 (3), 1769–1792. doi:10.3390/cancers6031769

Hu, D., Lou, X., Meng, N., Li, Z., Teng, Y., Zou, Y., et al. (2021). Peripheral Blood-Based DNA Methylation of Long Non-coding RNA H19 and Metastasis-Associated Lung Adenocarcinoma Transcript 1 Promoters Are Potential Non-invasive Biomarkers for Gastric Cancer Detection. *Cancer Control* 28, 1399505581. doi:10.1177/10732748211043667

Hu, W., Zheng, W., Liu, Q., Chu, H., Chen, S., Kim, J. J., et al. (2017). Diagnostic Accuracy of DNA Methylation in Detection of Gastric Cancer: A Meta-Analysis. *Oncotarget* 8 (68), 113142–113152. doi:10.18632/oncotarget.22613

Hu, X., Ling, Z., Hong, L., Yu, Q., Li, P., and Ling, Z. (2021). Circulating Methylated THBS1 DNAs as a Novel Marker for Predicting Peritoneal Dissemination in Gastric Cancer. *J. Clin. Lab. Anal.* 35 (9), e23936. doi:10.1002/jcla.23936

Huang, Z. B., Zhang, H. T., Yu, B., and Yu, D. H. (2021). Cell-free DNA as a Liquid Biopsy for Early Detection of Gastric Cancer. [Journal Article; Review]. *Oncol. Lett.* 21 (1), 3. doi:10.3892/ol.2020.12264

Hur, K., Niwa, T., Toyoda, T., Tsukamoto, T., Tatematsu, M., Yang, H. K., et al. (2011). Insufficient Role of Cell Proliferation in Aberrant DNA Methylation Induction and Involvement of Specific Types of Inflammation. *Carcinogenesis* 32 (1), 35–41. doi:10.1093/carcin/bgq219

Ikoma, H., Ichikawa, D., Koike, H., Ikoma, D., Tani, N., Okamoto, K., et al. (2006). Correlation between Serum DNA Methylation and Prognosis in Gastric Cancer Patients. *Anticancer Res.* 26 (3 B), 2313–2316.

Kaneda, A., Matsusaka, K., Aburatani, H., and Fukayama, M. (2012). Epstein-Barr Virus Infection as an Epigenetic Driver of Tumorigenesis. *Cancer Res.* 72 (14), 3445–3450. doi:10.1158/0008-5472.CAN-11-3919

Kanyama, Y., Hibi, K., Nakayama, H., Kodera, Y., Ito, K., Akiyama, S., et al. (2003). Detection of P16 Promoter Hypermethylation in Serum of Gastric Cancer Patients. [Comparative Study; Journal Article]. *Cancer Sci.* 94 (5), 418–420. doi:10.1111/j.1349-7006.2003.tb01457.x

Karamitrousis, E. I., Balgkouranidou, I., Xenidis, N., Amarantidis, K., Biziota, E., Koukaki, T., et al. (2021). Prognostic Role of RASSF1A, SOX17 and Wif-1 Promoter Methylation Status in Cell-free DNA of Advanced Gastric Cancer Patients. *Technol. Cancer Res. Treat.* 20, 1079240927. doi:10.1177/1533033820973279

Kim, N. (2019). Chemoprevention of Gastric Cancer by *Helicobacter pylori* Eradication and its Underlying Mechanism. [Journal Article; Review]. *J. Gastroenterol. Hepatol.* 34 (8), 1287–1295. doi:10.1111/jgh.14646

Ko, K., Kananazawa, Y., Yamada, T., Kakinuma, D., Matsuno, K., Ando, F., et al. (2021). Methylation Status and Long-Fragment Cell-free DNA Are Prognostic Biomarkers for Gastric Cancer. [Journal Article; Observational Study; Research Support, Non-U.S. Gov't]. *Cancer Med.* 10 (6), 2003–2012. doi:10.1002/cam4.3755

Kosumi, K., Baba, Y., Ishimoto, T., Harada, K., Miyake, K., Izumi, D., et al. (2015). Relationship between LINE-1 Hypomethylation and *Helicobacter pylori* Infection in Gastric Mucosae. [Comparative Study; Journal Article; Randomized Controlled Trial]. *Med. Oncol.* 32 (4), 117. doi:10.1007/s12032-015-0571-5

Lee, H. S., Hwang, S. M., Kim, T. S., Kim, D. W., Park, D. J., Kang, S. B., et al. (2013). Circulating Methylated Septin 9 Nucleic Acid in the Plasma of Patients with Gastrointestinal Cancer in the Stomach and colon. *Transl Oncol.* 6 (3), 290–296. doi:10.1593/tlo.13118

Lee, T., Leung, W. K., Chan, M. W. Y., Ng, E. K. W., Tong, J. H. M., Lo, K., et al. (2002). Detection of Gene Promoter Hypermethylation in the Tumor and Serum of Patients with Gastric Carcinoma. *Clin. Cancer Res.* 8 (6), 1761.

Leung, W. K., To, K. F., Chu, E. S., Chan, M. W., Bai, A. H., Ng, E. K., et al. (2005). Potential Diagnostic and Prognostic Values of Detecting Promoter Hypermethylation in the Serum of Patients with Gastric Cancer. *Br. J. Cancer* 92 (12), 2190–2194. doi:10.1038/sj.bjc.6602636

Li, Y., Wang, Y., Li, D., Zhang, Y., Zhao, T., and Li, C. (2018). Procaine Is a Specific DNA Methylation Inhibitor with Anti-tumor Effect for Human Gastric Cancer. *J. Cel. Biochem.* 119 (2), 2440–2449. doi:10.1002/jcb.26407

Li, Z., Li, D., Zhang, G., Xiong, J., Jie, Z., Cheng, H., et al. (2014). Methylation-associated Silencing of MicroRNA-335 Contributes Tumor Cell Invasion and Migration by Interacting with RASA1 in Gastric Cancer. [Journal Article]. *Am. J. Cancer Res.* 4 (6), 648–662.

Lin, Z., Luo, M., Chen, X., He, X., Qian, Y., Lai, S., et al. (2017). Combined Detection of Plasma ZIC1, HOXD10 and RUNX3 Methylation Is a Promising Strategy for Early Detection of Gastric Cancer and Precancerous Lesions. *J. Cancer* 8 (6), 1038–1044. doi:10.7150/jca.18169

Liu, C., Li, N., Lu, H., Wang, Z., Chen, C., Wu, L., et al. (2015). Circulating SFRP1 Promoter Methylation Status in Gastric Adenocarcinoma and Esophageal Squamous Cell Carcinoma. *Biomed. Rep.* 3 (1), 123–127. doi:10.3892/br.2014.388

Liu, J., Xie, Y., Wang, F., Zhang, L., Zhang, Y., and Luo, H. (2013). Cytotoxicity of 5-Aza-2′-Deoxycytidine against Gastric Cancer Involves DNA Damage in an ATM-P53 Dependent Signaling Pathway and Demethylation of P16INK4A. *Biomed. Pharmacother.* 67 (1), 78–87. doi:10.1016/j.biopha.2012.10.015

Liu, L., Liu, C., Fotouhi, O., Fan, Y., Wang, K., Xia, C., et al. (2017). TERT Promoter Hypermethylation in Gastrointestinal Cancer: A Potential Stool Biomarker. *[Journal Article; Res. Support, NonOncologist* 22 (10), 1178–1188. doi:10.1634/theoncologist.2017-0064

Lu, X., Yu, J., Ying, L., Han, J., Wang, S., Yu, Q., et al. (2012). Stepwise Cumulation of RUNX3 Methylation Mediated by *Helicobacter pylori* Infection Contributes to Gastric Carcinoma Progression. *Cancer-am. Cancer Soc.* 118 (22), 5507–5517. doi:10.1002/cncr.27604

Maeda, M., Moro, H., and Ushijima, T. (2017). Mechanisms for the Induction of Gastric Cancer by *Helicobacter pylori* Infection: Aberrant DNA Methylation Pathway. *Gastric Cancer* 20 (1), 8–15. doi:10.1007/s10120-016-0650-0

Matsusaka, K., Funata, S., Fukayama, M., and Kaneda, A. (2014). DNA Methylation in Gastric Cancer, Related to *Helicobacter pylori* and Epstein-Barr Virus. *World J. Gastroenterol.* 20 (14), 3916–3926. doi:10.3748/wjg.v20.i14.3916

Matsusaka, K., Kaneda, A., Nagae, G., Ushiku, T., Kikuchi, Y., Hino, R., et al. (2011). Classification of Epstein–Barr Virus–Positive Gastric Cancers by Definition of DNA Methylation Epigenotypes. *Cancer Res.* 71 (23), 7187. doi:10.1158/0008-5472.CAN-11-1349

Miao, J., Liu, Y., Zhao, G., Liu, X., Ma, Y., Li, H., et al. (2020). Feasibility of Plasma-Methylated SFRP2 for Early Detection of Gastric Cancer. [Journal Article]. *Cancer Control* 27 (2), 1148357823. doi:10.1177/1073274820922559

Min, J., Choi, B., Han, T. S., Lee, H. J., Kong, S. H., Suh, Y. S., et al. (2017). Methylation Levels of LINE-1 as a Useful Marker for Venous Invasion in Both FFPE and Frozen Tumor Tissues of Gastric Cancer. [Journal Article]. *Mol. Cell* 40 (5), 346–354. doi:10.14348/molcells.2017.0013

Moore, L. D., Le, T., and Fan, G. (2013). DNA Methylation and its Basic Function. *Neuropsychopharmacol* 38 (1), 23–38. doi:10.1038/npp.2012.112

Muhammad, J. S., Eladl, M. A., and Khoder, G. (2019). Helicobacter Pylori-Induced DNA Methylation as an Epigenetic Modulator of Gastric Cancer: Recent Outcomes and Future Direction. *Pathogens* 8, 23. doi:10.3390/pathogens8010023

Muretto, P., Ruzzo, A., Pizzagalli, F., Graziano, F., Maltese, P., Zingaretti, C., et al. (2008). Endogastric Capsule for E-Cadherin Gene (CDH1) Promoter Hypermethylation Assessment in DNA from Gastric Juice of Diffuse Gastric Cancer Patients. *Ann. Oncol.* 19 (3), 516–519. doi:10.1093/annonc/mdm493

Nakajima, T., Enomoto, S., Yamashita, S., Ando, T., Nakanishi, Y., Nakazawa, K., et al. (2010). Persistence of a Component of DNA Methylation in Gastric Mucosae after *Helicobacter pylori* Eradication. *J. Gastroenterol.* 45 (1), 37–44. doi:10.1007/s00535-009-0142-7

Namba-Fukuyo, H., Funata, S., Matsusaka, K., Fukuyo, M., Rahmutulla, B., Mano, Y., et al. (2016). TET2 Functions as a Resistance Factor against DNA Methylation Acquisition during Epstein-Barr Virus Infection. *Oncotarget* 7, 49. doi:10.18632/oncotarget.13130

Navada, S. C., Steinmann, J., Lübbert, M., and Silverman, L. R. (2014). Clinical Development of Demethylating Agents in Hematology. *J. Clin. Invest.* 124 (1), 40–46. doi:10.1172/JCI69739

Necula, L., Matei, L., Dragu, D., Neagu, A. I., Mambet, C., Nedeianu, S., et al. (2019). Recent Advances in Gastric Cancer Early Diagnosis. *World J. Gastroentero.* 25 (17), 2029–2044. doi:10.3748/wjg.v25.i17.2029

Ng, E. K., Leung, C. P., Shin, V. Y., Wong, C. L., Ma, E. S., Jin, H. C., et al. (2011). Quantitative Analysis and Diagnostic Significance of Methylated SLC19A3 DNA in the Plasma of Breast and Gastric Cancer patientsResearch Support, Non-U.S. Gov't]. *[Journal ArticlePLoS One* 6 (7), e22233. doi:10.1371/journal.pone.0022233

Niwa, T., Toyoda, T., Tsukamoto, T., Mori, A., Tatematsu, M., and Ushijima, T. (2013). Prevention of *Helicobacter pylori*–Induced Gastric Cancers in Gerbils by a DNA Demethylating Agent. *Cancer Prev. Res.* 6 (4), 263. doi:10.1158/1940-6207.CAPR-12-0369

Okada, T., Nakamura, M., Nishikawa, J., Sakai, K., Zhang, Y., Saito, M., et al. (2013). Identification of Genes Specifically Methylated in Epstein-Barr Virus-Associated Gastric carcinomasResearch Support, Non-U.S. Gov't]. *[Journal Articlecancer Sci.* 104 (10), 1309–1314. doi:10.1111/cas.12228

Pan, W., Zhu, S., Yuan, M., Cui, H., Wang, L., Luo, X., et al. (2010). MicroRNA-21 and MicroRNA-148a Contribute to DNA Hypomethylation in Lupus $CD4^+$ T Cells by Directly and Indirectly Targeting DNA Methyltransferase 1. *J. Immunol.* 184 (12), 6773. doi:10.4049/jimmunol.0904060

Pessoa, L. S., Heringer, M., and Ferrer, V. P. (2020). CtDNA as a Cancer Biomarker: A Broad Overview. [Journal Article; Review]. *Crit. Rev. Oncol. Hematol.* 155, 103109. doi:10.1016/j.critrevonc.2020.103109

Pimson, C., Ekalaksananan, T., Pientong, C., Promthet, S., Putthanachote, N., Suwanrungruang, K., et al. (2016). Aberrant Methylation of PCDH10 and RASSF1A Genes in Blood Samples for Non-invasive Diagnosis and Prognostic Assessment of Gastric Cancer. *PeerJ* 4, e2112. doi:10.7717/peerj.2112

Qi, Q., Pan, Y. F., Shen, J. J., Gu, X. Q., Han, S. W., Liao, H. H., et al. (2016). Circulating DNA for Detection of Gastric Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 20 (12), 2558–2564.

Qiu, H., Yashiro, M., Shinto, O., Matsuzaki, T., and Hirakawa, K. (2009). DNA Methyltransferase Inhibitor 5-Aza-CdR Enhances the Radiosensitivity of Gastric Cancer Cells. *Cancer Sci.* 100 (1), 181–188. doi:10.1111/j.1349-7006.2008.01004.x

Qu, Y., Dang, S., and Hou, P. (2013). Gene Methylation in Gastric Cancer. [Journal Article; Research Support, Non-U.S. Gov't; Review]. *Clin. Chim. Acta* 424, 53–65. doi:10.1016/j.cca.2013.05.002

Saito, M., Suzuki, K., Maeda, T., Kato, T., Kamiyama, H., Koizumi, K., et al. (2012). The Accumulation of DNA Demethylation in Sat α in normal Gastric Tissues with *Helicobacter pylori* Infection Renders Susceptibility to Gastric Cancer in Some Individuals. [Journal Article; Research Support, Non-U.S. Gov't]. *Oncol. Rep.* 27 (6), 1717–1725. doi:10.3892/or.2012.1718

Sakakura, C., Hamada, T., Miyagawa, K., Nishio, M., Miyashita, A., Nagata, H., et al. (2009). Quantitative Analysis of Tumor-Derived Methylated RUNX3 Sequences in the Serum of Gastric Cancer Patients. *Anticancer Res.* 29 (7), 2619–2625.

Schneider, B. J., Shah, M. A., Klute, K., Ocean, A., Popa, E., Altorki, N., et al. (2017). Phase I Study of Epigenetic Priming with Azacitidine Prior to Standard Neoadjuvant Chemotherapy for Patients with Resectable Gastric and Esophageal Adenocarcinoma: Evidence of Tumor Hypomethylation as an Indicator of Major Histopathologic Response. *Clin. Cancer Res.* 23 (11), 2673–2680. doi:10.1158/1078-0432.CCR-16-1896

Sexton, R. E., Al, H. M., Diab, M., and Azmi, A. S. (2020). Gastric Cancer: A Comprehensive Review of Current and Future Treatment Strategies. *Cancer Metastasis Rev.* 39 (4), 1179–1203. doi:10.1007/s10555-020-09925-3

Shao, L., Chen, Z., Peng, D., Soutto, M., Zhu, S., Bates, A., et al. (2018). Methylation of the HOXA10 Promoter Directs miR-196b-5p-dependent Cell Proliferation and Invasion of Gastric Cancer Cells. *Mol. Cancer Res.* 16 (4), 696–706. doi:10.1158/1541-7786.MCR-17-0655

Shin, C. M., Kim, N., Park, J. H., Kang, G. H., Kim, J. S., Jung, H. C., et al. (2012). Prediction of the Risk for Gastric Cancer Using Candidate Methylation Markers in the Non-neoplastic Gastric Mucosae. *J. Pathol.* 226 (4), 654–665. doi:10.1002/path.2990

Shinozaki-Ushiku, A., Kunita, A., and Fukayama, M. (2015). Update on Epstein-Barr Virus and Gastric Cancer (Review). *Int. J. Oncol.* 46 (4), 1421–1434. doi:10.3892/ijo.2015.2856

Shitara, K., Matsuo, K., Ito, S., Sawaki, A., Kawai, H., Yokota, T., et al. (2010). Effects of Genetic Polymorphisms in the ABCB1 Gene on Clinical Outcomes in Patients with Gastric Cancer Treated by Second-Line Chemotherapy. *Asian Pac. J. Cancer Prev.* 11 (2), 447–452.

Si, J., Boumber, Y. A., Shu, J., Qin, T., Ahmed, S., He, R., et al. (2010). Chromatin Remodeling Is Required for Gene Reactivation after Decitabine-Mediated DNA Hypomethylation. [Journal Article; Research Support, N.I.H., Extramural]. *Cancer Res.* 70 (17), 6968–6977. doi:10.1158/0008-5472.CAN-09-4474

Skvortsova, K., Stirzaker, C., and Taberlay, P. (2019). The DNA Methylation Landscape in Cancer. *Essays Biochem.* 63 (6), 797–811. doi:10.1042/EBC20190037

Sohn, B. H., Hwang, J. E., Jang, H. J., Lee, H. S., Oh, S. C., Shim, J. J., et al. (2017). Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by the Cancer Genome Atlas Project. [Journal Article]. *Clin. Cancer Res.* 23 (15), 4441–4449. doi:10.1158/1078-0432.CCR-16-2211

Sun, H., Zhou, H., Zhang, Y., Chen, J., Han, X., Huang, D., et al. (2018). Aberrant Methylation of FAT4 and SOX11 in Peripheral Blood Leukocytes and Their Association with Gastric Cancer Risk. *J. Cancer* 9 (13), 2275–2283. doi:10.7150/jca.24797

Tahara, T., and Arisawa, T. (2015). DNA Methylation as a Molecular Biomarker in Gastric Cancer. *Epigenomics-UK.* 7 (3), 475–486. doi:10.2217/epi.15.4

Tahara, T., Maegawa, S., Chung, W., Garriga, J., Jelinek, J., Estécio, M. R. H., et al. (2013). Examination of Whole Blood DNA Methylation as a Potential Risk Marker for Gastric Cancer. *Cancer Prev. Res.* 6 (10), 1093–1100. doi:10.1158/1940-6207.CAPR-13-0034

Tahara, T., and Arisawa, T. (2015). DNA methylation as a molecular biomarker in gastric cancer. *608 Epigenomics-UK* 7 (3), 476–486. doi:10.2217/epi.15.4

Tahara, T., Tahara, S., Horiguchi, N., Kawamura, T., Okubo, M., Yamada, H., et al. (2018). Methylation Status of IGF2 DMR and LINE1 in Leukocyte DNA Provides Distinct Clinicopathological Features of Gastric Cancer Patients. *Clin. Exp. Med.* 18 (2), 215–220. doi:10.1007/s10238-017-0471-4

Tan, W., Zhou, W., Yu, H., Luo, H., and Shen, L. (2013). The DNA Methyltransferase Inhibitor Zebularine Induces Mitochondria-Mediated Apoptosis in Gastric Cancer Cells *In Vitro* and *In Vivo*. *Biochem. Bioph. Res. Co.* 430 (1), 250–255. doi:10.1016/j.bbrc.2012.10.143

Usui, G., Matsusaka, K., Mano, Y., Urabe, M., Funata, S., Fukayama, M., et al. (2021). DNA Methylation and Genetic Aberrations in Gastric Cancer. *Digestion* 102 (1), 25–32. doi:10.1159/000511243

Wang, G., Zhang, W., Zhou, B., Jin, C., Wang, Z., Yang, Y., et al. (2015). The Diagnosis Value of Promoter Methylation of UCHL1 in the Serum for Progression of Gastric Cancer. *Biomed. Res. Int.* 2015, 741030. doi:10.1155/2015/741030

Wang, Y. C., Yu, Z. H., Liu, C., Xu, L. Z., Yu, W., Lu, J., et al. (2008). Detection of RASSF1A Promoter Hypermethylation in Serum from Gastric and Colorectal Adenocarcinoma Patients. *[Journal Article]world J. Gastroenterol.* 14 (19), 3074–3080. doi:10.3748/wjg.14.3074

Watanabe, Y., Kim, H. S., Castoro, R. J., Chung, W., Estecio, M. R. H., Kondo, K., et al. (2009). Sensitive and Specific Detection of Early Gastric Cancer with DNA Methylation Analysis of Gastric Washes. *Gastroenterology* 136 (7), 2149–2158. doi:10.1053/j.gastro.2009.02.085

Wu, F., Li, R., Yang, M., Yue, G., Wang, H., Liu, Q., et al. (2015). Gelatinases-stimuli Nanoparticles Encapsulating 5-fluorouridine and 5-Aza-2′-Deoxycytidine Enhance the Sensitivity of Gastric Cancer Cells to Chemical Therapeutics. *Cancer Lett.* 363 (1), 7–16. doi:10.1016/j.canlet.2015.01.006

Wu, F. L., Li, R. T., Yang, M., Yue, G. F., Wang, H. Y., Liu, Q., et al. (2015). Gelatinases-stimuli Nanoparticles Encapsulating 5-fluorouridine and 5-Aza-2'-Deoxycytidine Enhance the Sensitivity of Gastric Cancer Cells to Chemical therapeuticsResearch Support, Non-U.S. Gov't]. *[Journal Articlecancer Lett.* 363 (1), 7–16. doi:10.1016/j.canlet.2015.01.006

Xie, Q., Bai, Q., Zou, L., Zhang, Q., Zhou, Y., Chang, H., et al. (2014). Genistein Inhibits DNA Methylation and Increases Expression of Tumor Suppressor Genes in Human Breast Cancer Cells. *Genes, Chromosomes and Cancer* 53 (5), 422–431. doi:10.1002/gcc.22154

Xie, W., Zhou, H., Han, Q., Sun, T., Nie, C., Hong, J., et al. (2020). Relationship between DLEC1 and PBX3 Promoter Methylation and the Risk and Prognosis of Gastric Cancer in Peripheral Blood Leukocytes. [Journal Article]. *J. Cancer Res. Clin. Oncol.* 146 (5), 1115–1124. doi:10.1007/s00432-020-03171-4

Xue, W., Feng, Y., Wang, F., Li, P., Liu, Y., Guo, Y., et al. (2016). The Value of Serum RASSF10 Hypermethylation as a Diagnostic and Prognostic Tool for Gastric Cancer. *Tumor Biol.* 37 (8), 11249–11257. doi:10.1007/s13277-016-5001-6

Yamamoto, H., Watanabe, Y., Oikawa, R., Morita, R., Yoshida, Y., Maehata, T., et al. (2016). BARHL2 Methylation Using Gastric Wash DNA or Gastric Juice Exosomal DNA Is a Useful Marker for Early Detection of Gastric Cancer in an H. Pylori-independent Manner. *Clin. translational Gastroenterol.* 7 (7), e184. doi:10.1038/ctg.2016.40

Yamamoto, H., Watanabe, Y., Sato, Y., Maehata, T., and Itoh, F. (2020). Non-Invasive Early Molecular Detection of Gastric Cancers. *Cancers (Basel).* 12 (10). doi:10.3390/cancers12102880

Yan, H., Chen, W., Ge, K., Mao, X., Li, X., Liu, W., et al. (2021). Value of Plasma Methylated SFRP2 in Prognosis of Gastric Cancer. *Dig. Dis. Sci.* 66 (11), 3854–3861. doi:10.1007/s10620-020-06710-8

Yau, T. O., Tang, C. M., and Yu, J. (2014). Epigenetic Dysregulation in Epstein-Barr Virus-Associated Gastric Carcinoma: Disease and Treatments. *World J. Gastroenterol.* 20 (21), 6448–6456. doi:10.3748/wjg.v20.i21.6448

Yoshida, T., Yamashita, S., Takamura-Enya, T., Niwa, T., Ando, T., Enomoto, S., et al. (2011). Alu and Satα Hypomethylation in Helicobacter Pylori-Infected Gastric Mucosae. *Int. J. Cancer* 128 (1), 33–39. doi:10.1002/ijc.25534

Yu, Q., Wang, X., Luo, J., Wang, S., Fang, X., Yu, J., et al. (2012). CDH1 Methylation in Preoperative Peritoneal Washes Is an Independent Prognostic Factor for Gastric Cancer. *J. Surg. Oncol.* 106 (6), 765–771. doi:10.1002/jso.23116

Yuasa, Y. (2010). Epigenetics in Molecular Epidemiology of Cancer a New scopeResearch Support, Non-U.S. Gov't]. *[Journal Articleadv. Genet.* 71, 211–235. doi:10.1016/B978-0-12-380864-6.00007-9

Zeng, X. Q., Wang, J., and Chen, S. Y. (2017). Methylation Modification in Gastric Cancer and Approaches to Targeted Epigenetic Therapy (Review). *Int. J. Oncol.* 50 (6), 1921–1933. doi:10.3892/ijo.2017.3981

Zhang, G., Estève, P., Chin, H. G., Terragni, J., Dai, N., Corrêa, I. R. J., et al. (2015). Small RNA-Mediated DNA (Cytosine-5) Methyltransferase 1 Inhibition Leads to Aberrant DNA Methylation. *Nucleic Acids Res.* 43 (12), 6112–6124. doi:10.1093/nar/gkv518

Zhang, H., Song, Y., Xia, P., Cheng, Y., Guo, Q., Diao, D., et al. (2014). Detection of Aberrant Hypermethylated Spastic Paraplegia-20 as a Potential Biomarker and Prognostic Factor in Gastric Cancer. *Med. Oncol.* 31 (2), 830. doi:10.1007/s12032-013-0830-2

Zhang, X., Yashiro, M., Ohira, M., Ren, J., and Hirakawa, K. (2006). Synergic Antiproliferative Effect of DNA Methyltransferase Inhibitor in Combination with Anticancer Drugs in Gastric Carcinoma. *Cancer Sci.* 97 (9), 938–944. doi:10.1111/j.1349-7006.2006.00253.x

Zhang, Y., Ye, X., Geng, J., and Chen, L. (2010). Epigenetic Inactivation of Deleted in Lung and Esophageal Cancer 1 Gene by Promoter Methylation in Gastric and Colorectal Adenocarcinoma. *[Journal Article]Hepatogastroenterology* 57 (104), 1614–1619. doi:10.4149/neo_2010_03_228

Zhang, Y., Zhou, H., Sun, H., Chen, J., Huang, D., Han, X., et al. (2018). Association of Peripheral Blood Leukocyte KIBRA Methylation with Gastric Cancer Risk: A Case-Control Study. *Cancer Med.* 7 (6), 2682–2690. doi:10.1002/cam4.1474

Zhao, J., Liang, Q., Cheung, K. F., Kang, W., Lung, R. W., Tong, J. H., et al. (2013). Genome-wide Identification of Epstein-Barr Virus-Driven Promoter Methylation Profiles of Human Genes in Gastric Cancer Cells. *Cancer-am. Cancer Soc.* 119 (2), 304–312. doi:10.1002/cncr.27724

Zheng, Y., Chen, L., Li, J., Yu, B., Su, L., Chen, X., et al. (2011a). Hypermethylated DNA as Potential Biomarkers for Gastric Cancer Diagnosis. *Clin. Biochem.* 44 (17-18), 1405–1411. doi:10.1016/j.clinbiochem.2011.09.006

Zheng, Y., Zhang, Y., Huang, X., and Chen, L. (2011b). Analysis of the RUNX3 Gene Methylation in Serum DNA from Esophagus Squamous Cell Carcinoma, Gastric and Colorectal Adenocarcinoma Patients. *Hepatogastroenterology* 58 (112), 2007–2011. doi:10.5754/hge10016

Zhou, H., Sun, H., Liu, X., Chen, J., Zhang, L., Lin, S., et al. (2019). Combined Effect between WT1 Methylation and *Helicobacter pylori* Infection, Smoking, and Alcohol Consumption on the Risk of Gastric Cancer. *Helicobacter* 24 (5), e12650. doi:10.1111/hel.12650

Zong, L., and Seto, Y. (2014). CpG Island Methylator Phenotype, *Helicobacter pylori*, Epstein-Barr Virus, and Microsatellite Instability and Prognosis in Gastric Cancer: A Systematic Review and Meta-Analysis. [Journal Article; Meta-Analysis; Review; Systematic Review]. *PLoS One* 9 (1), e86097. doi:10.1371/journal.pone.0086097

Zuo, J., Xia, J., Ju, F., Yan, J., Zhu, A., Jin, S., et al. (2013). MicroRNA-148a Can Regulate Runt-Related Transcription Factor 3 Gene Expression via Modulation of DNA Methyltransferase 1 in Gastric Cancer. *Mol. Cell* 35 (4), 313–319. doi:10.1007/s10059-013-2314-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fgene.2022.833923

F2RL3 Methylation in the Peripheral Blood as a Potential Marker for the Detection of Coronary Heart Disease: A Case-Control Study

Xiaojing Zhao[1,2†], Liya Zhu[3†], Qiming Yin[3], Zhenguo Xu[4,5], Qian Jia[1,2], Rongxi Yang[3] and Kunlun He[1,2]**

[1]*Military Translational Medicine Lab, Medical Innovation Research Division, Chinese PLA General Hospital, Beijing, China,* [2]*Beijing Key Laboratory of Chronic Heart Failure Precision Medicine, Medical Innovation Research Division, Chinese PLA General Hospital, Beijing, China,* [3]*Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China,* [4]*The First Medical Center, Chinese PLA General Hospital, Beijing, China,* [5]*The Medical School of Chinese PLA, Beijing, China*

OPEN ACCESS

Edited by:
Nejat Dalay,
Istanbul University, Turkey

Reviewed by:
Silvio Zaina,
University of Guanajuato, Mexico
Dao Wen Wang,
Huazhong University of Science and Technology, China

***Correspondence:**
Rongxi Yang
rongxiyang@njmu.edu.cn
Kunlun He
kunlunhe@plagh.org

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *12 December 2021*
Accepted: *14 February 2022*
Published: *24 March 2022*

Citation:
Zhao X, Zhu L, Yin Q, Xu Z, Jia Q, Yang R and He K (2022) F2RL3 Methylation in the Peripheral Blood as a Potential Marker for the Detection of Coronary Heart Disease: A Case-Control Study. Front. Genet. 13:833923. doi: 10.3389/fgene.2022.833923

Background and Aims: Previous work has shown the association between blood-based methylation of coagulation factor II receptor-like 3 gene (*F2RL3*) and cardiovascular mortality in Caucasians. However, the diagnostic value of *F2RL3* methylation for CHD is still unknown. The aim of our study was to evaluate the association between blood-based *F2RL3* methylation and the risk of CHD in the Chinese population.

Methods: The methylation level of *F2RL3* was quantified by mass spectrometry in a case-control study with 180 CHD cases and 184 controls. The association between *F2RL3* methylation intensity and CHD was assessed by logistic regression models, controlling confounding factors.

Results: The hypomethylation in F2RL3_A amplicon was significantly associated with CHD (odds ratio (ORs) per -10% methylation: 1.22–1.42, $p < 0.035$ for six out of seven CpG loci). Specifically, this significant association was observed in elderly CHD patients (≥60 years), myocardial infarction (MI) patients, heart failure patients and the patients with minor to medium cardiac function impairment (NYHA I&II CHD cases) (ORs per -10% methylation: 1.35–1.58, 1.32–2.00, 1.29–1.43, 1.25–1.44; $p < 0.024$, 0.033, 0.035, 0.025, respectively). However, F2RL3_B CpG sites showed no or very weak association with CHD. The combination of F2RL3_A_CpG_1 and F2RL3_A_CpG_3 methylation levels could efficiently discriminate CHD, MI, heart failure, NYHA I&II CHD, and elderly CHD patients from controls (area under curve (AUC) = 0.75, 0.79, 0.75, 0.76, and 0.82, respectively).

Conclusion: We propose blood-based *F2RL3* methylation as a potential biomarker for CHD, especially for people with older age or with the status of MI. The combination of *F2RL3* methylation and conventional risk factors might be an approach to evaluate CHD at early stage.

Keywords: epigenomics, coronary heart disease, DNA methylation, coagulation factor II receptor-like 3 gene, biomarker, blood

INTRODUCTION

Coronary heart disease (CHD) is the leading cause of morbidity and mortality worldwide, producing immense health and economic burdens globally (Shaya et al., 2021; Virani et al., 2021). As a gene-environment interacted disease, CHD is characterized by endothelial dysfunction and chronic inflammation, and is mainly caused by atherosclerosis which progresses slowly and is usually asymptomatic in the early stage (Hansson, 2005; Talmud, 2007; Gatto and Prati, 2020). Currently available biomarkers, such as high-sensitivity C-reactive protein (hsCRP), interleukin-6, myeloperoxidase (MPO), pregnancy-associated plasma protein-A (PAPP-A), myeloperoxidase, leukocyte counts, are inadequate for the diagnosis of CHD due to their poor clinical practice (Danesh et al., 2004; Lobbes et al., 2010; Wang et al., 2017; Li et al., 2018). Recent studies have proposed plasma metabolomics and microRNAs as potential biomarkers for the diagnosis of CHD, but further validations with a larger sample size are still needed (Wang et al., 2017; Zhang et al., 2018; Fu et al., 2019). Nevertheless, these biomarkers are inadequate for the detection of early CHD due to their insufficient clinical practice. The identification and development of novel biomarkers are necessary and urgent for the early detection of CHD.

The term epigenetics is defined as changes in gene expression without altering the DNA sequence itself. Epigenetic silencing can mimic genetic mutations by impairing the expression of a gene, and aberrant epigenetic signatures are known as disease-related (Laird, 2003). DNA methylation is one of the most important epigenetic signatures, having critical roles in the control of gene activities and the architecture of the nucleus of the cells (Robertson and Wolffe, 2000; Weber et al., 2005). Unlike mutations and other genetic abnormalities, epigenetic modifications are reversible and could be modified by lifestyles and therapeutic methods (Arasaradnam et al., 2008). In recent years, epigenetic aspects are believed to play a significant role in cardiovascular biology with various epigenetic mechanisms involved in the physiological and pathophysiological vascular differentiation, proliferation, and related inflammatory processes (Schleithoff et al., 2012). In addition, the associations between CHD and blood-based hypermethylation of several genes, such as *FOXP3* (forkhead box P3), *ABCG1* (ATP binding cassette subfamily G member 1) and *GALNT2* (polypeptide N-acetylgalactosaminyltransferase 2), and hypomethylation of *IL-6* (interleukin 6) have been reported (Jia et al., 2013; Peng et al., 2014; Zuo et al., 2016). Therefore, DNA methylation in blood could be a potential biomarker for the detection of CHD.

Protease-activated receptors (PARs) are a group of receptors that could promote inflammation in intimal tissue, enhance the initiation of atherosclerotic plaques (Coughlin, 2000), and induce vascular smooth muscle proliferation, migration, and collagen synthesis leading to plaque progression (Bretschneider et al., 1999; Wei et al., 2019). Also, PAR-mediated platelet activation may play a significant role in the plaque complications (thrombosis) and allows adhesion to atherosclerotic lesions, involving in the recruitment of monocytes and lymphocytes, and thus undermines plaque stability (Vorchheimer and Becker, 2006). Consequently, arteries may be blocked, leading to acute ischemic events such as acute coronary syndrome (ACSS), stroke, and transient ischemic attack (Vorchheimer and Becker, 2006). There are four known PARs subtypes PAR-1, PAR-2, PAR-3, and PAR-4, which are expressed in various cell types of the cardiovascular system including platelets, endothelial cells, and smooth muscle cells (Coughlin, 2000; Leger et al., 2006). PAR4, coded by the *F2RL3* gene (coagulation factor II receptor-like 3), is a member of the protease-activated receptor subfamily and is known to be expressed in the leukocytes (Vergnolle et al., 2002). Evidence shows that over-expression of the wild-type PAR4 is correlated with a higher sensitivity of cardiomyocytes to apoptosis (Kolpakov et al., 2016). Breitling et al. (2011) first disclosed decreased methylation of cg03636183 at *F2RL3* in the blood of heavy smokers. Later, one prospective study indicated the hypomethylation of blood-based *F2RL3* at cg03636183 cytidine-phosphate-guanosine (CpG) loci and five adjacent CpG sites upstream to be strongly related to the mortality among patients with stable coronary heart disease (Breitling et al., 2012). Another prospective cohort study has revealed the association between increased mortality of cardiovascular disease (CVD, defined by either physician-reported coronary heart disease or a self-reported history of myocardial infarction, stroke, pulmonary embolism, or revascularization of coronary arteries) and the hypomethylation of blood-based *F2RL3* at the cg03636183 CpG loci and other three flanking CpG sites upstream (Zhang et al., 2014b). An epigenome-wide study further suggested that the methylation levels of *F2RL3* not only at the cg03636183 loci but also at the cg24704287 loci were associated with cardiovascular disease mortality (Zhang et al., 2017). Follow-up studies, however, have not been reported about the diagnostic value of *F2RL3* methylation for CHD in any population, especially for early CHD.

To investigate the relationship between CHD and the blood-derived methylation of *F2RL3* in the Chinese population, we hereby performed a case-control study with 180 CHD patients and 184 healthy individuals, aiming to evaluate the associations between the methylation intensities and the status of CHD diseases, lifestyles, and historical treatments. Two amplicons based on cg03636183 and cg24704287 respectively were designed by EpiDesigner and analyzed by mass spectrometry.

MATERIALS AND METHODS

Study Population

A total of 180 patients with CHD and 184 controls were collected from the Chinese PLA General Hospital from 2018 to 2019. All the CHD cases were confirmed according to the coronary angiography of the disease combined with clinical manifestations. Among the 180 CHD cases, 78 had MI, and 145 experienced heart failure. New York Heart Association (NYHA) (Yancy et al., 2017) cardiac function classifications were available from 161 CHD cases (NYHA Ⅰ CHD cases = 46; NYHA Ⅱ CHD cases = 78; NYHA Ⅲ CHD cases = 31; NYHA Ⅳ CHD cases = 6). CHD-free participants who participated in an annual health examination were randomly selected as controls. CHD cases and controls were matched by gender. The median age of patients with CHD was 66 years (58–73 years). Since the

TABLE 1 | Baseline characteristics of coronary heart disease (CHD) patients and healthy controls.

Clinical characteristics	Group	Controls (*N* = 184)	CHD cases (*N* = 180)	χ^2	*p*-value[a]
		N (%)	N (%)		
Gender	Female	70 (38.0)	71 (39.4)		
	Male	114 (62.0)	109 (60.6)	0.08	0.784
Smoking	No	127 (69.0)	107 (59.4)		
	Yes	53 (28.8)	73 (40.6)		
	Unknown	4 (2.2)	0 (0.0)	4.88	**0.027**
Drinking	No	114 (61.9)	128 (71.1)		
	Yes	66 (35.9)	52 (28.9)		
	Unknown	4 (2.2)	0 (0.0)	2.47	0.116
Hypertension	No	93 (50.5)	50 (27.8)		
	Yes	84 (45.7)	130 (72.2)		
	Unknown	7 (3.8)	0 (0.0)	22.79	**2.00E-06**
Diabetes	No	132 (71.7)	118 (65.6)		
	Yes	45 (24.5)	62 (34.4)		
	Unknown	7 (3.8)	0 (0.0)	3.46	0.063

Clinical characteristics	Controls (*N* = 184)	CHD cases (*N* = 180)	Z	*p*-value[b]
	Median (IQR)	Median (IQR)		
Age	63 (57–68)	66 (58–73)	−2.67	**0.008**
Total cholesterol (TC) (mmol/L)	4.26 (3.61–5.06)	3.81 (3.25–4.42)	−3.41	**0.001**
Triglyceride (TG) (mmol/L)	1.39 (1.06–2.17)	1.30 (0.95–1.86)	−1.45	0.147
High density lipoprotein (HDL) (mmol/L)	1.15 (0.91–1.36)	1.09 (0.91–1.31)	−1.06	0.290
Low density lipoprotein (LDL) (mmol/L)	2.69 (2.06–3.40)	2.28 (1.82–2.87)	−3.70	**2.00E-04**

[a]*The p-values were calculated by the Chi-square test, and significant p-values are in bold.*
[b]*The p-values were calculated by the Mann-Whitney test, and significant p-values are in bold.*

controls were recruited from the health examination center where most participants were under 70 years old, and thus the median age of controls was 63 years (57–68 years). All controls were self-report healthy, without a history of CHD, cancer, autoimmune diseases, and had normal blood accounts. No further exclusion or inclusion criteria were implemented for the controls. The detailed clinical characteristics of CHD cases and controls are listed in **Table 1**.

Sample Collection and Processing

Peripheral whole blood from CHD cases and healthy controls were deposited into the ethylene diamine tetraacetic acid (EDTA) tubes and kept at 4°C for up to 8 h before storing at −80°C till further usage. Genomic DNA was extracted from each sample using the Genomic DNA Extraction Kit (Zymo Research, Orange County, United States). Subsequently, DNA was bisulfite converted by the EZ-96 DNA Methylation Gold Kit according to the manufacturer's instruction (Zymo Research, Orange County, United States).

Agena Matrix-Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry

Agena MALDI-TOF mass spectrometry (Agena Bioscience, San Diego, California, United States) described by Yang et al. (2015), was used for the quantification of DNA methylation levels. Procedures of methylation assessment and quality controls have been described previously (Zhang et al., 2014b). The cg03636183 and cg24704287 loci reported by Zhang et al. (2017), are located at 19p13.11 (chr19:17,000,586, at the second exon of *F2RL3*) and 19p13.13 (chr19:13,951,482, at the 5′ upstream of *F2RL3*), respectively. We therefore designed two amplicons: F2RL3_A amplicon (206 bp, chr19:17,000,421-17,000,626) covers CpG cg03636183; F2RL3_B amplicon (377 bp, chr19:13951024-13951400) covers five adjacent CpG sites of cg24704287 since the amplicons covering cg24704287 are unstable for PCR. The schematic diagram and the sequence of amplicons are presented in **Supplementary Figure S1**. SNPs are located neither at the primer regions nor overlapped with any CpGs in the two amplicons (F2RL3_A, F2RL3_B). The EpiTyper assay determined the methylation levels of 7 CpGs in F2RL3_A amplicon and yielded 7 distinguishable mass peaks, and determined the methylation levels of 5 CpGs in F2RL3_B amplicon and yielded 4 distinguishable mass peaks. F2RL3_B_CpG_4 and F2RL3_B_CpG_5 are located at the same fragment after the EpiTyper treatment, and thus the mass peak shows the average methylation level of F2RL3_B_CpG_4 and F2RL3_B_CpG_5 (presented as F2RL3_B_CpG_4.5). Briefly, the bisulfite-converted DNA was amplified by bisulfite-specific primers. The polymerase chain reaction (PCR) products were treated in the light of the standard protocol of Agena EpiTyper Assay by shrimp alkaline phosphatase (SAP) treatment and RNAse A cleavage

TABLE 2 | Overall and age-specific methylation difference of *F2RL3* comparing CHD cases and controls.

A. Overall

CpG sites	Controls (N = 184)	CHD cases (N = 180)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.71 (0.49–0.82)	0.62 (0.46–0.78)	1.12 (1.01–1.24)	**0.028**	1.12 (1.01–1.24)	**0.029**	1.22 (1.08–1.38)	**0.001**
F2RL3_A_CpG_2/ cg03636183	0.83 (0.79–0.87)	0.82 (0.75–0.87)	1.22 (0.98–1.53)	0.078	1.23 (0.98–1.55)	0.079	1.32 (1.02–1.71)	**0.035**
F2RL3_A_CpG_3	0.71 (0.66–0.77)	0.71 (0.55–0.79)	1.29 (1.11–1.51)	**0.001**	1.33 (1.13–1.55)	**4.28E-04**	1.41 (1.18–1.68)	**1.57E-04**
F2RL3_A_CpG_4	0.66 (0.59–0.70)	0.62 (0.55–0.68)	1.34 (1.09–1.65)	**0.005**	1.36 (1.10–1.68)	**0.005**	1.42 (1.12–1.81)	**0.004**
F2RL3_A_CpG_5	0.86 (0.82–0.89)	0.85 (0.76–0.90)	1.25 (1.00–1.56)	0.053	1.24 (0.99–1.57)	0.064	1.26 (0.97–1.63)	0.080
F2RL3_A_CpG_6	0.66 (0.59–0.70)	0.62 (0.55–0.68)	1.34 (1.09–1.65)	**0.005**	1.36 (1.10–1.68)	**0.005**	1.42 (1.12–1.81)	**0.004**
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.66 (0.52–0.72)	1.26 (1.06–1.48)	**0.007**	1.27 (1.07–1.50)	**0.007**	1.29 (1.07–1.56)	**0.008**
F2RL3_B_CpG_2	0.67 (0.57–0.81)	0.71 (0.63–0.82)	0.89 (0.78–1.00)	**0.048**	0.87 (0.77–0.99)	**0.029**	0.86 (0.75–0.99)	**0.042**
F2RL3_B_CpG_4.5	0.10 (0.08–0.14)	0.10 (0.06–0.14)	1.27 (0.91–1.78)	0.163	1.23 (0.88–1.74)	0.229	1.16 (0.85–1.58)	0.354
F2RL3_B_CpG_6	0.45 (0.39–0.50)	0.44 (0.36–0.51)	1.11 (0.94–1.31)	0.206	1.10 (0.93–1.30)	0.272	1.09 (0.91–1.30)	0.365
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.04 (0.01–0.06)	1.94 (1.11–3.38)	**0.020**	1.97 (1.12–3.47)	**0.019**	1.92 (1.05–3.54)	**0.035**

B. Age < 60 years

CpG sites	Controls (*N* = 51)	CHD cases (*N* = 53)	Model 1 [a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.54 (0.43–0.70)	0.72 (0.47–0.82)	0.84 (0.70–1.02)	0.081	0.85 (0.70–1.04)	0.105	0.82 (0.64–1.05)	0.111
F2RL3_A_CpG_2/ cg03636183	0.84 (0.78–0.88)	0.84 (0.76–0.88)	1.13 (0.76–1.67)	0.551	1.08 (0.72–1.61)	0.711	1.13 (0.69–1.84)	0.634
F2RL3_A_CpG_3	0.72 (0.67–0.78)	0.71 (0.45–0.79)	1.36 (1.05–1.76)	**0.019**	1.34 (1.03–1.74)	**0.029**	1.61 (1.14–2.29)	**0.007**
F2RL3_A_CpG_4	0.67 (0.60–0.72)	0.63 (0.57–0.69)	1.44 (0.98–2.10)	0.064	1.38 (0.92–2.07)	0.119	1.57 (0.96–2.56)	0.070
F2RL3_A_CpG_5	0.86 (0.82–0.90)	0.85 (0.78–0.90)	1.48 (0.94–2.35)	0.094	1.44 (0.90–2.30)	0.124	1.66 (0.93–2.95)	0.085
F2RL3_A_CpG_6	0.67 (0.60–0.72)	0.63 (0.57–0.69)	1.44 (0.98–2.10)	0.064	1.38 (0.92–2.07)	0.119	1.57 (0.96–2.56)	0.070
F2RL3_A_CpG_7	0.68 (0.59–0.76)	0.67 (0.52–0.72)	1.30 (0.96–1.75)	0.090	1.25 (0.91–1.73)	0.173	1.34 (0.89–2.00)	0.159
F2RL3_B_CpG_2	0.77 (0.69–0.85)	0.70 (0.63–0.82)	1.01 (0.82–1.24)	0.947	1.02 (0.82–1.26)	0.882	1.01 (0.65–1.57)	0.964
F2RL3_B_CpG_4.5	0.12 (0.07–0.15)	0.10 (0.08–0.14)	1.20 (0.77–1.87)	0.433	1.18 (0.77–1.82)	0.451	1.34 (0.80–2.27)	0.267
F2RL3_B_CpG_6	0.47 (0.40–0.57)	0.44 (0.37–0.49)	1.24 (0.95–1.62)	0.120	1.23 (0.93–1.62)	0.144	1.36 (0.80–2.32)	0.255
F2RL3_B_CpG_7	0.05 (0.02–0.07)	0.03 (0.00–0.06)	3.16 (1.10–9.13)	**0.033**	2.88 (0.97–8.56)	0.058	1.54 (0.68–3.48)	0.296

C. Age ≥ 60 years

CpG sites	Controls (*N* = 133)	CHD cases (*N* = 127)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.76 (0.57–0.83)	0.61 (0.45–0.78)	1.27 (1.12–1.44)	**2.58E-04**	1.26 (1.10–1.44)	**0.001**	1.49 (1.25–1.77)	**8.00E-06**
F2RL3_A_CpG_2/ cg03636183	0.83 (0.79–0.87)	0.81 (0.74–0.86)	1.28 (0.97–1.68)	0.079	1.42 (1.05–1.90)	**0.021**	1.58 (1.12–2.21)	**0.009**
F2RL3_A_CpG_3	0.70 (0.66–0.77)	0.71 (0.58–0.80)	1.25 (1.03–1.52)	**0.025**	1.31 (1.07–1.61)	**0.009**	1.35 (1.08–1.70)	**0.009**
F2RL3_A_CpG_4	0.65 (0.59–0.70)	0.62 (0.54–0.68)	1.31 (1.03–1.67)	**0.030**	1.37 (1.05–1.78)	**0.021**	1.41 (1.05–1.91)	**0.024**
F2RL3_A_CpG_5	0.85 (0.81–0.89)	0.85 (0.76–0.89)	1.18 (0.92–1.52)	0.198	1.22 (0.93–1.61)	0.158	1.18 (0.88–1.59)	0.275
F2RL3_A_CpG_6	0.65 (0.59–0.70)	0.62 (0.54–0.68)	1.31 (1.03–1.67)	**0.030**	1.37 (1.05–1.78)	**0.021**	1.41 (1.05–1.91)	**0.024**
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.66 (0.52–0.72)	1.24 (1.02–1.51)	**0.034**	1.24 (1.00–1.54)	**0.046**	1.24 (0.98–1.57)	0.070
F2RL3_B_CpG_2	0.62 (0.55–0.77)	0.72 (0.62–0.82)	0.83 (0.72–0.97)	**0.017**	0.83 (0.71–0.97)	**0.018**	0.90 (0.75–1.07)	0.238
F2RL3_B_CpG_4.5	0.10 (0.08–0.13)	0.09 (0.06–0.12)	1.37 (0.86–2.17)	0.184	1.34 (0.83–2.18)	0.237	1.24 (0.74–2.10)	0.417
F2RL3_B_CpG_6	0.44 (0.38–0.49)	0.44 (0.35–0.52)	1.04 (0.84–1.29)	0.722	0.98 (0.78–1.24)	0.881	1.03 (0.80–1.32)	0.825
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.04 (0.01–0.06)	1.55 (0.80–3.03)	0.198	1.39 (0.70–2.79)	0.350	1.31 (0.61–2.81)	0.489

[a]*Model 1: Logistic regression without adjustment.*

[b]*Model 2: Logistic regression adjusted for age and gender.*

[c]*Model 3: Logistic regression adjusted for age, gender, smoking, hypertension, TC, LDL, and batch effect. Significant p-values are in bold.*

(so-called "T-cleavage") reaction. The samples were further cleaned by resin and then dispensed to a 384 SpectroCHIP using Nanodispenser. The chips were read by a MassARRAY system. Data were obtained by Spectro ACQUIRE v3.3.1.3 software and visualized with MassARRAY EpiTyper v1.2 software. For each batch of MassARRAY analysis, an equal number of cases and controls were treated and analyzed in parallel in all the processes.

Statistical Analyses

All the statistical analyses were conducted by SPSS Statistics 25. Spearman's rank correlation coefficient was carried out to evaluate the correlations. Differences between cases and controls were tested by non-parametric tests. ORs and 95% confidence intervals (CIs) were estimated by logistic regression models adjusted for covariates, especially for the significant covariates as indicated in **Table 1**. Cardiovascular-related quantitative variables are classified using appropriate cutoff values (Jellinger et al., 2017), including TC (5.0 mmol/L), TG (1.7 mmol/L), HDL (1.0 mmol/L), and LDL (3.0 mmol/L). ROC curve analysis was performed to assess the discriminatory power of altered *F2RL3* methylation levels for the diagnosis of CHD. The corresponding area under curve was calculated with 95% CIs. The statistical power was calculated by independent *t*-test using Power and Simple size software (http://powerandsamplesize.com/). All statistical tests were two-sided, and *p*-values less than 0.05 were defined as statistically significant.

RESULTS

Blood-Based *F2RL3* Hypomethylation is Associated With CHD

In this study, we quantitatively determined the methylation levels of *F2RL3* in the blood DNA of the 180 CHD patients and 184 controls using Agena MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometry. Two amplicons in *F2RL3*, namely F2RL3_A amplicon (harboring seven measurable CpG sites) and F2RL3_B amplicon (harboring five measurable CpG sites), were amplified and analyzed. Three logistic regression models adjusted for different covariants were performed to investigate the association between *F2RL3* methylation and the status of CHD (**Table 2**). Among which, all the baseline characteristics that had significant differences between the CHD cases and the controls (as listed in **Table 1**) were adjusted in the logistic regression model 3. Six out of the seven CpG loci in the F2RL3_A amplicon showed significantly lower methylation in the CHD cases than in the controls according to the logistic regression model 3 (Odds ratios (ORs) per -10% methylation ranging from 1.22 to 1.42, $p < 0.035$ for all by logistic regression adjusted for age, gender, smoking, hypertension, total cholesterol (TC) levels, low density lipoprotein (LDL) levels and batch effect; **Figure 1A** and **Table 2A**). Among the significant CpG loci, F2RL3_A_CpG_3 was the most significant one, and F2RL3_A_CpG_2/cg03636183 was the weakest (**Figure 1A** and **Table 2A**). Weak associations were also observed between two out of the five measurable CpG sites in the F2RL3_B amplicon and the CHD (F2RL3_B_CpG_2, ORs per -10% methylation = 0.86, *p*-value = 0.042; F2RL3_B_CpG_7, ORs per -10% methylation = 1.92, *p*-value = 0.035, logistic regression model 3; **Figure 1A** and **Table 2A**). The power for methylation difference is sufficient (power for F2RL3_A_CpG_1, F2RL3_A_CpG_3, F2RL3_A_CpG_4, F2RL3_A_CpG_5, F2RL3_A_CpG_6, F2RL3_A_CpG_7, F2RL3_B_CpG_2 and F2RL3_B_CpG_7 was 0.9003, 0.9954, 0.9721, 0.7826, 0.9721, 0.9838, 0.8739 and 0.9893, respectively). We also noticed that the methylation correlates better among close than among more distant CpG. More specific, the methylation correlates better among CpGs in the same amplicon than CpGs in different amplicons which have larger distance (**Supplementary Figure S2**). In addition, the methylation correlation among the CpG sites in the F2RL3_A amplicon is stronger than the correlation among the CpG sites in the F2RL3_B amplicon (**Supplementary Figure S2**).

The level of methylation has been known to be changed along with age (Horvath and Raj, 2018). We hereby stratified the subjects by the age of 60 years old. In the group younger than 60 years old, only one CpG site (F2RL3_A_CpG_3) was weakly associated with the CHD, whereas the other 11 measurable CpG sites showed no correlation (**Figure 1B** and **Table 2B**). In the group ≥60 years old, five out of the seven measurable CpG loci in F2RL3_A amplicon exhibited significantly lower methylation levels in the CHD cases than in the controls. Among which, F2RL3_A_CpG_1 showed the most significant difference (OR per -10% methylation = 1.49, $p = 8.00 \times 10^{-6}$ by logistic regression model 3; **Figure 1C** and **Table 2C**). In the F2RL3_B amplicon, none of the five measurable CpG loci indicated any association with CHD in people older than 60 years old (**Figure 1C** and **Table 2C**). When the subjects were stratified by 65 years old, we also found a similar pattern of the age-dependent *F2RL3* methylation for the risk of CHD (**Supplementary Table S1**). Moreover, the hypomethylation of F2RL3_A CpG sites showed even larger ORs for per -10% methylation in the group ≥65 years old than in the group ≥60 years old (**Table 2C** and **Supplementary Table 1B**), suggesting that age is a cofounder of the *F2RL3* hypomethylation associated risk for the CHD.

Decreased *F2RL3* Methylation is Mainly Associated With MI

Among the 180 CHD patients, 78 experienced myocardial infarction (MI). Thus, we further investigated whether MI played a role in the CHD associated *F2RL3* methylation in the blood. The F2RL3_A_CpG_3 site showed the most significant hypomethylation in the MI cases than in the controls (MI cases: median (interquartile range (IQR)) = 0.56 (0.44–0.80); controls: median (IQR) = 0.71 (0.66–0.77); OR per -10% methylation = 2.00, $p = 6.59 \times 10^{-8}$ by logistic regression adjusted for age, gender, smoking, hypertension, TC levels, LDL levels and batch effect; **Figure 2A** and **Table 3A**). Four additional CpG sites in the F2RL3_A amplicon located at the upstream of F2RL3_A_CpG_3 also exhibited significantly lower methylation levels in the MI cases than in the controls (OR per -10% methylation ranging

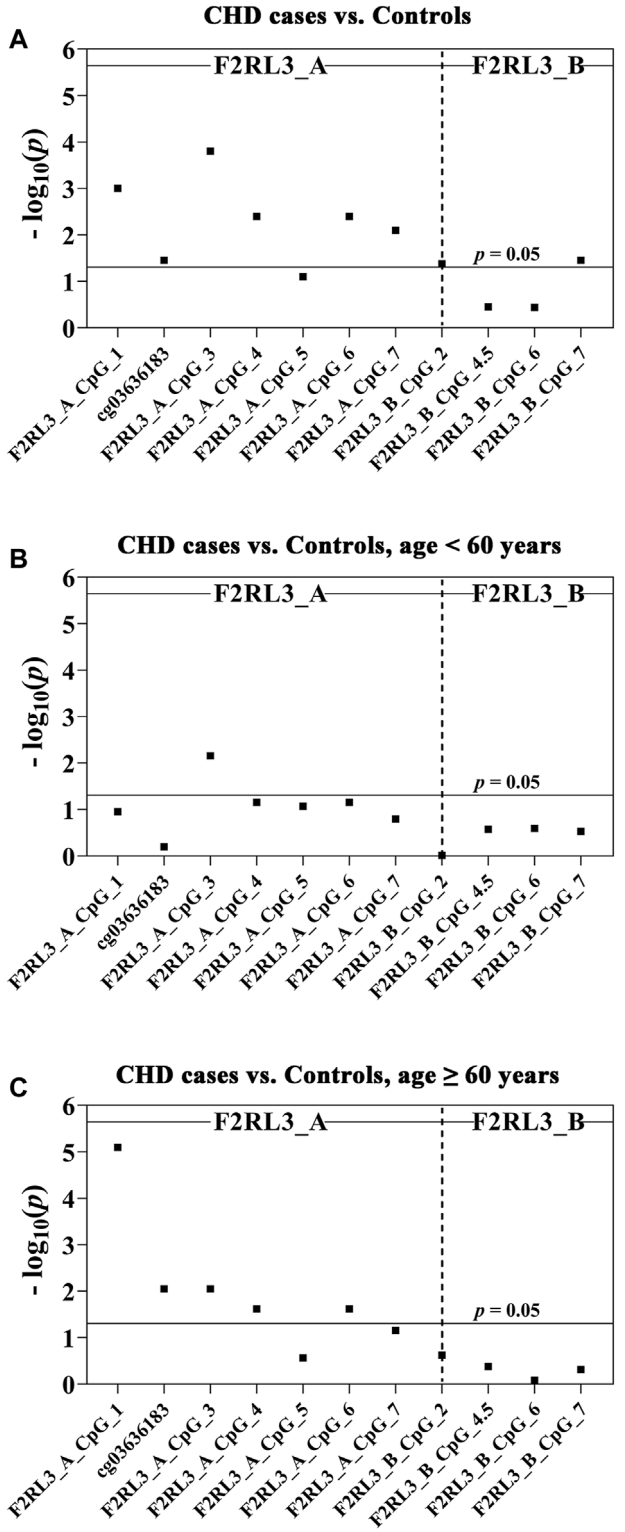


FIGURE 1 | The association between CHD and decreased *F2RL3* methylation in the peripheral blood. The point plots show the methylation differences of CpG sites in the amplicon of F2RL3_A and F2RL3_B **(A)** between 184 controls and 180 CHD cases, **(B)** between 51 controls and 53 CHD cases, age <60 years, **(C)** between 133 controls and 127 CHD cases, age ≥60 years. The *p*-values of all the 12 measurable CpG loci in the two amplicons were calculated by logistic regression adjusted for age, gender, smoking, hypertension, TC, LDL, and batch effect. The vertical dashed line separates the two amplicons. The solid lines indicate the thresholds of *p*-value = 0.05.

from 1.32 to 1.51, $p < 0.033$ for all by logistic regression adjusted for covariant; **Figure 2A** and **Table 3A**). The methylation levels of F2RL3_A_CpG_1 and F2RL3_A_CpG_2/cg03636183 were also lower in the MI cases than in the controls but without significance (**Figure 2A** and **Table 3A**). In the F2RL3_B amplicon, only one out of the five measurable CpG sites showed significant association with MI (F2RL3_B_CpG_7, ORs per -10% methylation = 2.99, $p = 0.019$ by logistic regression adjusted for covariant; **Figure 2A** and **Table 3A**). We further evaluated the *F2RL3* methylation difference between the non-MI CHD cases and the controls. Interestingly, unlike the MI cases, the only altered methylation for the non-MI CHD cases compared to the controls was detected in the F2RL3_A_CpG_1 site (OR per -10% methylation = 1.44, $p = 1.60 \times 10^{-5}$ by logistic regression adjusted for covariant; **Figure 2B** and **Table 3B**).

The Difference of *F2RL3* Methylation Level Between Heart Failure Cases Versus Controls

There were 145 heart failure cases in the 180 CHD patients. Hence, we also investigated the association between *F2RL3* methylation and the status of heart failure (**Table 3C**). Here, six out of the seven CpG loci in the F2RL3_A amplicon revealed significantly lower methylation levels in the heart failure cases than in the controls (OR per -10% methylation ranging from 1.29 to 1.43, $p < 0.035$ for all by logistic regression adjusted for age, gender, smoking, hypertension, TC levels, LDL levels and batch effect; **Figure 2C** and **Table 3C**). Among the significant CpG loci, F2RL3_A_CpG_1 showed the most significant difference (OR per -10% methylation = 1.30, $p = 7.30 \times 10^{-5}$ by logistic regression adjusted for covariant; **Figure 2C** and **Table 3C**). In the F2RL3_B amplicon, on the contrary, none of the five measurable CpG loci displayed any association with heart failure (**Figure 2C** and **Table 3C**). Since there are only 35 CHD patients without heart failure, we did not investigate the *F2RL3* methylation difference between the non-heart failure CHD cases and the controls using logistic regression analysis.

F2RL3 Methylation Difference Between NYHA I&II CHD Cases Versus Controls

Among the 180 CHD patients, patients with minor to medium cardiac function impairment (NYHA I&II CHD cases) were available from 124 CHD cases (NYHA I CHD cases = 46, NYHA II CHD cases = 78). Compared to the healthy controls, the status of *F2RL3* methylation was also associated with NYHA I&II CHD cases. The methylation level of five out of the seven CpG loci in the F2RL3_A amplicon was also significantly decreased for NYHA I&II CHD cases with ORs >1.25 per

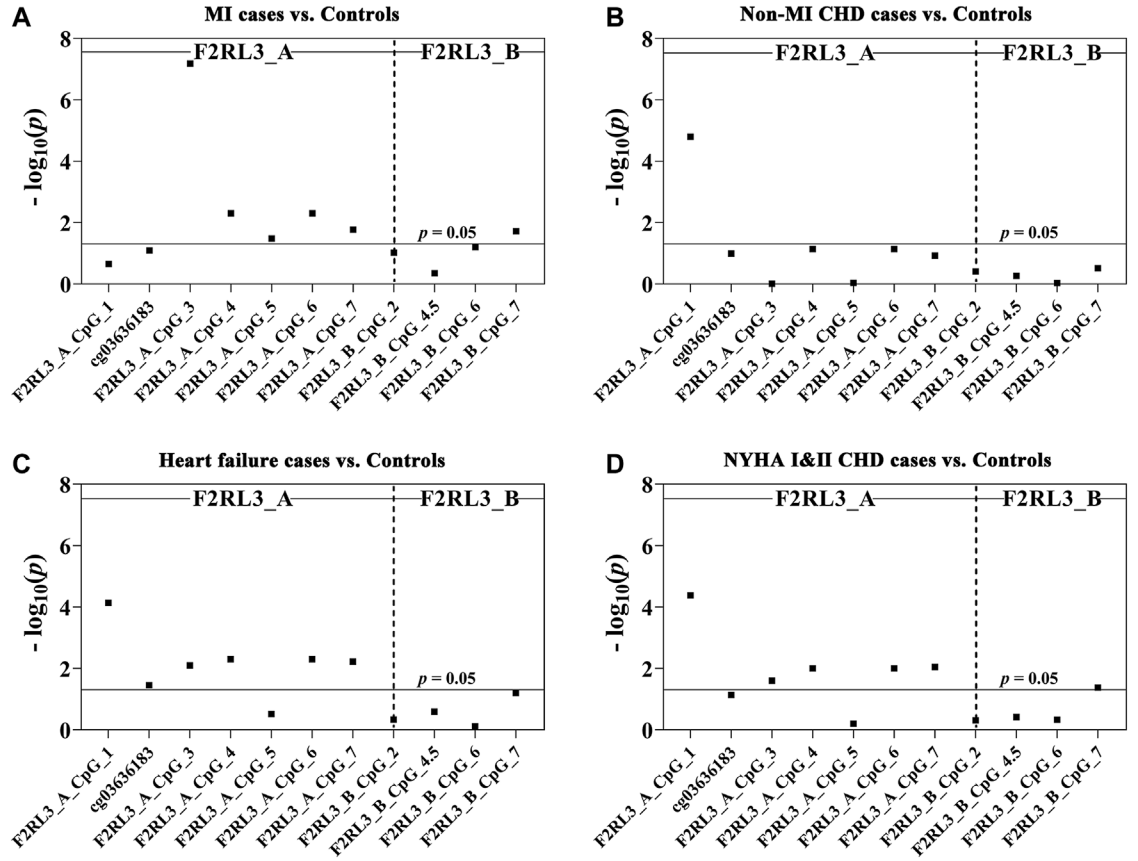


FIGURE 2 | The association between decreased *F2RL3* methylation in the peripheral blood and MI cases, non-MI CHD cases, heart failure cases, and patients with minor to medium cardiac function impairment (NYHA I&II CHD cases). The point plots show the methylation differences of CpG sites in the amplicon of F2RL3_A and F2RL3_B **(A)** between 184 controls and 78 MI cases, **(B)** between 184 controls and 102 non-MI CHD cases, **(C)** between 184 controls and 145 heart failure cases, **(D)** between 184 controls and 124 NYHA I&II CHD cases. The *p*-values of all the 12 measurable CpG loci in the two amplicons were calculated by logistic regression adjusted for age, gender, smoking, hypertension, TC, LDL, and batch effect. The vertical dashed line separates the two amplicons. The solid lines indicate the thresholds of *p*-value = 0.05.

-10% methylation ($p < 0.025$ for all by logistic regression adjusted for age, gender, smoking, hypertension, TC levels, LDL levels and batch effect; **Figure 2D** and **Table 3D**). Here, the most significant locus was F2RL3_A_CpG_1 with an OR of 1.35 per -10% methylation and a *p*-value of 4.20×10^{-5} (**Figure 2D** and **Table 3D**). Only one out of five measurable CpG sites in the F2RL3_B amplicon exhibited a weak significant association with the NYHA I&II CHD cases (F2RL3_B_CpG_7, ORs per -10% methylation = 2.11, $p = 0.042$, logistic regression adjusted for covariant; **Figure 2D** and **Table 3D**). Although the sample size of NYHA III&IV CHD cases was very small (only 37 cases), to investigate the correlation between *F2RL3* methylation and the level of cardiac function impairment, logistic regression was also applied. Compared with the healthy controls, NYHA III&IV CHD patients showed significantly decreased methylation in the F2RL3_A amplicon (**Supplementary Table S2**). To note, the ORs of all significant F2RL3_A loci in the NYHA III&IV CHD patients were larger than that in the NYHA I&II CHD patients (**Figure 2D**; **Table 3D** and **Supplementary Table S2**). Additionally, we attempted to explore the methylation differences between NYHA I&II CHD cases and NYHA III&IV CHD cases using the Mann-Whitney test. The NYHA III&IV CHD cases showed significantly decreased methylation than the NYHA I&II CHD cases at the sites of F2RL3_A_CpG_3, F2RL3_A_CpG_5, and F2RL3_B_CpG_4.5 ($p < 0.032$ for all; **Supplementary Table S3**). This indicated that the aberrant *F2RL3* methylation would be enhanced along with the impairment of cardiac function.

F2RL3 Methylation and CHD-Related Characteristics

To explore the relationship between the blood-based *F2RL3* methylation and the CHD-related characteristics, the subjects (including both CHD cases and controls) with available data were interpreted. In agreement with previous reports (Breitling et al., 2011; Wan et al., 2012; Sun et al., 2013; Zhang et al., 2014a), smoking has a tremendous influence on the *F2RL3* methylation especially at the F2RL3_A amplicon which covers F2RL3_A_CpG_2/cg036361837 and six flanking CpG sites

TABLE 3 | *F2RL3* methylation in MI cases, non-MI CHD cases, heart failure cases, and patients with minor to medium cardiac function impairment (NYHA I&II CHD cases) compared to controls.

A. MI cases vs. controls

CpG sites	Controls (*N* = 184)	MI cases (*N* = 78)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.71 (0.49–0.82)	0.71 (0.46–0.80)	1.06 (0.94–1.21)	0.340	1.07 (0.94–1.22)	0.303	1.10 (0.94–1.28)	0.223
F2RL3_A_CpG_2/ cg03636183	0.83 (0.79–0.87)	0.82 (0.73–0.86)	1.39 (1.06–1.82)	**0.017**	1.28 (0.97–1.69)	0.079	1.31 (0.97–1.78)	0.081
F2RL3_A_CpG_3	0.71 (0.66–0.77)	0.56 (0.44–0.80)	1.72 (1.39–2.12)	**4.73E-07**	1.71 (1.39–2.12)	**6.77E-07**	2.00 (1.55–2.57)	**6.59E-08**
F2RL3_A_CpG_4	0.66 (0.59–0.70)	0.61 (0.48–0.68)	1.57 (1.22–2.02)	**0.001**	1.44 (1.11–1.87)	**0.006**	1.51 (1.13–2.02)	**0.005**
F2RL3_A_CpG_5	0.86 (0.82–0.89)	0.85 (0.71–0.89)	1.49 (1.14–1.96)	**0.004**	1.39 (1.06–1.82)	**0.019**	1.39 (1.03–1.88)	**0.033**
F2RL3_A_CpG_6	0.66 (0.59–0.70)	0.61 (0.48–0.68)	1.57 (1.22–2.02)	**0.001**	1.44 (1.11–1.87)	**0.006**	1.51 (1.13–2.02)	**0.005**
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.61 (0.46–0.71)	1.41 (1.15–1.74)	**0.001**	1.33 (1.07–1.64)	**0.009**	1.32 (1.05–1.67)	**0.017**
F2RL3_B_CpG_2	0.67 (0.57–0.81)	0.69 (0.61–0.78)	0.93 (0.79–1.08)	0.335	0.92 (0.78–1.08)	0.310	0.85 (0.71–1.03)	0.096
F2RL3_B_CpG_4.5	0.10 (0.08–0.14)	0.08 (0.06–0.14)	1.38 (0.85–2.22)	0.193	1.26 (0.80–2.00)	0.319	1.18 (0.78–1.77)	0.441
F2RL3_B_CpG_6	0.45 (0.39–0.50)	0.40 (0.29–0.49)	1.29 (1.03–1.61)	**0.025**	1.27 (1.02–1.59)	**0.035**	1.24 (0.99–1.56)	0.063
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.03 (0.00–0.06)	2.98 (1.32–6.69)	**0.008**	2.68 (1.15–6.23)	**0.022**	2.99 (1.20–7.47)	**0.019**

B. non-MI CHD cases vs. controls

CpG sites	Controls (*N* = 184)	non-MI CHD cases (*N* = 102)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.71 (0.49–0.82)	0.59 (0.47–0.76)	1.18 (1.04–1.33)	**0.011**	1.22 (1.07–1.39)	**0.003**	1.44 (1.22–1.70)	**1.60E-05**
F2RL3_A_CpG_2/ cg03636183	0.83 (0.79–0.87)	0.83 (0.77–0.87)	1.10 (0.83–1.47)	0.504	1.17 (0.86–1.58)	0.314	1.35 (0.94–1.92)	0.102
F2RL3_A_CpG_3	0.71 (0.66–0.77)	0.73 (0.63–0.79)	1.04 (0.84–1.31)	0.704	1.03 (0.82–1.30)	0.776	1.00 (0.77–1.29)	0.980
F2RL3_A_CpG_4	0.66 (0.59–0.70)	0.65 (0.57–0.69)	1.19 (0.92–1.53)	0.183	1.21 (0.92–1.59)	0.165	1.33 (0.97–1.82)	0.073
F2RL3_A_CpG_5	0.86 (0.82–0.89)	0.85 (0.80–0.90)	1.04 (0.79–1.36)	0.798	1.02 (0.76–1.35)	0.914	1.02 (0.75–1.38)	0.920
F2RL3_A_CpG_6	0.66 (0.59–0.70)	0.65 (0.57–0.69)	1.19 (0.92–1.53)	0.183	1.21 (0.92–1.59)	0.165	1.33 (0.97–1.82)	0.073
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.67 (0.55–0.73)	1.15 (0.94–1.41)	0.169	1.18 (0.94–1.47)	0.151	1.22 (0.95–1.56)	0.119
F2RL3_B_CpG_2	0.67 (0.57–0.81)	0.73 (0.65–0.83)	0.86 (0.74–0.99)	**0.035**	0.83 (0.71–0.96)	**0.014**	0.92 (0.77–1.11)	0.389
F2RL3_B_CpG_4.5	0.10 (0.08–0.14)	0.10 (0.07–0.13)	1.21 (0.81–1.81)	0.346	1.21 (0.77–1.89)	0.410	1.12 (0.77–1.64)	0.548
F2RL3_B_CpG_6	0.45 (0.39–0.50)	0.47 (0.38–0.52)	0.99 (0.80–1.24)	0.953	0.96 (0.76–1.21)	0.714	1.01 (0.78–1.31)	0.933
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.04 (0.02–0.06)	1.51 (0.81–2.81)	0.197	1.50 (0.78–2.88)	0.225	1.37 (0.75–2.51)	0.308

C. Heart failure cases vs. controls

CpG sites	Controls (*N* = 184)	Heart failure cases (*N* = 145)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.71 (0.49–0.82)	0.56 (0.42–0.78)	1.20 (1.07–1.33)	**0.001**	1.20 (1.08–1.34)	**0.001**	1.30 (1.14–1.47)	**7.30E-05**
F2RL3_A_CpG_2/ cg03636183	0.83 (0.79–0.87)	0.82 (0.76–0.86)	1.24 (0.98–1.58)	0.079	1.23 (0.96–1.58)	0.103	1.35 (1.02–1.79)	**0.035**
F2RL3_A_CpG_3	0.71 (0.66–0.77)	0.72 (0.59–0.81)	1.17 (0.99–1.37)	0.064	1.21 (1.02–1.43)	**0.028**	1.29 (1.07–1.55)	**0.008**
F2RL3_A_CpG_4	0.66 (0.59–0.70)	0.62 (0.55–0.68)	1.35 (1.09–1.68)	**0.006**	1.34 (1.07–1.67)	**0.010**	1.43 (1.12–1.83)	**0.005**
F2RL3_A_CpG_5	0.86 (0.82–0.89)	0.85 (0.78–0.90)	1.17 (0.92–1.49)	0.198	1.14 (0.89–1.46)	0.313	1.16 (0.88–1.53)	0.304
F2RL3_A_CpG_6	0.66 (0.59–0.70)	0.62 (0.55–0.68)	1.35 (1.09–1.68)	**0.006**	1.34 (1.07–1.67)	**0.010**	1.43 (1.12–1.83)	**0.005**
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.66 (0.51–0.72)	1.31 (1.10–1.56)	**0.003**	1.30 (1.08–1.56)	**0.006**	1.33 (1.08–1.62)	**0.006**
F2RL3_B_CpG_2	0.67 (0.57–0.81)	0.70 (0.62–0.81)	0.91 (0.80–1.04)	0.149	0.90 (0.79–1.02)	0.106	0.94 (0.81–1.10)	0.462
F2RL3_B_CpG_4.5	0.10 (0.08–0.14)	0.10 (0.06–0.13)	1.34 (0.91–1.98)	0.137	1.26 (0.86–1.86)	0.239	1.22 (0.86–1.73)	0.257
F2RL3_B_CpG_6	0.45 (0.39–0.50)	0.45 (0.37–0.52)	1.05 (0.88–1.25)	0.580	1.02 (0.85–1.22)	0.840	1.03 (0.85–1.25)	0.779
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.04 (0.01–0.06)	1.99 (1.09–3.63)	**0.025**	1.95 (1.06–3.61)	**0.033**	1.85 (0.97–3.55)	0.063

D. NYHA I&II CHD cases vs. controls

CpG sites	Controls (*N* = 184)	NYHA I&II CHD cases (*N* = 124)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_1	0.71 (0.49–0.82)	0.56 (0.40–0.78)	1.21 (1.08–1.36)	**0.001**	1.23 (1.09–1.38)	**0.001**	1.35 (1.17–1.55)	**4.20E-05**
	0.83 (0.79–0.87)	0.82 (0.77–0.87)	1.14 (0.88–1.47)	0.335	1.18 (0.90–1.55)	0.229	1.33 (0.97–1.81)	0.073

(Continued on following page)

TABLE 3 | (*Continued*) *F2RL3* methylation in MI cases, non-MI CHD cases, heart failure cases, and patients with minor to medium cardiac function impairment (NYHA I&II CHD cases) compared to controls.

D. NYHA I&II CHD cases vs. controls

CpG sites	Controls (*N* = 184)	NYHA I&II CHD cases (*N* = 124)	Model 1[a]		Model 2[b]		Model 3[c]	
	Median (IQR)	Median (IQR)	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value	OR (95%CI) per-10% methylation	*p*-value
F2RL3_A_CpG_2/ cg03636183								
F2RL3_A_CpG_3	0.71 (0.66–0.77)	0.73 (0.61–0.82)	1.12 (0.95–1.33)	0.178	1.16 (0.97–1.38)	0.098	1.25 (1.03–1.53)	**0.025**
F2RL3_A_CpG_4	0.66 (0.59–0.70)	0.63 (0.57–0.68)	1.27 (1.01–1.60)	**0.045**	1.32 (1.03–1.68)	**0.027**	1.44 (1.09–1.89)	**0.010**
F2RL3_A_CpG_5	0.86 (0.82–0.89)	0.87 (0.80–0.90)	0.93 (0.71–1.22)	0.611	0.94 (0.71–1.24)	0.640	0.93 (0.69–1.25)	0.629
F2RL3_A_CpG_6	0.66 (0.59–0.70)	0.63 (0.57–0.68)	1.27 (1.01–1.60)	**0.045**	1.32 (1.03–1.68)	**0.027**	1.44 (1.09–1.89)	**0.010**
F2RL3_A_CpG_7	0.69 (0.58–0.75)	0.66 (0.53–0.72)	1.25 (1.04–1.50)	**0.018**	1.29 (1.07–1.57)	**0.009**	1.34 (1.08–1.66)	**0.009**
F2RL3_B_CpG_2	0.67 (0.57–0.81)	0.71 (0.62–0.81)	0.90 (0.78–1.03)	0.114	0.88 (0.76–1.01)	0.074	0.94 (0.79–1.12)	0.493
F2RL3_B_CpG_4.5	0.10 (0.08–0.14)	0.10 (0.07–0.14)	1.17 (0.82–1.67)	0.398	1.15 (0.79–1.66)	0.471	1.16 (0.83–1.64)	0.387
F2RL3_B_CpG_6	0.45 (0.39–0.50)	0.45 (0.39–0.51)	1.03 (0.85–1.25)	0.741	1.03 (0.84–1.25)	0.799	1.08 (0.88–1.34)	0.467
F2RL3_B_CpG_7	0.04 (0.02–0.07)	0.04 (0.01–0.06)	2.10 (1.11–4.00)	**0.024**	2.09 (1.09–4.00)	**0.026**	2.11 (1.03–4.31)	**0.042**

[a]*Model 1: Logistic regression without adjustment.*

[b]*Model 2: Logistic regression adjusted for age and gender.*

[c]*Model 3: Logistic regression adjusted for age, gender, smoking, hypertension, TC, LDL, and batch effect. Significant p-values are in bold. MI, myocardial infarction; NYHA, new york heart association.*

(**Table 4**). In contrast, only one CpG locus in the F2RL3_B amplicon showed a borderline association with smoking in our study (**Table 4**). When stratified by the status of smoking, we unexpectedly found that the methylation of F2RL3_A amplicon was better associated with non-smokers than smokers with larger ORs and more significant *p*-values, whereas the methylation of F2RL3_B amplicon was associated with smokers with larger ORs than non-smokers **(Supplementary Table S4)**. The males showed lower methylation levels than the females but also mainly in the F2RL3_A amplicon (**Table 4**). As shown in **Table 4**, weak methylation differences were observed in a few CpG sites when stratified by age groups, diabetes, TC levels, and high density lipoprotein (HDL) levels. The methylation levels of all the 12 CpG sites in the F2RL3_A amplicon and F2RL3_B amplicon showed no correlation with drinking, hypertension, levels of triglyceride (TG), and LDL (**Table 4**).

F2RL3 Methylation as a Potential Biomarker for the Detection of CHD

Aiming to estimate the potential clinical utility of *F2RL3* methylation as a marker for the presence of CHD, receiver operating characteristic (ROC) curve analysis was performed adjusted for age, gender, smoking, hypertension, TC levels, LDL levels, and batch effect by logistic regression. Among all the investigated 11 distinguished *F2RL3* CpG groups, F2RL3_A_CpG_3 exhibited the best discriminatory power for general CHD cases, old than 60 years CHD cases, MI cases, heart failure cases, and NYHA I&II CHD cases from healthy controls (area under curve (AUC) = 0.71, 0.75, 0.79, 0.69 and 0.71, respectively; **Figures 3A,C,E,G,I**, **Supplementary Table S5**). The combination of F2RL3_A_CpG_1 and F2RL3_A_CpG_3 could improve the model, and dramatically elevate the efficiency for the distinguishing of CHD cases, old than 60 years CHD cases, MI cases, heart failure cases, and NYHA I&II CHD cases from healthy controls (AUC = 0.75, 0.82, 0.79, 0.75 and 0.76, respectively; **Figures 3B,D,F,H,J**, **Supplementary Table S5**). However, the discriminatory power could hardly be further improved when all additional CpG sites in F2RL3_A amplicon were included in the model (**Supplementary Table S5**).

DISCUSSION

CHD is a prevalent and chronic life-threatening disease. However, there is no reliable way for early detection and risk prediction of CHD so far. DNA methylation plays a critical role in the development of cardiovascular disease with the potential to predict fundamental pathogenic processes. The previous study has demonstrated CVD mortality-related *F2RL3* (cg03636183 and cg24704287) methylation in the Caucasian population (Zhang et al., 2017). The diagnostic value of *F2RL3* methylation for CHD has not been addressed in the different ethnic populations. In the present study, we analyzed blood-based *F2RL3* methylation levels in 180 CHD patients and 184 healthy subjects in the Chinese population and proposed *F2RL3* methylation, especially methylation at CpG sites adjacent to cg03636183, as an independent biomarker for the detection of CHD controlling variant CHD-related risk factors. Moreover, we also firstly disclosed that the aberrant *F2RL3* methylation is mainly correlated with CHD in elder people, MI status, and heart failure status, and could be detected when patients have minor to medium cardiac function impairment (NYHA I&II CHD cases).

The *F2RL3* gene encoding for PAR-4 has been shown to play a crucial role in mediating the activation of platelet (Kahn et al., 1999; Coughlin, 2000), and multiple signaling pathways, such as immune response, the regulation of vascular endothelial cell activity, and inflammatory reactions (Vergnolle et al., 2002; Kataoka et al., 2003; Steinhoff et al., 2005). Hypomethylation is usually associated with

TABLE 4 | The association between *F2RL3* methylation and CHD-related characteristics in the study subjects.

Characteristics (N)	Median (IQR) of methylation levels											
	Group (N)	F2RL3_A_CpG_1	F2RL3_A_CpG_2/ cg036361837	F2RL3_A_CpG_3	F2RL3_A_CpG_4	F2RL3_A_CpG_5	F2RL3_A_CpG_6	F2RL3_A_CpG_7	F2RL3_B_CpG_2	F2RL3_B_CpG_4.5	F2RL3_B_CpG_6	F2RL3_B_CpG_7
Age (380)	<60 (110)	0.60 (0.44–0.78)	0.84 (0.76–0.88)	0.70 (0.58–0.78)	0.65 (0.58–0.70)	0.86 (0.80–0.90)	0.65 (0.58–0.70)	0.67 (0.55–0.74)	0.74 (0.64–0.82)	0.11 (0.07–0.15)	0.45 (0.39–0.52)	0.04 (0.01–0.07)
	≥60 (270)	0.71 (0.49–0.82)	0.83 (0.76–0.87)	0.70 (0.63–0.78)	0.64 (0.56–0.69)	0.85 (0.79–0.89)	0.64 (0.56–0.69)	0.67 (0.56–0.74)	0.68 (0.58–0.78)	0.10 (0.07–0.13)	0.44 (0.37–0.50)	0.04 (0.02–0.06)
	p-value*	**0.042**	0.209	0.600	0.436	0.366	0.436	0.936	**0.002**	0.124	0.427	0.325
Gender (380)	Female (152)	0.72 (0.49–0.82)	0.84 (0.80–0.87)	0.71(0.62–0.78)	0.67 (0.61–0.71)	0.87 (0.84–0.91)	0.67 (0.61–0.71)	0.68 (0.62–0.75)	0.69 (0.58–0.82)	0.10 (0.07–0.14)	0.45 (0.38–0.52)	0.04 (0.02–0.07)
	Male (228)	0.67 (0.48–0.81)	0.81 (0.74–0.86)	0.69 (0.63–0.77)	0.63 (0.53–0.68)	0.84 (0.77–0.89)	0.63 (0.53–0.68)	0.65 (0.52–0.72)	0.69 (0.60–0.80)	0.09 (0.07–0.13)	0.43 (0.37–0.50)	0.04 (0.01–0.06)
	p-value*	0.316	**1.22E-04**	0.329	**1.60E-05**	**6.00E-06**	**1.60E-05**	**7.20E-05**	0.651	**0.044**	0.178	0.669
Smoking (376)	No (248)	0.70 (0.49–0.82)	0.84 (0.79–0.87)	0.71 (0.63–0.77)	0.66 (0.61–0.70)	0.87 (0.83–0.90)	0.66 (0.61–0.70)	0.69 (0.62–0.75)	0.69 (0.59–0.80)	0.10 (0.07–0.14)	0.45 (0.38–0.50)	0.04 (0.02–0.07)
	Yes (128)	0.65 (0.44–0.80)	0.79 (0.72–0.85)	0.69 (0.62–0.79)	0.59 (0.51–0.65)	0.82 (0.75–0.87)	0.59 (0.51–0.65)	0.58 (0.48–0.69)	0.70 (0.61–0.81)	0.09 (0.06–0.13)	0.44 (0.36–0.51)	0.04 (0.01–0.06)
	p-value*	0.123	**8.72E-07**	0.943	**6.97E-11**	**2.76E-08**	**6.97E-11**	**1.46E-09**	0.508	**0.046**	0.677	0.422
Drinking (376)	No (254)	0.70 (0.49–0.82)	0.83 (0.77–0.87)	0.70 (0.61–0.77)	0.65 (0.59–0.69)	0.86 (0.80–0.90)	0.65 (0.59–0.69)	0.67 (0.57–0.74)	0.68 (0.58–0.79)	0.10 (0.07–0.14)	0.45 (0.37–0.51)	0.04 (0.02–0.07)
	Yes (122)	0.67 (0.47–0.81)	0.83 (0.75–0.86)	0.72 (0.65–0.79)	0.62 (0.57–0.70)	0.85 (0.79–0.89)	0.62 (0.57–0.70)	0.66 (0.53–0.73)	0.71 (0.62–0.83)	0.10 (0.06–0.13)	0.44 (0.38–0.50)	0.04 (0.01–0.07)
	p-value*	0.415	0.310	0.104	0.187	0.279	0.187	0.282	0.052	0.500	0.923	0.807
Hypertension (373)	No (148)	0.67 (0.48–0.79)	0.82 (0.74–0.87)	0.70 (0.63–0.78)	0.65 (0.57–0.70)	0.85 (0.79–0.89)	0.65 (0.57–0.70)	0.67 (0.56–0.74)	0.69 (0.60–0.80)	0.10 (0.07–0.14)	0.44 (0.38–0.52)	0.04 (0.01–0.07)
	Yes (225)	0.70 (0.49–0.82)	0.83 (0.78–0.87)	0.70 (0.62–0.78)	0.65 (0.57–0.69)	0.86 (0.80–0.90)	0.65 (0.57–0.69)	0.67 (0.56–0.73)	0.69 (0.59–0.81)	0.10 (0.07–0.13)	0.44 (0.36–0.50)	0.04 (0.02–0.06)
	p-value*	0.212	0.237	0.858	0.805	0.637	0.805	0.982	0.933	0.166	0.440	0.742
Diabetes (373)	No (264)	0.70 (0.49–0.81)	0.82 (0.76–0.86)	0.69 (0.62–0.77)	0.64 (0.57–0.70)	0.85 (0.79–0.89)	0.64 (0.57–0.70)	0.66 (0.55–0.74)	0.69 (0.61–0.81)	0.10 (0.07–0.13)	0.43 (0.37–0.50)	0.04 (0.01–0.07)
	Yes (109)	0.62 (0.45–0.82)	0.84 (0.79–0.88)	0.72 (0.65–0.79)	0.65 (0.59–0.69)	0.87 (0.81–0.90)	0.65 (0.59–0.69)	0.68 (0.59–0.74)	0.68 (0.57–0.79)	0.10 (0.07–0.14)	0.45 (0.36–0.51)	0.04 (0.02–0.06)
	p-value*	0.253	**0.017**	0.073	0.375	0.055	0.375	0.235	0.202	0.622	0.583	0.200
TC (376)	<5.0 mmol/ L (296)	0.70 (0.49–0.82)	0.83 (0.77–0.87)	0.70 (0.62–0.78)	0.65 (0.57–0.69)	0.86 (0.80–0.89)	0.65 (0.57–0.69)	0.67 (0.55–0.74)	0.68 (0.60–0.78)	0.10 (0.07–0.13)	0.44 (0.37–0.50)	0.04 (0.01–0.06)
	≥5.0 mmol/ L (80)	0.56 (0.45–0.79)	0.83 (0.75–0.86)	0.70 (0.65–0.76)	0.64 (0.55–0.70)	0.85 (0.81–0.90)	0.64 (0.55–0.70)	0.66 (0.57–0.73)	0.77 (0.62–0.90)	0.10 (0.07–0.15)	0.45 (0.38–0.53)	0.05 (0.02–0.09)
	p-value*	0.081	0.398	0.914	0.756	0.934	0.756	0.804	**0.003**	0.579	0.600	**0.035**
TG (374)	<1.70 mmol/ L (247)	0.70 (0.49–0.81)	0.83 (0.76–0.87)	0.71 (0.63–0.78)	0.65 (0.57–0.69)	0.85 (0.80–0.89)	0.65 (0.57–0.69)	0.67 (0.56–0.74)	0.69 (0.59–0.78)	0.10 (0.07–0.14)	0.45 (0.38–0.50)	0.04 (0.02–0.07)
	≥1.70 mmol/ L (127)	0.66 (0.45–0.82)	0.83 (0.76–0.87)	0.70 (0.62–0.77)	0.65 (0.57–0.70)	0.86 (0.80–0.90)	0.65 (0.57–0.70)	0.66 (0.54–0.73)	0.70 (0.60–0.83)	0.10 (0.07–0.14)	0.44 (0.34–0.51)	0.04 (0.01–0.07)
	p-value*	0.374	0.557	0.642	0.940	0.941	0.940	0.416	0.231	0.727	0.281	0.427
HDL (376)	<1.0 mmol/ L (135)	0.63 (0.44–0.81)	0.82 (0.76–0.86)	0.70 (0.62–0.79)	0.65 (0.56–0.70)	0.86 (0.79–0.90)	0.65 (0.56–0.70)	0.67 (0.54–0.74)	0.71 (0.60–0.80)	0.09 (0.07–0.13)	0.44 (0.38–0.52)	0.04 (0.01–0.06)
	≥1.0 mmol/ L (241)	0.70 (0.49–0.81)	0.83 (0.77–0.87)	0.71 (0.63–0.77)	0.64 (0.57–0.69)	0.86 (0.80–0.89)	0.64 (0.57–0.69)	0.67 (0.56–0.74)	0.68 (0.60–0.82)	0.10 (0.07–0.14)	0.45 (0.36–0.50)	0.04 (0.02–0.07)
	p-value*	0.102	0.204	0.860	0.951	0.955	0.951	0.599	0.511	0.185	0.374	**0.023**
LDL (376)	<3.0 mmol/ L (262)	0.70 (0.49–0.82)	0.83 (0.76–0.87)	0.71 (0.61–0.78)	0.65 (0.57–0.70)	0.86 (0.80–0.89)	0.65 (0.57–0.70)	0.67 (0.55–0.74)	0.69 (0.60–0.78)	0.10 (0.07–0.13)	0.44 (0.37–0.50)	0.04 (0.01–0.06)
	≥3.0 mmol/ L (114)	0.67 (0.46–0.80)	0.82 (0.76–0.87)	0.70 (0.65–0.77)	0.64 (0.55–0.70)	0.85 (0.80–0.90)	0.64 (0.55–0.70)	0.65 (0.56–0.73)	0.71 (0.59–0.87)	0.10 (0.07–0.15)	0.47 (0.38–0.53)	0.04 (0.02–0.08)
	p-value*	0.562	0.662	0.847	0.773	0.810	0.773	0.499	0.237	0.429	0.249	0.206

**The p-values were calculated by the Mann-Whitney test, and significant p-values are in bold.*

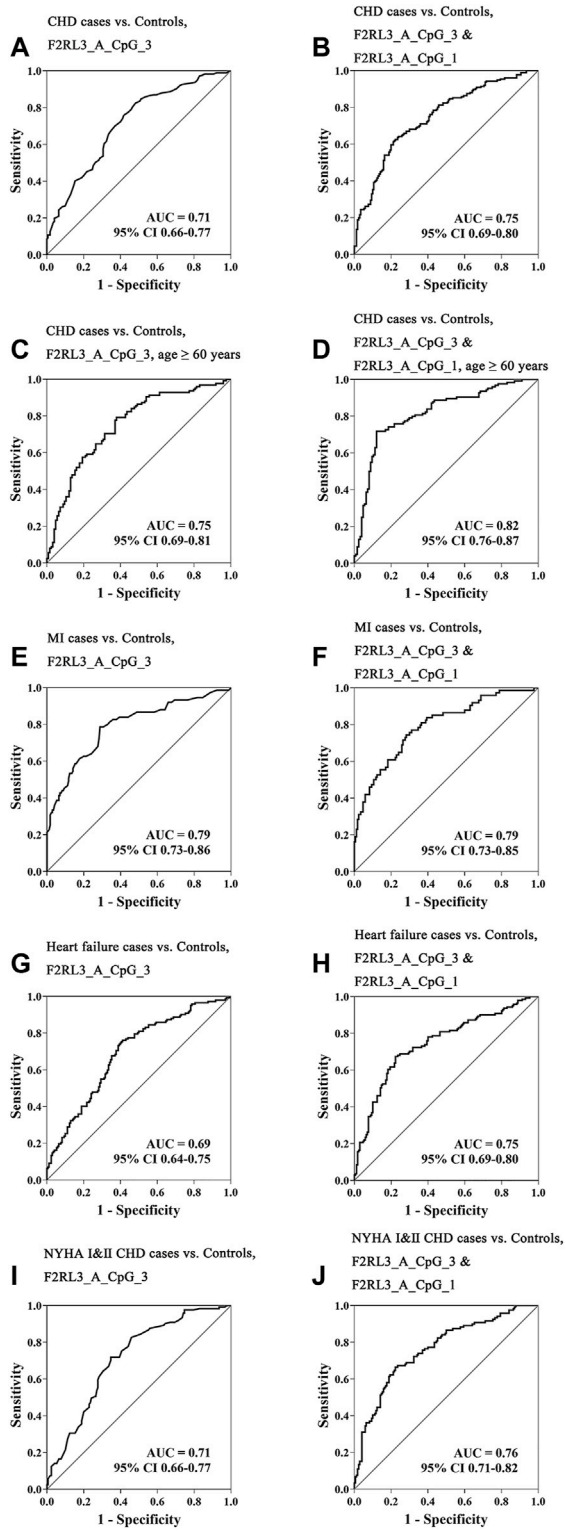


FIGURE 3 | The methylation level of *F2RL3* in the peripheral blood DNA as a marker for the detection of CHD in general, age ≥60 years CHD cases, MI cases, heart failure cases, and patients with minor to medium cardiac function impairment (NYHA I&II CHD cases). **(A,C,E,G,I)** ROC curve analyses for the discriminatory power of F2RL3_A_CpG_3 methylation to distinguish CHD (*Continued*)

FIGURE 3 | cases, age ≥60 years CHD cases, MI cases, heart failure cases, and NYHA I&II CHD cases from controls. **(B,D,F,H,J)** The combination of F2RL3_A_CpG_3 and F2RL3_A_CpG_1 for the discrimination of CHD cases, age ≥60 years CHD cases, MI cases, heart failure cases, and NYHA I&II CHD cases from controls. The ROC analyses were calculated by logistic regression adjusted for age, gender, smoking, hypertension, TC, LDL, and batch effect. The gray lines represent the line of no discrimination.

increased gene expression, which, if this is the case with *F2RL3*, may cause increased inflammation and coagulation (Shenker et al., 2013). Several studies have reported the association between hypomethylation of *F2RL3* in the whole blood and the prognosis of CHD or increased risk of cardiovascular-related mortality, but so far there is no report about the methylation of *F2RL3* in blood and the diagnosis of CHD (Breitling et al., 2012; Zhang et al., 2014b; Zhang et al., 2017; Gao et al., 2018). In our study, we reported the association between the status of CHD and the hypomethylation of *F2RL3* in blood, more specifically at the F2RL3_A amplicon region covering cg036361837, but not at the F2RL3_B amplicon region. In addition, we also firstly observed that *F2RL3* methylation in the peripheral blood was mostly associated with CHD in people older than 60 years (especially at F2RL3_A_CpG_1; **Figure 1C** and **Table 2C**), with MI (especially at F2RL3_A_CpG_3; **Figure 2A** and **Table 3A**), heart failure patients (especially at F2RL3_A_CpG_1; **Figure 2C** and **Table 3C**) and NYHA I&II CHD patients (especially at F2RL3_A_CpG_1; **Figure 2D** and **Table 3D**). This region or CpG sites specific pattern of CHD-related *F2RL3* methylation may extend our understanding of methylation signatures.

Previous studies have revealed major differences in CHD occurring in young and old people owing to patient demographics, cardiopulmonary function, and molecular biological characteristics (Ekblom-Bak et al., 2019). As a result, the incidence of CHD is increased in people old than 65 years (Moran et al., 2008). In our study, we found that the hypomethylation of *F2RL3* is mainly associated with the risk for CHD in people above 60 years old, and this was further enhanced when people became older than 65 years old. These altered DNA methylation patterns in the blood-based *F2RL3* could be detected in minor to medium cardiac function impairment (NYHA I&II CHD cases) and became even more aberrant in the patients with advanced cardiac function impairment (NYHA III&IV CHD cases). Thus, our observation was consistent with the aging-related risk of CHD and suggested that the alternations of DNA methylation in blood, or say in the blood leukocytes, may play a role in the occurrence and even the progress of CHD. Nonetheless, we admitted that age can hardly be fully matched in case-control studies, and future prospective nested case-control study shall provide more robust evidence for the age-related DNA methylation alternation for the risk of CHD.

Very recently, it has been shown that blood leukocyte DNA methylation could predict the risk of future MI and CHD across diverse populations in a large-scale cohort study involving 11,461 individuals (Agha et al., 2019). Here, we specified the significant hypomethylation of *F2RL3* in the blood leukocyte DNA of MI and heart failure CHD cases compared to controls and suggested its potential clinical application as a biomarker for the prediction of MI and heart failure.

The signatures of methylation could be influenced by environmental factors and treatment (Lax and Szyf, 2018; Martin and Fry, 2018). In this study, we confirmed the strong association between the behavior of smoking and hypomethylation of *F2RL3* at cg036361837 and flanking CpG sites as reported previously (Breitling et al., 2011; Wan et al., 2012; Sun et al., 2013; Zhang et al., 2014a). But the association between hypomethylation of *F2RL3* at cg036361837 and flanking CpG sites and CHD is independent from the status of smoking since it only appeared in the non-smokers (**Supplementary Table S4**). We also found that the males had lower *F2RL3* methylation levels than the females whereas the hypomethylation of *F2RL3* is an indicator of high risk for CHD. This may explain why the males have a higher incidence and mortality of CHD than the females (Barrett-Connor, 2013; Benjamin et al., 2018; Ma et al., 2020). However, it seems that the blood-based *F2RL3* methylation was not influenced by drinking, hypertension, diabetes, levels of TC, TG, HDL, and LDL. Most CHD patients have a history of medication. Out of our expectations, we did not observe the obvious influence of 11 common drugs on the methylation level of *F2RL3* (**Supplementary Table S6**). And the correlation with digoxin should also be taken with caution since there were only 11 patients who took this drug (**Supplementary Table S6**). According to Breitling and the following studies (Breitling et al., 2011; Breitling et al., 2012; Zhang et al., 2014a), the hypomethylation of *F2RL3* is reversible but very slow. Even 10 years after quitting smoking, the methylation patterns of *F2RL3* in the formal smokers are still close to the smokers, and significantly differ from the non-smokers (Breitling et al., 2011; Breitling et al., 2012; Zhang et al., 2014a). Thus, *F2RL3* methylation may have a slow response to the exposure. In our study, the duration of treatment to the CHD patients is unknown, but unlikely has lasted for more than 5 years. Therefore, we can hardly observe methylation changes in the *F2RL3* gene. Whether there is drug-induced reversibility of *F2RL3* methylation, a long-term follow-up study is needed. Taken together, our intensive investigation suggested that the methylation of *F2RL3* in blood could hardly be influenced by most of the environmental factors and common medical treatment, and thus, it might be a robust and stable biomarker for the baseline initial diagnosis of CHD. Nevertheless, subjects with environmental and treatment information are still limited. The influence of CHD-related factors and medication on the *F2RL3* methylation warrants further investigations in multi-center studies with larger sample size. Moreover, the role of troponin and brain natriuretic peptide (BNP) as diagnostic biomarkers of cardiovascular disease is well established (Gaggin and Januzzi, 2013; Garg et al., 2017). However, the information of troponin and BNP are not available in the present study and should be collected in future studies. Also, the combination of *F2RL3* methylation and other types of markers, especially the markers representing different pathways or mechanisms, might provide better insight for the detection of CHD.

Differences in methylation profiles might be influenced by the proportions of the leukocyte subpopulations if cell distribution differed by disease status. However, we do not have the information of the blood cell composition in our study. Given that the altered *F2RL3* methylation was purely due to the change of leukocyte proportion in the blood, there should be a similar pattern of altered methylation in the same gene. However, we observed CHD-related methylation changes mostly in the F2RL3_A amplicon, but only slightly in the F2RL3_B amplicon. The data from the Netherlands Twin Register biobank project suggested that the interindividual differences in the cellular composition were independent of the variation observed in DNA methylation or explained only a minor proportion of this variation (Talens et al., 2010). Therefore, the observed association between *F2RL3* methylation and CHD in our study may be partly independent from the variations of blood cell composition. Moreover, there are barely reported CHD-associated mutations or common single nucleotide polymorphisms (SNPs) in the gene of *F2RL3*. Thus, we proposed that the aberrant *F2RL3* methylation in blood might be an independent risk factor for CHD. But we could not conclude if the altered *F2RL3* methylation is a causative factor or a consequence of CHD. Following studies in multi-center studies with enlarged sample size and even prospective studies are needed. Studies for the mechanism of *F2RL3* involving the circulating leukocytes would be meaningful. Unfortunately, due to the limited sample materials, we could not further explore which blood cell component plays a key role in the altered *F2RL3* methylation in blood. It is also meaningful to know if the *F2RL3* methylation contributes to the altered expression of *F2RL3* in blood. However, without fresh blood, RNA extraction is not possible in the present study. The RNA materials and the information of the blood cell proportion should be considered and collected in future studies. When possible, the *F2RL3* methylation and expression in each major cell component of cases and controls should be evaluated and compared. Meanwhile, functional studies of *F2RL3* in the cell lines and animal models would be rather helpful.

In conclusion, our study disclosed the correlation between CHD and blood-based hypomethylation of *F2RL3*, especially at cg03636183 and flanking CpG sites. This correlation appears at the early stage of cardiovascular dysfunction, and is strengthened by older age and the occurrence of MI and heart failure, but is not or just weakly influenced by most of the environmental factors and common medical treatment. We hereby suggested the blood-based *F2RL3* methylation as an objective and stable biomarker for the detection of CHD, especially for people with older age or with the status of MI. The combination of *F2RL3* methylation and conventional risk factors might be an approach to improve the risk evaluation and detection of CHD at early stage.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Chinese PLA General Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KH and RY contributed to conception and design of the study, revision of the manuscript, and revised it critically for important intellectual content. XZ, LY, and QY performed the experiments and wrote the first draft of the manuscript. LZ, QY, and RY performed the statistical analysis. XZ, ZX, QJ, and KH provided the materials and supervised the patient enrollment and acquisition of biological samples and clinical date. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the Nanjing Medical University Research Support Funding (No. 2018RC0003), National Natural Science Foundation of China (No. 82001994), Chinese PLA General Hospital Clinical Research Support Funding (No. 2018FC-WJFWZX-1-21), and Chinese PLA General Hospital Youth Development Project (No. QNC19058).

ACKNOWLEDGMENTS

We thank Chunlan Liu and Shuifang Lei for their kind technique support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.833923/full#supplementary-material

REFERENCES

Agha, G., Mendelson, M. M., Ward-Caviness, C. K., Joehanes, R., Huan, T., Gondalia, R., et al. (2019). Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. *Circulation* 140 (8), 645–657. doi:10.1161/CIRCULATIONAHA.118.039357

Arasaradnam, R. P., Commane, D. M., Bradburn, D., and Mathers, J. C. (2008). A Review of Dietary Factors and its Influence on DNA Methylation in Colorectal Carcinogenesis. *Epigenetics* 3 (4), 193–198. doi:10.4161/epi.3.4.6508

Barrett-Connor, E. (2013). Gender Differences and Disparities in All-Cause and Coronary Heart Disease Mortality: Epidemiological Aspects. *Best Pract. Res. Clin. Endocrinol. Metab.* 27 (4), 481–500. doi:10.1016/j.beem.2013.05.013

Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., et al. (2018). Heart Disease and Stroke Statistics-2018 Update: A Report from the American Heart Association. *Circulation* 137 (12), e67–e492. doi:10.1161/CIR.0000000000000558

Breitling, L. P., Salzmann, K., Rothenbacher, D., Burwinkel, B., and Brenner, H. (2012). Smoking, F2RL3 Methylation, and Prognosis in Stable Coronary Heart Disease. *Eur. Heart J.* 33 (22), 2841–2848. doi:10.1093/eurheartj/ehs091

Breitling, L. P., Yang, R., Korn, B., Burwinkel, B., and Brenner, H. (2011). Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication. *Am. J. Hum. Genet.* 88 (4), 450–457. doi:10.1016/j.ajhg.2011.03.003

Bretschneider, E., Kaufmann, R., Braun, M., Wittpoth, M., Glusa, E., Nowak, G., et al. (1999). Evidence for Proteinase-Activated Receptor-2 (PAR-2)-Mediated Mitogenesis in Coronary Artery Smooth Muscle Cells. *Br. J. Pharmacol.* 126 (8), 1735–1740. doi:10.1038/sj.bjp.0702509

Coughlin, S. R. (2000). Thrombin Signalling and Protease-Activated Receptors. *Nature* 407 (6801), 258–264. doi:10.1038/35025229

Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., Rumley, A., et al. (2004). C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. *N. Engl. J. Med.* 350 (14), 1387–1397. doi:10.1056/NEJMoa032804

Ekblom-Bak, E., Ekblom, B., Söderling, J., Börjesson, M., Blom, V., Kallings, L. V., et al. (2019). Sex- and Age-Specific Associations between Cardiorespiratory Fitness, CVD Morbidity and All-Cause Mortality in 266.109 Adults. *Prev. Med.* 127, 105799. doi:10.1016/j.ypmed.2019.105799

Fu, H., Zhu, K., Zhou, D., Guan, Y., Li, W., and Xu, S. (2019). Identification and Validation of Plasma Metabolomics Reveal Potential Biomarkers for Coronary Heart Disease. *Int. Heart J.* 60 (6), 1387–1397. doi:10.1536/ihj.19-059

Gaggin, H. K., and Januzzi, J. L., Jr (2013). Biomarkers and Diagnostics in Heart Failure. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1832 (12), 2442–2450. doi:10.1016/j.bbadis.2012.12.014

Gao, B. F., Shen, Z. C., Bian, W. S., Wu, S. X., Kang, Z. X., and Gao, Y. (2018). Correlation of Hypertension and F2RL3 Gene Methylation with Prognosis of Coronary Heart Disease. *J. Biol. Regul. Homeost Agents* 32 (6), 1539–1544.

Garg, P., Morris, P., Fazlanie, A. L., Vijayan, S., Dancso, B., Dastidar, A. G., et al. (2017). Cardiac Biomarkers of Acute Coronary Syndrome: from History to High-Sensitivity Cardiac Troponin. *Intern. Emerg. Med.* 12 (2), 147–155. doi:10.1007/s11739-017-1612-1

Gatto, L., and Prati, F. (2020). Subclinical Atherosclerosis: How and When to Treat it? *Eur. Heart J. Suppl.* 22 (Suppl. E), E87–E90. doi:10.1093/eurheartj/suaa068

Hansson, G. K. (2005). Inflammation, Atherosclerosis, and Coronary Artery Disease. *N. Engl. J. Med.* 352 (16), 1685–1695. doi:10.1056/NEJMra043430

Horvath, S., and Raj, K. (2018). DNA Methylation-Based Biomarkers and the Epigenetic Clock Theory of Ageing. *Nat. Rev. Genet.* 19 (6), 371–384. doi:10.1038/s41576-018-0004-3

Jellinger, P. S., Handelsman, Y., Rosenblit, P. D., Bloomgarden, Z. T., Fonseca, V. A., Garber, A. J., et al. (2017). American Association of Clinical Endocrinologists and American College of Endocrinology Guidelines for Management of Dyslipidemia and Prevention of Cardiovascular Disease. *Endocr. Pract.* 23 (Suppl. 2), 1–87. doi:10.4158/EP171764.APPGL

Jia, L., Zhu, L., Wang, J. Z., Wang, X. J., Chen, J. Z., Song, L., et al. (2013). Methylation of FOXP3 in Regulatory T Cells Is Related to the Severity of Coronary Artery Disease. *Atherosclerosis* 228 (2), 346–352. doi:10.1016/j.atherosclerosis.2013.01.027

Kahn, M. L., Nakanishi-Matsui, M., Shapiro, M. J., Ishihara, H., and Coughlin, S. R. (1999). Protease-Activated Receptors 1 and 4 Mediate Activation of Human Platelets by Thrombin. *J. Clin. Invest.* 103 (6), 879–887. doi:10.1172/JCI6042

Kataoka, H., Hamilton, J. R., McKemy, D. D., Camerer, E., Zheng, Y.-W., Cheng, A., et al. (2003). Protease-Activated Receptors 1 and 4 Mediate Thrombin Signaling in Endothelial Cells. *Blood* 102 (9), 3224–3231. doi:10.1182/blood-2003-04-1130

Kolpakov, M. A., Rafiq, K., Guo, X., Hooshdaran, B., Wang, T., Vlasenko, L., et al. (2016). Protease-Activated Receptor 4 Deficiency Offers Cardioprotection after Acute Ischemia Reperfusion Injury. *J. Mol. Cell Cardiol.* 90, 21–29. doi:10.1016/j.yjmcc.2015.11.030

Laird, P. W. (2003). The Power and the Promise of DNA Methylation Markers. *Nat. Rev. Cancer* 3 (4), 253–266. doi:10.1038/nrc1045

Lax, E., and Szyf, M. (2018). The Role of DNA Methylation in Drug Addiction: Implications for Diagnostic and Therapeutics. *Prog. Mol. Biol. Transl Sci.* 157, 93–104. doi:10.1016/bs.pmbts.2018.01.003

Leger, A. J., Covic, L., and Kuliopulos, A. (2006). Protease-Activated Receptors in Cardiovascular Diseases. *Circulation* 114 (10), 1070–1077. doi:10.1161/CIRCULATIONAHA.105.574830

Li, H., Sun, K., Zhao, R., Hu, J., Hao, Z., Wang, F., et al. (2018). Inflammatory Biomarkers of Coronary Heart Disease. *Front. Biosci. (Schol Ed)* 10, 185–196. doi:10.2741/s508

Lobbes, M. B., Kooi, M. E., Lutgens, E., Ruiters, A. W., Lima Passos, V., Braat, S. H., et al. (2010). Leukocyte Counts, Myeloperoxidase, and Pregnancy-Associated

Plasma Protein a as Biomarkers for Cardiovascular Disease: Towards a Multi-Biomarker Approach. *Int. J. Vasc. Med.* 2010, 726207. doi:10.1155/2010/726207

Ma, L. Y., Chen, W. W., Gao, R. L., Liu, L. S., Zhu, M. L., Wang, Y. J., et al. (2020). China Cardiovascular Diseases Report 2018: An Updated Summary. *J. Geriatr. Cardiol.* 17 (1), 1–8. doi:10.11909/j.issn.1671-5411.2020.01.001

Martin, E. M., and Fry, R. C. (2018). Environmental Influences on the Epigenome: Exposure- Associated DNA Methylation in Human Populations. *Annu. Rev. Public Health* 39, 309–333. doi:10.1146/annurev-publhealth-040617-014629

Moran, A., Zhao, D., Gu, D., Coxson, P., Chen, C.-S., Cheng, J., et al. (2008). The Future Impact of Population Growth and Aging on Coronary Heart Disease in China: Projections from the Coronary Heart Disease Policy Model-China. *BMC Public Health* 8, 394. doi:10.1186/1471-2458-8-394

Peng, P., Wang, L., Yang, X., Huang, X., Ba, Y., Chen, X., et al. (2014). A Preliminary Study of the Relationship between Promoter Methylation of the ABCG1, GALNT2 and HMGCR Genes and Coronary Heart Disease. *PLoS One* 9 (8), e102265. doi:10.1371/journal.pone.0102265

Robertson, K. D., and Wolffe, A. P. (2000). DNA Methylation in Health and Disease. *Nat. Rev. Genet.* 1 (1), 11–19. doi:10.1038/35049533

Schleithoff, C., Voelter-Mahlknecht, S., Dahmke, I. N., and Mahlknecht, U. (2012). On the Epigenetics of Vascular Regulation and Disease. *Clin. Epigenet* 4 (1), 7. doi:10.1186/1868-7083-4-7

Shaya, G. E., Leucker, T. M., Jones, S. R., Martin, S. S., and Toth, P. P. (2021). Coronary Heart Disease Risk: Low-Density Lipoprotein and Beyond. *Trends Cardiovasc. Med.* S1050-1738 (21), 00046–53. doi:10.1016/j.tcm.2021.04.002

Shenker, N. S., Polidoro, S., van Veldhoven, K., Sacerdote, C., Ricceri, F., Birrell, M. A., et al. (2013). Epigenome-Wide Association Study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) Identifies Novel Genetic Loci Associated with Smoking. *Hum. Mol. Genet.* 22 (5), 843–851. doi:10.1093/hmg/dds488

Steinhoff, M., Buddenkotte, J., Shpacovitch, V., Rattenholl, A., Moormann, C., Vergnolle, N., et al. (2005). Proteinase-Activated Receptors: Transducers of Proteinase-Mediated Signaling in Inflammation and Immune Response. *Endocr. Rev.* 26 (1), 1–43. doi:10.1210/er.2003-0025

Sun, Y. V., Smith, A. K., Conneely, K. N., Chang, Q., Li, W., Lazarus, A., et al. (2013). Epigenomic Association Analysis Identifies Smoking-Related DNA Methylation Sites in African Americans. *Hum. Genet.* 132 (9), 1027–1037. doi:10.1007/s00439-013-1311-6

Talens, R. P., Boomsma, D. I., Tobi, E. W., Kremer, D., Jukema, J. W., Willemsen, G., et al. (2010). Variation, Patterns, and Temporal Stability of DNA Methylation: Considerations for Epigenetic Epidemiology. *FASEB j.* 24 (9), 3135–3144. doi:10.1096/fj.09-150490

Talmud, P. J. (2007). Gene-Environment Interaction and its Impact on Coronary Heart Disease Risk. *Nutr. Metab. Cardiovasc. Dis.* 17 (2), 148–152. doi:10.1016/j.numecd.2006.01.008

Vergnolle, N., Derian, C. K., D'Andrea, M. R., Steinhoff, M., and Andrade-Gordon, P. (2002). Characterization of Thrombin-Induced Leukocyte Rolling and Adherence: A Potential Proinflammatory Role for Proteinase-Activated Receptor-4. *J. Immunol.* 169 (3), 1467–1473. doi:10.4049/jimmunol.169.3.1467

Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., et al. (2021). Heart Disease and Stroke Statistics-2021 Update: A Report from the American Heart Association. *Circulation* 143 (8), e254–e743. doi:10.1161/CIR.0000000000000950

Vorchheimer, D. A., and Becker, R. (2006). Platelets in Atherothrombosis. *Mayo Clinic Proc.* 81 (1), 59–68. doi:10.4065/81.1.59

Wan, E. S., Qiu, W., Baccarelli, A., Carey, V. J., Bacherman, H., Rennard, S. I., et al. (2012). Cigarette Smoking Behaviors and Time since Quitting Are Associated with Differential DNA Methylation across the Human Genome. *Hum. Mol. Genet.* 21 (13), 3073–3082. doi:10.1093/hmg/dds135

Wang, J., Tan, G. J., Han, L. N., Bai, Y. Y., He, M., and Liu, H. B. (2017). Novel Biomarkers for Cardiovascular Risk Prediction. *J. Geriatr. Cardiol.* 14 (2), 135–150. doi:10.11909/j.issn.1671-5411.2017.02.008

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., et al. (2005). Chromosome-Wide and Promoter-Specific Analyses Identify Sites of Differential DNA Methylation in normal and Transformed Human Cells. *Nat. Genet.* 37 (8), 853–862. doi:10.1038/ng1598

Wei, M., Liu, Y., Zheng, M., Wang, L., Ma, F., Qi, Y., et al. (2019). Upregulation of Protease-Activated Receptor 2 Promotes Proliferation and Migration of Human Vascular Smooth Muscle Cells (VSMCs). *Med. Sci. Monit.* 25, 8854–8862. doi:10.12659/MSM.917865

Yancy, C. W., Jessup, M., Bozkurt, B., Butler, J., Casey, D. E., Jr., Colvin, M. M., et al. (2017). 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *J. Card. Fail.* 23 (8), 628–651. doi:10.1016/j.cardfail.2017.04.014

Yang, R., Pfütze, K., Zucknick, M., Sutter, C., Wappenschmidt, B., Marme, F., et al. (2015). DNA Methylation Array Analyses Identified Breast Cancer-AssociatedHYAL2methylation in Peripheral Blood. *Int. J. Cancer* 136 (8), 1845–1855. doi:10.1002/ijc.29205

Zhang, L., Zhang, Y., Zhao, Y., Wang, Y., Ding, H., Xue, S., et al. (2018). Circulating miRNAs as Biomarkers for Early Diagnosis of Coronary Artery Disease. *Expert Opin. Ther. Patents* 28 (8), 591–601. doi:10.1080/13543776.2018.1503650

Zhang, Y., Wilson, R., Heiss, J., Breitling, L. P., Saum, K.-U., Schöttker, B., et al. (2017). DNA Methylation Signatures in Peripheral Blood Strongly Predict All-Cause Mortality. *Nat. Commun.* 8, 14617. doi:10.1038/ncomms14617

Zhang, Y., Yang, R., Burwinkel, B., Breitling, L. P., and Brenner, H. (2014a). F2RL3 Methylation as a Biomarker of Current and Lifetime Smoking Exposures. *Environ. Health Perspect.* 122 (2), 131–137. doi:10.1289/ehp.1306937

Zhang, Y., Yang, R., Burwinkel, B., Breitling, L. P., Holleczek, B., Schöttker, B., et al. (2014b). F2RL3methylation in Blood DNA Is a strong Predictor of Mortality. *Int. J. Epidemiol.* 43 (4), 1215–1225. doi:10.1093/ije/dyu006

Zuo, H. P., Guo, Y. Y., Che, L., and Wu, X. Z. (2016). Hypomethylation of Interleukin-6 Promoter Is Associated with the Risk of Coronary Heart Disease. *Arq Bras Cardiol.* 107 (2), 131–136. doi:10.5935/abc.20160124

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fgene.2022.844141

Association of DNA Methylation Patterns in 7 Novel Genes With Ischemic Stroke in the Northern Chinese Population

Hongwei Sun[1], Jia Xu[2], Bifeng Hu[1], Yue Liu[1], Yun Zhai[1], Yanyan Sun[1], Hongwei Sun[1], Fang Li[1], Jiamin Wang[2], Anqi Feng[2], Ying Tang[1] and Jingbo Zhao[2]**

[1]Department of Neurology, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Epidemiology, School of Public Health, Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:
Cunyou Zhao,
Southern Medical University, China

Reviewed by:
Yi Huang,
Zhejiang University School of Medicine, China
Yingying Xu,
The Second Hospital of Shandong University, China

***Correspondence:**
Jingbo Zhao
bojingzhao@126.com
Ying Tang
hydtangying@hotmail.com

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: 27 December 2021
Accepted: 18 March 2022
Published: 11 April 2022

Citation:
Sun H, Xu J, Hu B, Liu Y, Zhai Y, Sun Y, Sun H, Li F, Wang J, Feng A, Tang Y and Zhao J (2022) Association of DNA Methylation Patterns in 7 Novel Genes With Ischemic Stroke in the Northern Chinese Population. Front. Genet. 13:844141. doi: 10.3389/fgene.2022.844141

Background: Ischemic stroke is a highly complex disorder. This study aims to identify novel methylation changes in ischemic stroke.

Methods: We carried out an epigenome-wide study of ischemic stroke using an Infinium HumanMethylation 850K array (cases:controls = 4:4). 10 CpG sites in 8 candidate genes from gene ontology analytics top-ranked pathway were selected to validate 850K BeadChip results (cases:controls = 20:20). We further qualified the methylation level of promoter regions in 8 candidate genes (cases:controls = 188:188). Besides, we performed subgroup analysis, dose-response relationship and diagnostic prediction polygenic model of candidate genes.

Results: In the discovery stage, we found 462 functional DNA methylation positions to be associated with ischemic stroke. Gene ontology analysis highlighted the "calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules" item, including 8 candidate genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB6/PCDHB9*). In the replication stage, we identified 5 differentially methylated loci in 20 paired samples and 7 differentially methylated genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) in 188 paired samples. Subgroup analysis showed that the methylation level of above 7 genes remained significantly different in the male subgroup, large-artery atherosclerosis subgroup and right hemisphere subgroup. The methylation level of each gene was grouped into quartiles, and Q4 groups of the 7 genes were associated with higher risk of ischemic stroke than Q1 groups ($p < 0.05$). Besides, the polygenic model showed high diagnostic specificity (0.8723), sensitivity (0.883), and accuracy (0.8777).

Abbreviations: CDHs, cadherins; CE, cardioembolism; CpG, cytosine phosphate guanine; DMPs, DNA methylation positions; GO, gene ontology; HDL, high-density lipoprotein; IS, ischemic stroke; KEGG, Kyoto Encyclopedia of Genes and Genomes; LAA, large-artery atherosclerosis; LDL, low-density lipoprotein; ODC, other determined cause; PCDHs, protocadherins; ROC, receiver operating characteristic; SVD, small vessel disease; TC, total cholesterol; TG, total triglyceride; TOAST, Trial of ORG 10172 in Acute Stroke Treatment; UND, undetermined; 5′UTR, 5′untranslated region; AUC, area under the curve.

Conclusion: Our results demonstrate that DNA methylation plays a crucial part in ischemic stroke. The methylation of these 7 genes may be potential diagnostic biomarker for ischemic stroke.

Keywords: ischemic stroke, epigenome-wide association study, DNA methylation, gene ontology, cadherins

INTRODUCTION

Stroke is a devastating disease due to its high morbidity, disability, and recurrence. The prevalence of stroke is increasing over time, and the affected population is becoming younger (Katan and Luft, 2018). Ischemic stroke (IS) is the dominant subtype of stroke with a proportion more than 80% (Ajoolabady et al., 2021). Many of the environmental and genetic risk factors are associated with IS, and the genetic risk is believed to be in the order of 37.9%. However, genetic variants associated with IS found to date only account for 5–10% of that genetic risk (Bevan et al., 2012; Krupinski et al., 2017), which suggests that more associated heritable risk factors have not yet been discovered. Epigenetic modifications is one of these possible heritable changes (Mahjoubin-Tehran et al., 2021).

Epigenetics refer to chemical modifications of DNA structure without affecting the DNA sequence that may provide a link between environment and gene expression (Majnik and Lane, 2014). Generally, epigenetic modifications of gene expression occur by three main forms: DNA methylation, histone modification and microRNA expression (Schiano et al., 2020; Liu et al., 2020). As a major type of epigenetic process, DNA methylation mainly occurs at the cytosine of a cytosine-phosphate-guanine (CpG) dinucleotide forming 5-methylcytosine, which can inhibit gene expression through transcriptional silencing (Bird, 2007). DNA methylation is a crucial epigenetic mechanism involved in normal and pathological cellular processes (Krupinski et al., 2017).

The current epigenome-wide studies revealed that DNA methylation plays a vital part in the pathogenesis and recurrence of IS (Soriano-Tárraga et al., 2020; Davis Armstrong et al., 2018; Gómez-Úriz et al., 2015; Shen et al., 2019), which have mainly been conducted in American and European populations (Soriano-Tárraga et al., 2020; Davis Armstrong et al., 2018; Gómez-Úriz et al., 2015). An epigenetic study in a Chinese population revealed that hypomethylation of the *MTRNR2L8* gene is associated with IS (Shen et al., 2019). It is undeniable that this study is significant; however, it focused on the large-artery atherosclerosis stroke, a subgroup of IS, which cannot fully demonstrate the role of DNA methylation in IS. Thus far, epigenome-wide alterations of IS have not been systematically investigated in a Chinese population.

In our research, we designed a two-stage case-control study to perform integrated analysis of genome-wide DNA methylation profiles to identify novel candidate genes and pathways for IS in a Chinese population.

MATERIALS AND METHODS

Ethical Approval

We performed this study according to the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. The study procedure was approved by the Ethics Committee of the First Affiliated Hospital of Harbin Medical University. All participants voluntarily gave written informed consent.

Study Population

In this study, a two-stage case-control DNA methylation study was designed, including discovery and replication analyses. The sample consisted of 192 IS patients and 192 age-(±3 years) and sex-matched controls from the First Affiliated Hospital of Harbin Medical University. **Figure 1** demonstrates the study design and selection criteria for cases and controls. The control group was from healthy people who underwent physical examination or patients hospitalized from the First Affiliated Hospital of Harbin Medical University at the same period.

Data and Sample Collection

Data for subjects included questionnaire, laboratory investigations, and clinical characteristics. We collected data using a questionnaire related to demographic information (such as gender, age, and marital status), lifestyle habits (such as tobacco smoking and alcohol drinking), and past medical history (hypertension and type 2 diabetes mellitus) of the participants (Qin et al., 2019).

Laboratory investigations were done at the First Affiliated Hospital of Harbin Medical University. We collected blood samples in tubes with clot activator and gel separator from all participants in the morning after fasting and measured using an auto biochemical analyzer (Biobase BK-600; Shandong, China). The levels of total triglyceride (TG), total cholesterol (TC), high-density lipoprotein (HDL), and low-density lipoprotein (LDL) was collected from all participants.

Clinical characteristics included Trial of Org 10172 in Acute Stroke Treatment (TOAST) type and stroke location. Stratification analysis by TOAST criteria was performed for large-artery atherosclerosis (LAA), cardioembolism (CE), small vessel disease (SVD), other determined cause (ODC), and undetermined (UND) aetiologies. Stroke location was categorized as anterior, posterior, and both anterior and posterior circulation. Anterior circulation was further subdivided as left and right hemisphere based on computerized tomography or magnetic resonance imaging scans.

Two milliliters of blood samples were collected in EDTA tubes from all participants and stored at −80°C until DNA extraction.

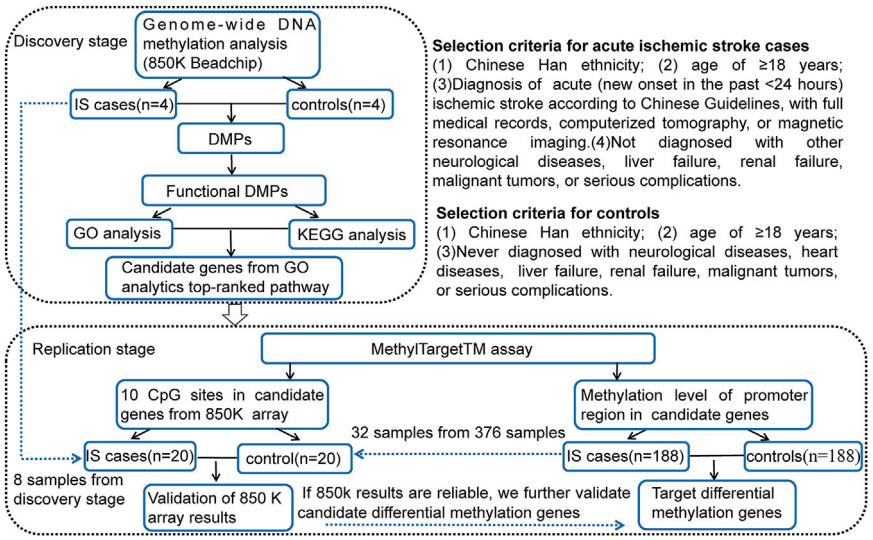


FIGURE 1 | Two-stage case-control study design and selection criteria for cases and controls. IS, ischemic stroke; DMPs, DNA methylation positions; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; CpG, cytosine phosphate guanine.

Discovery Stage: Genome-Wide DNA Methylation Profiles

Genomic DNA was extracted from the leukocytes in the peripheral blood using an DNA Methylation kit (Zymo, Irvine, CA), following the protocol (Genesky Biotechnologies Inc., Shanghai, China). Genome-wide DNA methylation was assessed using Infinium Human Methylation 850K BeadChip (Illumina Inc., San Diego, CA, United States)). More than 853,000 CpG sites were included in each chip. CpG sites containing documented single-nucleotide polymorphisms and mapping to X and Y chromosomes were removed to avoid potential confounding by single-nucleotide polymorphisms and gender.

Discovery Stage: Bioinformatics Analysis of DNA Methylation Profiles

To screen functional DNA methylation positions (DMPs), DMPs located in promoter regions (upstream or 5′untranslated region, 5′UTR), with |methylation difference|>0.1 and $p < 0.05$ were selected. We performed Gene ontology (GO, http://www.geneontology.org) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/) pathway enrichment analysis to clarify the function and biological pathways of differential methylation loci. Through GO enrichment analysis, the differentially methylated genes were classified according to cellular component, molecular function and biological process. In GO and KEGG enrichment analysis, $p < 0.05$ was significantly enriched by differential methylation loci-related genes.

Replication Stage: MethylTargetTM Assay

Considering the *p*-values and biological functions of the genes, we selected 8 candidate genes from GO analytics top-ranked pathway including 10 CpG sites to validate 850 K BeadChips

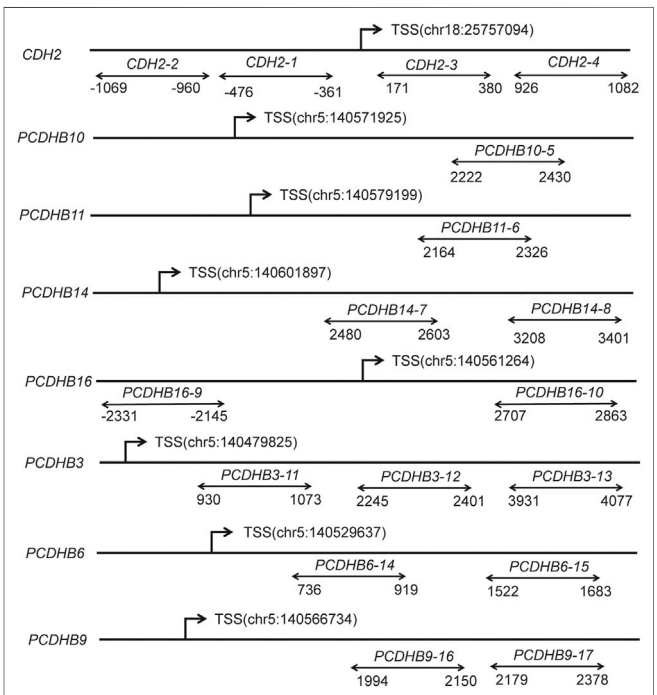


FIGURE 2 | Diagram of the structure of the 8 genes and 17 CpG islands. The line represents the gene, and double arrows represent CpG islands. CpG, cytosine phosphate guanine; TSS, transcription start sites; Chr, chromosome.

results using MethylTargetTM assay, which was a multi-targeted CpG methylation analysis method based on next-generation sequencing (Wan et al., 2021). As the 10 CpG sites in 850 K BeadChips results are not top-ranked loci and FDR *p* values of the 10 CpG sites are not statistically significant, we replicated 10 CpGs in a small replication sample (20 cases:20 matched controls,

8 samples from DNA methylation chip analysis and 32 samples from validation analysis). Moreover, we further qualified the methylation level of promoter regions in 8 candidate genes, containing 17 CpG islands and 308 CpG sites, using a large replication sample (188 cases:188 individually matched controls). The locations of the 17 CpG islands in 8 genes are shown in **Figure 2**. The primers were carefully designed (**Additional File S1**).

Validation of the mRNA Expression of Differentially Methylated Genes

The dataset GSE22255 was acquired from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) and was utilized as the validation sample. Gene expression profiling was performed in peripheral blood of 20 IS patients and 20 sex- and age-matched controls. From this dataset, the downloaded data format was MINIML. Box plots of expression were drawn by the R software package ggplot2, and principal component analysis graphs were drawn by the R software package ggord. The Wilcoxon rank-sum test was used to detect whether the mRNA expression of differentially methylated genes was different between stroke patients and controls.

Statistical Analysis

Quantitative variables were reported as the mean and standard deviation, while qualitative variables were expressed as frequencies and percentages (%). The normality of quantitative variables was assessed using the Shapiro–Wilk test.

First, we compared the differences of baseline demographic characteristics, lifestyle habits, medical history, and lipid levels between IS cases and controls. The differences of the categorical variables were measured with a chi-square test. We performed a paired t test for the normal distributed continuous variables. Non-parametric test was used for the skewed distributed continuous variables.

Second, the methylation status was compared between IS cases and marched controls using conditional logistic regression analysis. Smoking, drinking, medical history of diabetes mellitus and hypertension, and lipid levels were added as adjusted factors. The association of the methylation level of genes was assessed using Pearson's correlation. All subjects were classified into four groups (Q1, Q2, Q3, and Q4) based on the quartile value of methylation level. Q1 group having the lowest methylation level is regarded as the reference group, and all other groups (Q2 having 25–50% of the values, Q3 having 50–75%, and Q4 having 75–100%) were compared with Q1 group (Qin et al., 2019).

The methylation level of each CpG site was measured as the percentage of the methylated cytosines over total tested cytosines. The average methylation level of all CpG sites was worked out as the methylation level of the DNA segments and genes. Results were considered statistically significant when the p values were less than 0.05. All statistical analyses were performed with the SAS version 9.4 for Windows (SAS Institute, Inc., Cary, NC).

RESULTS

Characteristics of Participants

The distribution of characteristics for the participants are presented in **Table 1**. Compared with controls, IS patients in validation stage had a higher prevalence of smoking, drinking and hypertension ($p < 0.0001$), and higher level of triglyceride ($p = 0.0288$).

Discovery Stage: Epigenome-Wide Association Study of IS

In the discovery analysis, 2656 DMPs exhibited differences ($p < 0.05$, |methylation difference|>0.1) in DNA methylation between the two groups. The results are shown in another unpublished article (Xu J, et al.). After screening, a total of 462 functional DMPs (located in the promoter region, | methylation difference| > 0.1 and $p < 0.05$) corresponding to 373 genes were selected (**Additional File S2**). Among them, the majority of the DMPs were in CpG island (26.4%) and N-shore (28.6%). In addition, 87.7% (405 of 462) CpG cites corresponding to 327 genes were found to be hypomethylated in IS cases compared to controls.

Discovery Stage: Bioinformatics Analysis of DNA Methylation Profiles

We screened and finally included 12 KEGG pathways ($p < 0.05$, **Figure 3**). A total of 315 significant GO terms ($p < 0.05$) were selected. The top 10 GO items were shown in **Table 2**. Among them, the "calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules" item showed the most significant fold enrichment. There are 8 genes in this item including *CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB6/PCDHB9*, and the 8 genes were selected for replication.

Replication Stage

In the replication stage, 5 of 10 CpG sites in 8 candidate genes from 850K BeadChip results were verified to be statistically different ($p < 0.05$). The trend of the methylation level of the rest 5 CpG sites with no significant differences is consistent with the 850k BeadChip results (**Additional File S3**), suggesting that the 850k results are reliable. The genes screened by GO analysis are worthy of further verification.

In 188 paired samples, 203 of 308 CpG sites from 17 CpG islands in the promoter regions of 8 candidate genes were significantly different ($p < 0.05$) between the IS and control group (**Additional File S4**). Consistent with the 850k BeadChip results, the methylation level of all 203 CpG sites was lower in the IS group than in the control group. 13 of 17 CpG islands were differentially methylated between IS patients and controls ($p < 0.05$, **Additional File S5**). 7 genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) showed differences ($p < 0.05$) between the two groups even if adjusting for smoking, drinking, history of hypertension and diabetes, and blood lipid

TABLE 1 | Characteristics of participants.

	DNA methylation chip analysis		*P*	DNA methylation validation analysis		*P*
	IS (*n* = 4)	Control (*n* = 4)		IS (*n* = 188)	Control (*n* = 188)	
Age	65.2397 ± 2.6561	65.9904 ± 2.9258	0.7171	62.2155 ± 10.6660	62.1466 ± 10.5739	0.9499
Female	2 (50%)	2 (50%)	1.000	44 (23.40%)	44 (23.40%)	1.000
Smoking (%)	1 (25%)	1 (25%)	1.000	106 (56.38%)	41 (21.81%)	<0.0001*
Drinking (%)	1 (25%)	1 (25%)	1.000	91 (48.40%)	39 (20.74%)	<0.0001*
Hypertension (%)	0 (0%)	0 (0%)	1.000	112 (59.57%)	69 (36.70%)	<0.0001*
Diabetes (%)	0 (0%)	0 (0%)	1.000	35 (18.62)	26 (13.83%)	0.2080
Total cholesterol (mmol/L)	5.4900 ± 0.8676	5.1325 ± 0.2749	0.4620	5.4598 ± 4.5605	4.8362 ± 1.0503	0.0691
Triglyceride (mmol/L)	1.7350 ± 0.8916	1.4225 ± 1.0174	0.6604	1.9978 ± 1.3728	1.7139 ± 1.1217	0.0288*
High-density lipoprotein (mmol/L)	1.1650 ± 0.2883	1.5425 ± 0.5916	0.2950	1.2963 ± 0.9580	1.2990 ± 0.4034	0.9715
Low-density lipoprotein (mmol/L)	3.5500 ± 0.4497	2.9450 ± 0.3692	0.0828	3.2304 ± 0.7974	2.9960 ± 1.8316	0.1089

**Statistically significant difference ($p < 0.05$). IS, ischemic stroke.*

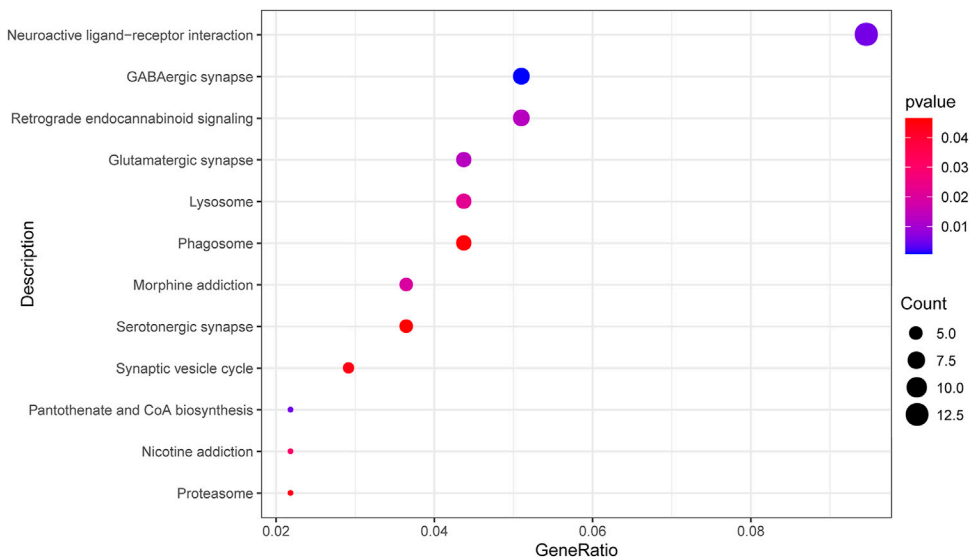


FIGURE 3 | Top 12 enriched KEGG pathways of differential methylation loci-related genes. The dot size represents the number of genes in the pathway, and the dot color represents the enrichment *p* value. KEGG, Kyoto Encyclopedia of Genes and Genomes; GABA, gamma-aminobutyric acid.

levels, and adjust FDR *p* values of 6 genes (*CDH2/PCDHB10/PCDHB11/PCDHB16/PCDHB3/PCDHB9*) were statistically significant ($p < 0.05$, **Table 3**). The methylation levels of the 7 genes were highly correlated in both the case group and the control group ($p < 0.05$, **Additional File S6**). Besides, we further split the samples by sex (male or female). There were 7 differentially methylated genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) in the male subgroup and 4 (*CDH2/PCDHB10/PCDHB11/PCDHB9*) in the female subgroup, which is shown in **Table 4**.

Stratified Analysis by IS Subtypes and Location

We split the case samples by IS subtypes (TOAST classification), stratified by LAA ($n = 67$), CE ($n = 13$), SVD ($n = 64$), ODC ($n = 5$), and UND ($n = 39$). However, no analysis was conducted in the CE and ODC groups because of the small sample size. We also did not perform analysis in the UND group as it is hard to identify the cause of IS. There were 8 differentially methylated genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB6/PCDHB9*) in the LAA subgroup and 1 (*CDH2*) in the SVD subgroup after adjusting for smoking, drinking, history of hypertension and diabetes, and blood lipid levels ($p < 0.05$, **Table 5**). Besides, the case samples were seperated according to the location of the lesion (anterior circulation-left hemisphere, $n = 72$; anterior circulation-right hemisphere, $n = 63$; posterior circulation, $n = 48$; and both anterior and posterior circulation, $n = 5$). We also did not perform analysis in the "both anterior and posterior circulation" group because of the small sample size. Four genes (*CDH2/PCDHB10/PCDHB11/PCDHB9*) in the anterior circulation-left hemisphere subgroup, 8 genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB6/PCDHB9*) in

TABLE 2 | Top 10 enriched GO pathways of differential methylation loci-related genes.

ID	Description	Genes	*P*
BP			
GO:0016339	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	*PCDHB14/PCDHB11/PCDHB3/PCDHB16/PCDHB9/CDH2/PCDHB10/PCDHB6*	$p < 0.0001$
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	*DSC3/PCDHGA4/PCDHA6/PCDHB14/PCDHGA11/PCDHB11/PCDHB3/PCDHGA1/PCDHA1/PCDHB16/PCDHB9/CDH2/PCDHB7/PCDHB10/PCDHB6*	$p < 0.0001$
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	*DSC3/PCDHGA4/PCDHA6/CBLN1/PCDHB14/PCDHGA11/PCDHB11/PCDHB3/PCDHGA1/PCDHA1/PCDHB16/PCDHB9/CDH2/PCDHB7/PCDHB10/CD164/PCDHB6*	$p < 0.0001$
GO:0007416	synapse assembly	*CBLN1/NRG1/PCDHB14/THBS2/PCDHB11/PCDHB3/PCDHB16/PCDHB9/CDH2/NRXN2/PCDHB7/PCDHB10/PCDHB6*	$p < 0.0001$
GO:0050808	synapse organization	*ANK3/CBLN1/NRG1/PCDHB14/THBS2/CTNND2/CACNB2/PCDHB11/PCDHB3/PCDHB16/PCDHB9/CDH2/NRXN2/PCDHB7/PCDHB10/PCDHB6*	$p < 0.0001$
GO:0048863	stem cell differentiation	*PSMA8/NRG1/KIT/EDN3/TP73/MYOCD/TAL1/PSMD1/PSMB2/SFRP1/SEMA3C/MSX1/YTHDF2*	$p = 0.0008$
CC			
GO:0045211	postsynaptic membrane	*PDLIM4/ANK3/GRIK4/GRIK1/CBLN1/TANC1/CHRM2/ABI1/GABRA5/SSPN/GABRG1/GRM7/NETO1/GABRB3*	$p < 0.0001$
GO:0098794	postsynapse	*ADD2/PDLIM4/ANK3/GRIK4/GRIK1/CBLN1/CALB1/TANC1/CHRM2/ABI1/GABRA5/SSPN/GABRG1/GRM7/NETO1/CDH2/KCNN2/GABRB3*	$p = 0.0004$
GO:0097060	synaptic membrane	*PDLIM4/ANK3/GRIK4/GRIK1/CBLN1/TANC1/CHRM2/ABI1/GABRA5/SSPN/GABRG1/GRM7/NETO1/GABRB3*	$p = 0.0006$
MF			
GO:0005338	nucleotide-sugar transmembrane transporter activity	*SLC35D2/SLC35C1/SLC35A1*	$p = 0.0005$

GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function.

the anterior circulation-right hemisphere subgroup and 4 genes (*CDH2/PCDHB10/PCDHB11/PCDHB9*) in the posterior circulation subgroup showed differences after adjusting for smoking, drinking, history of hypertension and diabetes, and blood lipid levels ($p < 0.05$, **Table 6**).

The Association Between DNA Methylation in 7 Differentially Methylated Genes and IS

The methylation level of each gene was grouped into quartiles. The DNA methylation level of the 7 differentially methylated genes in Q4 groups was associated with higher risk of IS than Q1 groups ($p < 0.05$), which are listed in **Table 7**.

Biomarker Potential of the 7 Differentially Methylated Genes for IS

We tested the biomarker potential of the 7 differentially methylated genes using receiver operating characteristic (ROC) analysis. The sensitivity, specificity, accuracy and area under the curve (AUC) in *CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9* genes was determined based on methylation values of replication stage (**Table 8**, **Figure 4A**). To enhance the diagnostic ability of these differentially methylated genes in IS, we tested the performance of the polygenic methylation model. The sensitivity, specificity, accuracy and AUC of polygenic methylation model was shown in **Table 8** and **Figure 4B** using ROC analysis. The polygenic methylation model performed better than each of the individual genes,

TABLE 3 | Validation of 8 candidate genes in the ischemic stroke group and control group.

Target	Group difference	*p* value	Adjust *p* value	Adjust FDR *p* value
CDH2	−0.0127	$p < 0.0001$*	$p < 0.0001$*	$p = 0.0001$*
PCDHB10	−0.0228	$p < 0.0001$*	$p < 0.0001$*	$p = 0.0002$*
PCDHB11	−0.0232	$p < 0.0001$*	$p = 0.0001$*	$p = 0.0006$*
PCDHB14	−0.0193	$p = 0.0002$*	$p = 0.0498$*	$p = 0.0940$
PCDHB16	−0.0162	$p = 0.0117$*	$p = 0.0148$*	$p = 0.0315$*
PCDHB3	−0.0158	$p = 0.0016$*	$p = 0.0089$*	$p = 0.0217$*
PCDHB6	−0.0087	$p = 0.3376$	$p = 0.3399$	$p = 0.5268$
PCDHB9	−0.0251	$p < 0.0001$*	$p = 0.0002$*	$p = 0.0008$*

*Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). *Statistically significant difference ($p < 0.05$).*

TABLE 4 | Stratified analysis by gender.

Target	Group difference	*p* value	Adjust *p* value
Male (case:control = 144:144)			
CDH2	−0.0118	$p < 0.0001$*	$p = 0.0006$*
PCDHB10	−0.0211	$p = 0.0001$*	$p < 0.0001$*
PCDHB11	−0.0216	$p = 0.0005$*	$p = 0.0003$*
PCDHB14	−0.0201	$p = 0.0002$*	$p = 0.0139$*
PCDHB16	−0.0153	$p = 0.0327$*	$p = 0.0070$*
PCDHB3	−0.0139	$p = 0.0081$*	$p = 0.0038$*
PCDHB6	−0.0038	$p = 0.7111$	$p = 0.3113$
PCDHB9	−0.0230	$p = 0.0003$*	$p = 0.0003$*
Female (case:control = 44:44)			
CDH2	−0.0158	$p = 0.0014$*	$p = 0.0299$*
PCDHB10	−0.0283	$p = 0.0103$*	$p = 0.0270$*
PCDHB11	−0.0283	$p = 0.0129$*	$p = 0.0160$*
PCDHB14	−0.0164	$p = 0.2108$	$p = 0.0562$
PCDHB16	−0.0192	$p = 0.1810$	$p = 0.2095$
PCDHB3	−0.0220	$p = 0.0875$	$p = 0.0907$
PCDHB6	−0.0247	$p = 0.2093$	$p = 0.3205$
PCDHB9	−0.0317	$p = 0.0124$*	$p = 0.0195$*

*Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). *Statistically significant difference ($p < 0.05$).*

TABLE 5 | Stratified analysis by ischemic stroke subtypes.

Target	Group difference	*p* value	Adjust *p* value
LAA (case:control = 67:67)			
CDH2	−0.0138	$p = 0.0016$*	$p = 0.0059$*
PCDHB10	−0.0249	$p = 0.0015$*	$p = 0.0028$*
PCDHB11	−0.0264	$p = 0.0035$*	$p = 0.0034$*
PCDHB14	−0.0273	$p = 0.0016$*	$p = 0.0070$*
PCDHB16	−0.0204	$p = 0.0597$	$p = 0.0043$*
PCDHB3	−0.0182	$p = 0.0247$*	$p = 0.0035$*
PCDHB6	−0.0146	$p = 0.3849$	$p = 0.0434$*
PCDHB9	−0.0292	$p = 0.0014$*	$p = 0.0015$*
SVD (case:control = 64:64)			
CDH2	−0.0101	$p = 0.0022$*	$p = 0.0343$*
PCDHB10	−0.0190	$p = 0.0322$*	$p = 0.1069$
PCDHB11	−0.0195	$p = 0.0502$	$p = 0.1195$
PCDHB14	−0.0131	$p = 0.1816$	$p = 0.3445$
PCDHB16	−0.0151	$p = 0.1758$	$p = 0.3626$
PCDHB3	−0.0093	$p = 0.3182$	$p = 0.5970$
PCDHB6	−0.0031	$p = 0.8273$	$p = 0.9188$
PCDHB9	−0.0117	$p = 0.2385$	$p = 0.4120$

*Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). *Statistically significant difference ($p < 0.05$). LAA, large-artery atherosclerosis; SVD, small vessel disease.*

especially with higher specificity (0.8723), sensitivity (0.883), accuracy (0.8777), and AUC (0.9384).

Validation of the mRNA Expression of Differentially Methylated Genes

In the replication stage, there were 7 differentially hypomethylated genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) in the IS group. Generally, a decrease in DNA methylation leads to abnormal gene transcription, resulting in the upregulation of gene expression. We validated the expression levels of these 7 genes in the GSE22255 dataset. The expression of two genes (*PCDHB9* and *PCDHB11*) increased, while the methylation level of the same 2 genes decreased in IS patients compared to controls ($p < 0.05$, **Figure 5**). The expression levels of the other three genes (*CDH2/PCDHB14/PCDHB16*) also increased in the IS group, but there were no significant differences between the two groups ($p > 0.05$, **Additional Files S7–11**).

TABLE 6 | Stratified analysis by ischemic stroke location.

Target	Group difference	*p* value	Adjust *p* value
Anterior circulation-left hemisphere (case:control = 72:72)			
CDH2	−0.0133	$p = 0.0005$*	$p = 0.0015$*
PCDHB10	−0.0200	$p = 0.0201$*	$p = 0.0278$*
PCDHB11	−0.0191	$p = 0.0419$*	$p = 0.0400$*
PCDHB14	−0.0155	$p = 0.0997$	$p = 0.2873$
PCDHB16	−0.0111	$p = 0.2982$	$p = 0.1031$
PCDHB3	−0.0093	$p = 0.3007$	$p = 0.1201$
PCDHB6	0.0034	$p = 0.8128$	$p = 0.3847$
PCDHB9	−0.0237	$p = 0.0214$*	$p = 0.0144$*
Anterior circulation-right hemisphere (case:control = 63:63)			
CDH2	−0.0103	$p = 0.0005$*	$p = 0.0058$*
PCDHB10	−0.0249	$p = 0.0026$*	$p = 0.0037$*
PCDHB11	−0.0257	$p = 0.0075$*	$p = 0.0060$*
PCDHB14	−0.0219	$p = 0.0067$*	$p = 0.0470$*
PCDHB16	−0.0204	$p = 0.0644$	$p = 0.0243$*
PCDHB3	−0.0224	$p = 0.0051$*	$p = 0.0076$*
PCDHB6	−0.0248	$p = 0.1181$	$p = 0.0276$*
PCDHB9	−0.0257	$p = 0.0052$*	$p = 0.0090$*
Posterior circulation (case:control = 48:48)			
CDH2	−0.0174	$p = 0.0022$*	$p = 0.0363$*
PCDHB10	−0.0243	$p = 0.0066$*	$p = 0.0113$*
PCDHB11	−0.0265	$p = 0.0064$*	$p = 0.0196$*
PCDHB14	−0.0187	$p = 0.0557$	$p = 0.0717$
PCDHB16	−0.0190	$p = 0.1315$	$p = 0.1532$
PCDHB3	−0.0172	$p = 0.0661$	$p = 0.1380$
PCDHB6	−0.0112	$p = 0.5626$	$p = 0.9945$
PCDHB9	−0.0268	$p = 0.0075$*	$p = 0.0205$*

*Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). *Statistically significant difference ($p < 0.05$).*

DISCUSSION

Certainly, IS is a heterogeneous disease with a genetic predisposition (Acosta et al., 2021). Emerging evidence has revealed the importance of epigenetic regulations in IS, particularly DNA methylation (Gómez-Úriz et al., 2015; Davis Armstrong et al., 2018; Shen et al., 2019; Soriano-Tárraga et al., 2020). Our study presented genome-wide alterations in DNA methylation in IS patients and controls by identifying a total of 462 functional DMPs corresponding to 373 annotated genes and revealed that hypomethylated sites were eightfold more numerous than DNA hypermethylation sites in IS cases, demonstrating that the hypomethylated modification was predominant in IS.

For DNA methylation profiling, GO analysis showed that these genes form interconnected networks involved in the

TABLE 7 | Associations between DNA methylation in 7 differentially methylated genes and risk of ischemic stroke.

Target	OR	95%CI	*p* value	Adjust *p* value
CDH2				
Q2 vs. Q1	8.844	3.776–20.710	$p < 0.0001$*	$p < 0.0001$*
Q3 vs. Q1	6.092	2.593–14.309	$p < 0.0001$*	$p < 0.0001$*
Q4 vs. Q1	24.615	9.518–63.660	$p < 0.0001$*	$p < 0.0001$*
PCDHB10				
Q2 vs. Q1	2.634	1.363–5.089	$p = 0.0040$*	$p = 0.0004$*
Q3 vs. Q1	3.627	1.775–7.413	$p = 0.0004$*	$p = 0.0001$*
Q4 vs. Q1	7.797	3.707–16.399	$p < 0.0001$*	$p < 0.0001$*
PCDHB11				
Q2 vs. Q1	2.694	1.371–5.296	$p = 0.0040$*	$p = 0.0016$*
Q3 vs. Q1	4.103	1.982–8.497	$p = 0.0001$*	$p = 0.0003$*
Q4 vs. Q1	6.346	2.963–13.592	$p < 0.0001$*	$p < 0.0001$*
PCDHB14				
Q2 vs. Q1	1.402	0.774–2.540	$p = 0.2653$	$p = 0.3518$
Q3 vs. Q1	4.035	2.051–7.937	$p < 0.0001$*	$p = 0.0055$*
Q4 vs. Q1	4.048	2.053–7.983	$p < 0.0001$*	$p = 0.0028$*
PCDHB16				
Q2 vs. Q1	2.862	1.537–5.327	$p = 0.0009$*	$p = 0.0007$*
Q3 vs. Q1	1.319	0.671–2.595	$p = 0.4221$	$p = 0.2949$
Q4 vs. Q1	3.355	1.743–6.456	$p = 0.0003$*	$p = 0.0002$*
PCDHB3				
Q2 vs. Q1	1.390	0.764–2.529	$p = 0.2813$	$p = 0.0897$
Q3 vs. Q1	2.032	1.125–3.671	$p = 0.0187$*	$p = 0.0055$*
Q4 vs. Q1	2.965	1.561–5.630	$p = 0.0009$*	$p = 0.0016$*
PCDHB9				
Q2 vs. Q1	2.402	1.221–4.725	$p = 0.0111$*	$p = 0.0008$*
Q3 vs. Q1	3.500	1.799–6.807	$p = 0.0002$*	$p = 0.0010$*
Q4 vs. Q1	5.336	2.656–10.720	$p < 0.0001$*	$p < 0.0001$*

*Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). *Statistically significant difference ($p < 0.05$). OR odds ratio; CI, confidence interval.*

calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules, homophilic cell adhesion via plasma membrane adhesion molecules, cell-cell adhesion via plasma-membrane adhesion molecules and other networks. Among them, the most enriched term in GO analysis was "calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules" item including 8 candidate genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB6/PCDHB9*), which were related to the cadherins.

Cadherins are a superfamily of calcium-dependent adhesion molecules mainly involved in tissue and embryonic cell development (Schaarschuch and Hertel, 2018). They can be classified into several subfamilies such as classic cadherins (CDHs), protocadherins (PCDHs), desmosomal cadherins and so on (Pancho et al., 2020). The *PCDH* gene family comprises three gene clusters (*PCDHA, PCDHB*, and *PCDHG*, respectively) (Peek et al., 2017; Canzio and Maniatis, 2019). PCDHs are highly expressed in the central nervous system in neurons, astrocytes, pericytes and brain microvascular endothelial cells, and mediate various developmental processes, including synaptic maintenance, neuronal survival, and spatial patterning of axons and dendrites (Gabbert et al., 2020; Mancini et al., 2020; Ing-Esteves et al., 2018). PCDHs have been reported to be relevant to neurodegenerative diseases, epilepsy, schizophrenia and mood disorders (Schaarschuch and Hertel, 2018; Langfelder et al., 2016; Li et al., 2017; Pederick et al., 2018; Lencz et al., 2021; Flaherty and Maniatis, 2020). However, the association between PCDHs and IS is rarely reported. Ana et al. found altered *PCDHGA3* gene expression was strongly associated with reduced stroke volume (Ortega et al., 2016). Some studies revealed that cadherins are involved in the regulation of angiogenesis and inflammation in endothelial cells and lack of PCDHs may lead to the destruction of the blood-brain barrier (Gabbert et al., 2020; Nanes et al., 2012), which are the mechanisms of IS. However, no direct evidence of the association between *PCDHB* genes and IS has been shown in previous studies. The semi-stochastic expression of *PCDH* genes is regulated by DNA methylation (Hirayama and Yagi, 2017). In our results, the methylation level of *PCDHB (PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) genes was lower in IS cases compared with controls, which suggested the methylation status of *PCDHB* genes may be involved in the pathogenesis of IS and be potential diagnostic biomarkers for IS.

Like the PCDHs, classic CDHs have been implicated in neurulation, brain development, and regulation of synaptic function (Polanco et al., 2021; de Agustín-Durán et al., 2021; Sanes and Zipursky, 2020). They are also involved in the formation of atherosclerotic plaques (He et al., 2017), which is an important clinical feature of IS. One of the most important members of CDHs is the neuronal cadherin (N-cadherin, *CDH2*) (Schaarschuch and Hertel, 2018; Martinez-Garay, 2020). So far, there has been no direct evidence to show the relationship

TABLE 8 | ROC analysis of 7 differentially methylated genes and polygenic methylation model.

Target	Adjust AUC	Adjust Sensitivity	Adjust Specificity	Adjust Accuracy
CDH2	0.9067	0.883	0.7872	0.8351
PCDHB10	0.9015	0.8617	0.7979	0.8298
PCDHB11	0.8924	0.8298	0.8191	0.8245
PCDHB14	0.8679	0.8191	0.8191	0.8191
PCDHB16	0.874	0.8245	0.8191	0.8218
PCDHB3	0.8793	0.8245	0.8245	0.8245
PCDHB9	0.891	0.8617	0.7872	0.8245
Polygenic model	0.9384	0.883	0.8723	0.8777

Adjusted factors: smoking, drinking, previous medical history of hypertension and diabetes mellitus, and plasma lipid levels (total triglyceride, total cholesterol, high-density lipoprotein, and low-density lipoprotein). AUC area under the curve; ROC receiver operating characteristic.

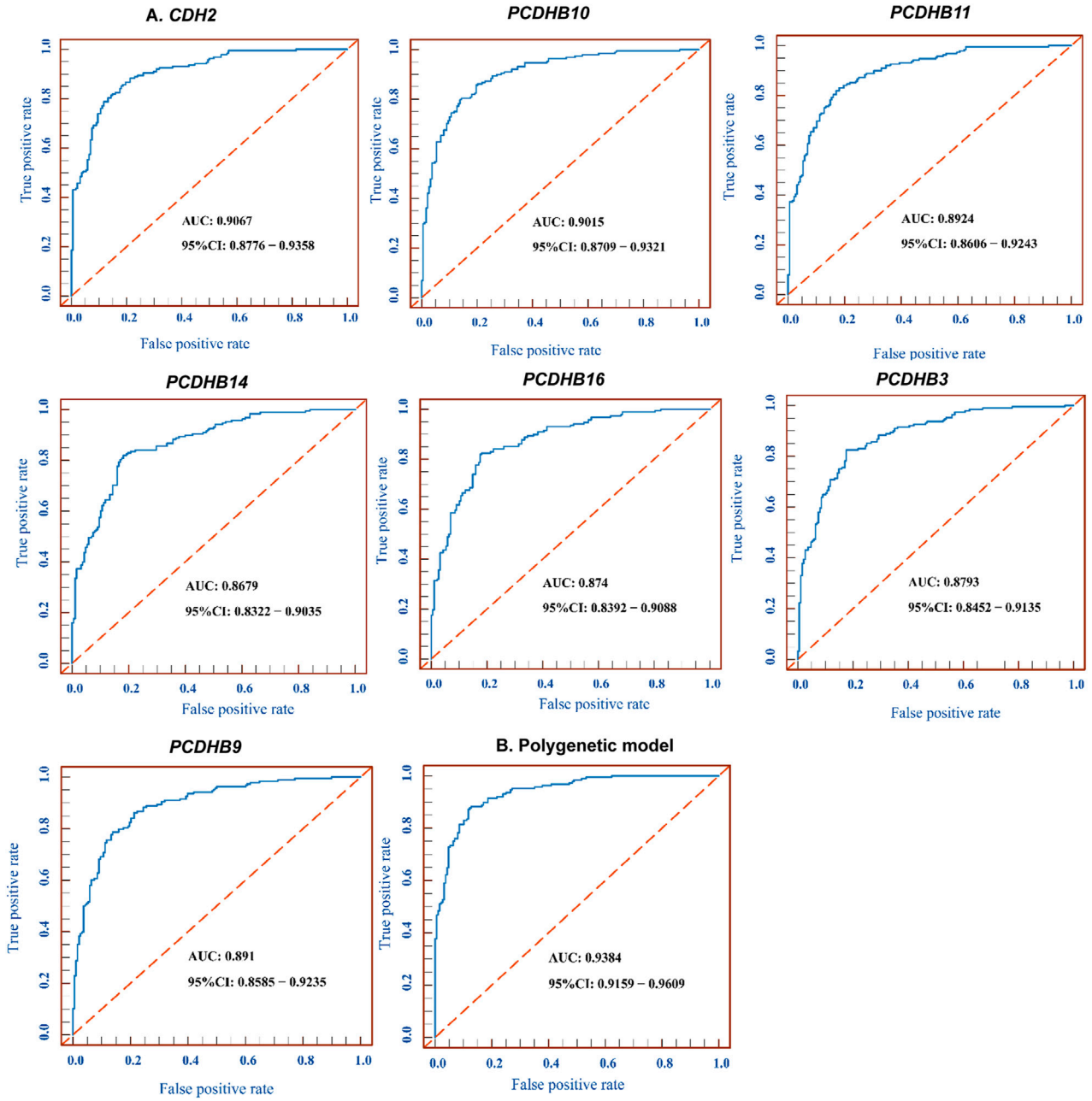


FIGURE 4 | ROC for 7 differentially methylated genes and polygenic methylation model. **(A)** ROC curve measuring sensitivity (*Y* axis) and 1-specificity (*X* axis) of methylation of *CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9* genes as biomarkers for discriminating IS from controls. **(B)** ROC curve of the polygenic methylation model. AUC, area under the curve; ROC, receiver operating characteristic; CI, confidence interval.

between *CDH2* gene and stroke. A Study about myocardial infarction revealed overexpression of *CDH2*, *CDH12*, *PCDH17*, and *PCDH18* in myocardial infarction vascular smooth muscle cells compared with controls (Derda et al., 2018). In this study we found decreased methylation level of the promoter region in *CDH2* gene. Generally, hypomethylation in promotor is believed to upregulate gene transcription, which is consistent with the previous study of myocardial infarction in a way. This is the first time we have found the correlation between the methylation of *CDH2* gene and IS.

There are also some interesting findings in our study. First, when we split the samples by sex, we found fewer differentially methylated genes (4 genes) in females, which we thought might be because the sample size of women is much smaller than that of men. What's more, compared with SVD-control groups, there are more differentially methylated genes in LAA-control groups. We

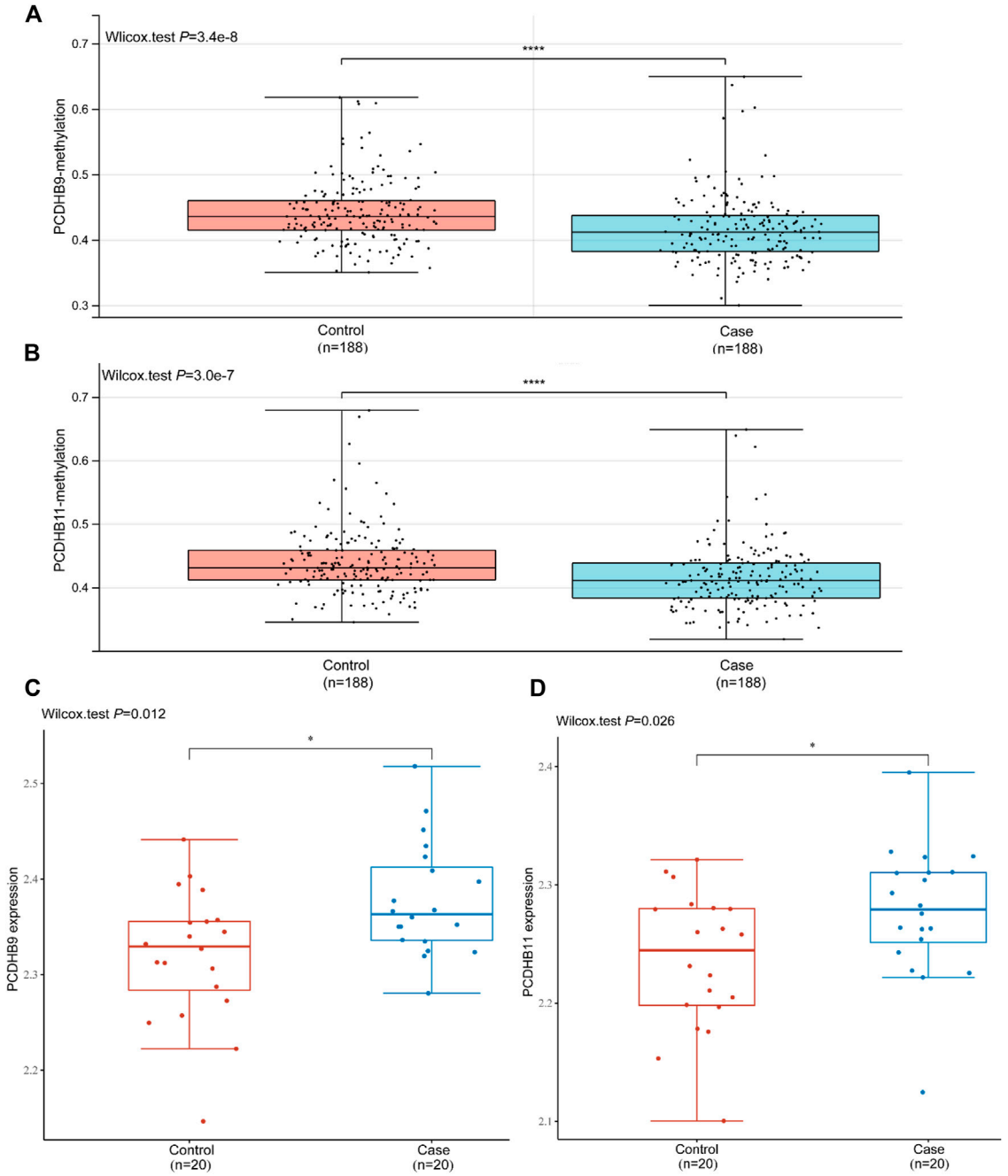


FIGURE 5 | Validation of the mRNA expression of *PCDHB9* and *PCDHB11* in GSE22255. **(A)**: Methylation level of the *PCDHB9* gene in the control and case (IS) groups; **(B)**: Methylation level of the *PCDHB11* gene in the control and case (IS) groups; **(C)**: mRNA expression of the *PCDHB9* gene in the control and case (IS) groups; **(D)**: mRNA expression of the *PCDHB11* gene in the control and case (IS) groups. *Statistically significant difference ($p < 0.05$). ****Statistically significant difference ($p < 0.0001$).

thought it might hint that the LAA subtype may be more susceptible to epigenetic regulation, and further studies are needed. Lastly, there existed more differentially methylated genes in the right hemisphere-control groups than the other two paired groups, which need further studies in a large sample size.

The peripheral blood is a good choice for epigenetic research of IS, as it is easy to obtain with minor invasion (Qin et al., 2019). Moreover, IS is a disease related to the vasculature and interrupting blood supply to the brain. Liu et al. found that stroke patients had lower methylation level of the *ACTB* gene in blood (Liu et al., 2021), which revealed that peripheral blood could identify the methylation aberrations associated with stroke.

There were certain limitations in our study. First, gene expression was verified in a public dataset, but not explored

in our blood samples. Our study aimed to describe the global DNA methylation patterns in IS and to explore potential diagnostic biomarkers for IS in a Chinese population. Although gene and protein expression was not the main purpose of our research, it will be carried out in follow-up studies. In addition, we couldn't detect the methylation level of more CpG sites due to the limited funds. However, they will be the directions for our future study.

CONCLUSION

The present study demonstrated the changes in genome-wide DNA methylation between IS cases and controls and identified 7 novel DNA methylation genes (*CDH2/PCDHB10/PCDHB11/PCDHB14/PCDHB16/PCDHB3/PCDHB9*) related to IS in the replication stage. These data may provide preliminary evidence for further exploring the role of DNA methylation in IS.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of ethical restrictions. Requests to access the datasets should be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Study design: JZ and YT. Blood collection: HS, JX, BH and YL. Data check: YZ, YS, HS, FL, JW and AF. Manuscript drafting: HS. Manuscript edition: HS. Administrative and material support: JZ and YT. Study supervision: JZ and YT. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant number 81771508); Scientific research innovation fund of the First Affiliated Hospital of Harbin Medical University (Grant number 2020B05).

ACKNOWLEDGMENTS

We thank all participants for their cooperation throughout the whole study. We acknowledged the technical support from the Shanghai Genesky Biotechnology Company (Shanghai, China).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.844141/full#supplementary-material

REFERENCES

Acosta, J. N., Szejko, N., and Falcone, G. J. (2021). Mendelian Randomization in Stroke: A Powerful Approach to Causal Inference and Drug Target Validation. *Front. Genet.* 12, 683082. doi:10.3389/fgene.2021.683082

Ajoolabady, A., Wang, S., Kroemer, G., Penninger, J. M., Uversky, V. N., and Pratico, D. (2021). Targeting Autophagy in Ischemic Stroke: From Molecular Mechanisms to Clinical Therapeutics. *Pharmacol. Ther.* 225, 107848. doi:10.1016/j.pharmthera.2021.107848

Bevan, S., Traylor, M., Adib-Samii, P., Malik, R., Paul, N. L. M., Jackson, C., et al. (2012). Genetic Heritability of Ischemic Stroke and the Contribution of Previously Reported Candidate Gene and Genomewide Associations. *Stroke* 43, 3161–3167. doi:10.1161/STROKEAHA.112.665760

Bird, A. (2007). Perceptions of Epigenetics. *Nature* 447, 396–398. doi:10.1038/nature05913

Canzio, D., and Maniatis, T. (2019). The Generation of a Protocadherin Cell-Surface Recognition Code for Neural Circuit Assembly. *Curr. Opin. Neurobiol.* 59, 213–220. doi:10.1016/j.conb.2019.10.001

Davis Armstrong, N. M., Chen, W. M., Brewer, M. S., Williams, S. R., Sale, M. M., Worrall, B. B., et al. (2018). Epigenome-Wide Analyses Identify Two Novel Associations with Recurrent Stroke in the Vitamin Intervention for Stroke Prevention Clinical Trial. *Front. Genet.* 9, 358. doi:10.3389/fgene.2018.00358

de Agustín-Durán, D., Mateos-White, I., Fabra-Beser, J., and Gil-Sanz, C. (2021). Stick Around: Cell–Cell Adhesion Molecules during Neocortical Development. *Cells* 10, 118. doi:10.3390/cells10010118

Derda, A. A., Woo, C. C., Wongsurawat, T., Richards, M., Lee, C. N., Kofidis, T., et al. (2018). Gene Expression Profile Analysis of Aortic Vascular Smooth Muscle Cells Reveals Upregulation of Cadherin Genes in Myocardial Infarction Patients. *Physiol. Genomics* 50, 648–657. doi:10.1152/physiolgenomics.00042.2017

Flaherty, E., and Maniatis, T. (2020). The Role of Clustered Protocadherins in Neurodevelopment and Neuropsychiatric Diseases. *Curr. Opin. Genet. Develop.* 65, 144–150. doi:10.1016/j.gde.2020.05.041

Gabbert, L., Dilling, C., Meybohm, P., and Burek, M. (2020). Deletion of Protocadherin Gamma C3 Induces Phenotypic and Functional Changes in Brain Microvascular Endothelial Cells *In Vitro*. *Front. Pharmacol.* 11, 590144. doi:10.3389/fphar.2020.590144

Gómez-Úriz, A. M., Milagro, F. I., Mansego, M. L., Cordero, P., Abete, I., De Arce, A., et al. (2015). Obesity and Ischemic Stroke Modulate the Methylation Levels of KCNQ1 in white Blood Cells. *Hum. Mol. Genet.* 24, 1432–1440. doi:10.1093/hmg/ddu559

He, X., Li, D. R., Cui, C., and Wen, L. J. (2017). Clinical Significance of Serum MCP-1 and VE-Cadherin Levels in Patients with Acute Cerebral Infarction. *Eur. Rev. Med. Pharmacol. Sci.* 21, 804–808.

Hirayama, T., and Yagi, T. (2017). Regulation of Clustered Protocadherin Genes in Individual Neurons. *Semin. Cel Develop. Biol.* 69, 122–130. doi:10.1016/j.semcdb.2017.05.026

Ing-Esteves, S., Kostadinov, D., Marocha, J., Sing, A. D., Joseph, K. S., Laboulaye, M. A., et al. (2018). Combinatorial Effects of Alpha- and Gamma-Protocadherins on Neuronal Survival and Dendritic Self-Avoidance. *J. Neurosci.* 38, 2713. doi:10.1523/JNEUROSCI.3035-17.2018

Katan, M., and Luft, A. (2018). Global Burden of Stroke. *Semin. Neurol.* 38, 208–211. doi:10.1055/s-0038-1649503

Krupinski, J., Carrera, C., Muiño, E., Torres, N., Al-Baradie, R., Cullell, N., et al. (2017). DNA Methylation in Stroke. Update of Latest Advances. *Comput. Struct. Biotechnol. J.* 16, 1–5. doi:10.1016/j.csbj.2017.12.001

Langfelder, P., Cantle, J. P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., et al. (2016). Integrated Genomics and Proteomics Define Huntingtin CAG Length–dependent Networks in Mice. *Nat. Neurosci.* 19, 623–633. doi:10.1038/nn.4256

Lencz, T., Yu, J., Khan, R. R., Flaherty, E., Carmi, S., Lam, M., et al. (2021). Novel Ultra-rare Exonic Variants Identified in a Founder Population Implicate Cadherins in Schizophrenia. *Neuron* 109, 1465–1478. doi:10.1016/j.neuron.2021.03.004

Li, Y., Chen, Z., Gao, Y., Pan, G., Zheng, H., Zhang, Y., et al. (2017). Synaptic Adhesion Molecule Pcdh-Γc5 Mediates Synaptic Dysfunction in Alzheimer's Disease. *J. Neurosci.* 37, 9259. doi:10.1523/jneurosci.1051-17.2017

Liu, C., Yin, Q., Li, M., Fan, Y., Shen, C., and Yang, R. (2021). ACTB Methylation in Blood as a Potential Marker for the Pre-clinical Detection of Stroke: A Prospective Nested Case-Control Study. *Front. Neurosci.* 15, 644943. doi:10.3389/fnins.2021.644943

Liu, X., Fan, B., Chopp, M., and Zhang, Z. (2020). Epigenetic Mechanisms Underlying Adult Post Stroke Neurogenesis. *Int. J. Mol. Sci.* 21, 6179. doi:10.3390/ijms21176179

Mahjoubin-Tehran, M., Rezaei, S., Jesmani, A., Birang, N., Morshedi, K., Khanbabaei, H., et al. (2021). New Epigenetic Players in Stroke Pathogenesis: From Non-coding RNAs to Exosomal Non-coding RNAs. *Biomed. Pharmacother.* 140, 111753. doi:10.1016/j.biopha.2021.111753

Majnik, A. V., and Lane, R. H. (2014). Epigenetics: where Environment, Society and Genetics Meet. *Epigenomics* 6, 1–4. doi:10.2217/epi.13.83

Mancini, M., Bassani, S., and Passafaro, M. (2020). Right Place at the Right Time: How Changes in Protocadherins Affect Synaptic Connections Contributing to the Etiology of Neurodevelopmental Disorders. *Cells* 9, 2711. doi:10.3390/cells9122711

Martinez-Garay, I. (2020). Molecular Mechanisms of Cadherin Function during Cortical Migration. *Front. Cel. Dev. Biol.* 8, 588152. doi:10.3389/fcell.2020.588152

Nanes, B. A., Chiasson-MacKenzie, C., Lowery, A. M., Ishiyama, N., Faundez, V., Ikura, M., et al. (2012). p120-catenin Binding Masks an Endocytic Signal Conserved in Classical Cadherins. *J. Cel Biol* 199, 365–380. doi:10.1083/jcb.201205029

Ortega, A., Gil-Cayuela, C., Tarazón, E., García-Manzanares, M., Montero, J. A., Cinca, J., et al. (2016). New Cell Adhesion Molecules in Human Ischemic Cardiomyopathy. PCDHGA3 Implications in Decreased Stroke Volume and Ventricular Dysfunction. *PLoS One* 11, e0160168. doi:10.1371/journal.pone.0160168

Pancho, A., Aerts, T., Mitsogiannis, M. D., and Seuntjens, E. (2020). Protocadherins at the Crossroad of Signaling Pathways. *Front. Mol. Neurosci.* 13, 117. doi:10.3389/fnmol.2020.00117

Pederick, D. T., Richards, K. L., Piltz, S. G., Kumar, R., Mincheva-Tasheva, S., Mandelstam, S. A., et al. (2018). Abnormal Cell Sorting Underlies the Unique X-Linked Inheritance of PCDH19 Epilepsy. *Neuron* 97, 59–66. e55. doi:10.1016/j.neuron.2017.12.005

Peek, S. L., Mah, K. M., and Weiner, J. A. (2017). Regulation of Neural Circuit Formation by Protocadherins. *Cell Mol. Life Sci.* 74, 4133–4157. doi:10.1007/s00018-017-2572-3

Polanco, J., Reyes-Vigil, F., Weisberg, S. D., Dhimitruka, I., and Brusés, J. L. (2021). Differential Spatiotemporal Expression of Type I and Type II Cadherins Associated with the Segmentation of the Central Nervous System and Formation of Brain Nuclei in the Developing Mouse. *Front. Mol. Neurosci.* 14, 633719. doi:10.3389/fnmol.2021.633719

Qin, X., Li, J., Wu, T., Wu, Y., Tang, X., Gao, P., et al. (2019). Overall and Sex-specific Associations between Methylation of the ABCG1 and APOE Genes and Ischemic Stroke or Other Atherosclerosis-Related Traits in a Sibling Study of Chinese Population. *Clin. Epigenetics* 11, 189. doi:10.1186/s13148-019-0784-0

Sanes, J. R., and Zipursky, S. L. (2020). Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits. *Cell* 181, 536–556. doi:10.1016/j.cell.2020.04.008

Schaarschuch, A., and Hertel, N. (2018). Expression Profile of N-Cadherin and Protocadherin-19 in Postnatal Mouse Limbic Structures. *J. Comp. Neurol.* 526, 663–680. doi:10.1002/cne.24359

Schiano, C., Benincasa, G., Franzese, M., Della Mura, N., Pane, K., Salvatore, M., et al. (2020). Epigenetic-sensitive Pathways in Personalized Therapy of Major Cardiovascular Diseases. *Pharmacol. Ther.* 210, 107514. doi:10.1016/j.pharmthera.2020.107514

Shen, Y., Peng, C., Bai, Q., Ding, Y., Yi, X., Du, H., et al. (2019). Epigenome-Wide Association Study Indicates Hypomethylation of MTRNR2L8 in Large-Artery Atherosclerosis Stroke. *Stroke* 50, 1330–1338. doi:10.1161/STROKEAHA.118.023436

Soriano-Tárraga, C., Lazcano, U., Giralt-Steinhauer, E., Avellaneda-Gómez, C., Ois, Á., Rodríguez-Campello, A., et al. (2020). Identification of 20 Novel Loci Associated with Ischaemic Stroke. Epigenome-wide Association Study. *Epigenetics* 15, 988–997. doi:10.1080/15592294.2020.1746507

Wan, S., Liu, L., Ren, B., Qu, M., Wu, H., Jiang, W., et al. (2021). DNA Methylation Patterns in the HLA-DPB1 and PDCD1LG2 Gene Regions in Patients with Autoimmune Thyroiditis from Different Water Iodine Areas. *Thyroid* 31, 1741–1748. doi:10.1089/thy.2021.0221

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cell and Developmental Biology

TYPE Methods
PUBLISHED 31 August 2022
DOI 10.3389/fcell.2022.995590

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Wanjun Zhang,
National Center for Protein Science, China
Zhongwei Xu,
Logistics University of People's Armed Police Force, China
Zexian Liu,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Minjia Tan,
mjtan@simm.ac.cn
Yinan Zhang,
yinanzhang@njucm.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 16 July 2022
ACCEPTED 01 August 2022
PUBLISHED 31 August 2022

CITATION
Zhai L, Wang L, Hu H, Liu Q, Lee S, Tan M and Zhang Y (2022), PBC, an easy and efficient strategy for high-throughput protein C-terminome profiling. *Front. Cell Dev. Biol.* 10:995590. doi: 10.3389/fcell.2022.995590

PBC, an easy and efficient strategy for high-throughput protein C-terminome profiling

Linhui Zhai[1,2,3†], Le Wang[1,3†], Hao Hu[3], Quan Liu[3], Sangkyu Lee[4], Minjia Tan[1,3*] and Yinan Zhang[1,2*]

[1]School of Chinese Materia Medica, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China, [2]Jiangsu Key Laboratory for Functional Substances of Chinese Medicine, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China, [3]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China, [4]College of Pharmacy and Research Institute of Pharmaceutical Sciences, Kyungpook National University, Daegu, South Korea

High-throughput profiling of protein C-termini is still a challenging task. Proteomics provides a powerful technology for systematic and high-throughput study of protein C-termini. Various C-terminal peptide enrichment strategies based on chemical derivatization and chromatography separation have been reported. However, they are still costly and time-consuming, with low enrichment efficiency for C-terminal peptides. In this study, by taking advantage of the high reaction selectivity of 2-pyridinecarboxaldehyde (2-PCA) with an α-amino group on peptide N-terminus and high affinity between biotin and streptavidin, we developed a 2-PCA- and biotin labeling–based C-terminomic (PBC) strategy for a high-efficiency and high-throughput analysis of protein C-terminome. Triplicates of PBC experiments identified a total of 1,975 C-terminal peptides corresponding to 1,190 proteins from 293 T cell line, which is 180% higher than the highest reported number of C-terminal peptides identified from mammalian cells by chemical derivatization–based C-terminomics study. The enrichment efficiency (68%) is the highest among the C-terminomics methods currently reported. In addition, we not only uncovered 50 proteins with truncated C-termini which were significantly enriched in extracellular exosome, vesicle, and ribosome by a bioinformatic analysis but also systematically characterized the whole PTMs on C-terminal in 293 T cells, suggesting PBC as a powerful tool for protein C-terminal degradomics and PTMs investigation. In conclusion, the PBC strategy would benefit high-efficiency and high-throughput profiling of protein C-terminome.

KEYWORDS

C-terminomics, chemical derivatization, enrichment, high-efficiency, post-translation modification (PTM)

Introduction

Protein N-termini and C-termini play important roles in diverse biological processes such as protein stability, protein localization, protein–protein interaction, and macromolecular complexes formation (Marino et al., 2015; Perrar et al., 2019; Winter et al., 2021). The high-throughput study of protein termini and their posttranslational modifications (PTMs) is important for understanding their functions. (Marino et al., 2015; Klein et al., 2018; Chen and Kashina, 2021; Chi et al., 2021). Proteomics technologies have emerged as a powerful tool for the systematical and high-throughput analysis of protein termini. Thus far, various strategies have been developed for protein termini enrichment (Huesgen and Overall, 2012; Rogers and Overall, 2013; Koudelka et al., 2021). However, current methods for protein C-terminome profiling still lag far behind the N-terminomics technologies. More than 7,400 N-terminal peptides were reported to be identified without pre-fractionation in human lymphoblastoid B cell line (Klein et al., 2015), while the reported highest number of identified C-terminal peptides was 3,129 with the requirement of extensive off-line HPLC fractionation (24 fractions) in HeLa cells. (Wang et al., 2021) There are several technical difficulties leading to the lower coverage of C-terminome than that of N-terminome. First, it was reported that more than 60% human protein lack lysine or arginine residues near C-termini (Wang et al., 2021). For these proteins, the C-terminal peptides generated by the widely used proteases in proteomics study (i.e., trypsin and LysC) do not contain appropriate length for efficient LC-MS/MS detection. Second, the lack of basic amino acid residue will affect the ionization efficiency in positive mode for mass spectrometry detection. Third, because of the low chemical reaction reactivity of the C-terminal α-carboxyl group, the development of chemical derivatization–based C-terminal enrichment approach is largely restricted.

Current C-terminal peptide enrichment strategies are mainly classified into two different types. One is chromatography-based C-terminal peptide direct enrichment. This strategy is based on the physiochemical difference between C-terminal peptides and other internal peptides, such as the hydrophilic/hydrophobic properties and isoelectric point (Dormeyer et al., 2007; Van Damme et al., 2010; Wang et al., 2021). In order to increase the physiochemical difference between the C-terminal peptides and non-C-terminal peptides, the amidation of carboxyl group on protein C-terminal or propionylation of amino group on peptide was performed prior to chromatographic separation (Kaleja et al., 2019; Li et al., 2020a). Though the chromatography-based strategy provides an easy way for C-terminal peptides enrichment, the selectivity and efficiency of these methods are still not satisfactory, due to the low separation resolution for highly complex peptide mixtures. This strategy also suffers from intensive labor and instrumentation cost, which requires lots of off-line prefractions and high MS instrument time.

Another C-terminal peptide enrichment strategy is the chemical derivatization–based negative enrichment method. This strategy is based on chemical derivatization to protect carboxyl group on original protein C-terminal at the protein level; then, proteins were digested into peptides and the internal peptides with free carboxyl group were removed by the polyallylamine polymer. The C-TAILS (C-terminal amine–based isotope labeling of substrate) method was first reported for C-terminal peptide enrichment by Schilling et al. (2010). In this method, the α-amine groups on protein N-termini were first blocked by dimethylation, and the carboxyl groups on protein C-termini were then blocked by ethanolamidation. After the digestion of proteins with trypsin, the neo-α-amine groups exposed from the internal peptides were further blocked by dimethylation. Finally, the neo-internal peptides containing free carboxyl groups were coupled and depleted with the poly(allylamine) polymer. The C-terminal peptides were then enriched. After the introduction of the C-TAILS method, different kinds of C-TAILS–based methods were further developed. Zhang et al. (2015) used Ac-NHS to block the α-amine group on the protein level and used ethanolamine to block the carboxyl group before using a high molecular polymer to negatively enrich C-terminal peptides. This method achieved a higher yield of chemical derivatization and identified more C-terminal peptides than the original C-TAILS method. We developed the LAACTer method on the basis of C-TAILS (Hu et al., 2019). Our LAACTer method combined LysargiNase digestion, chemical reaction, and ion-aided proteome database searching for an in-depth C-terminomic study and finally identified 164% and quantified 300% more C-terminal peptides than those using the original C-TAILS method from 293 T cells.

Although the reported chemical derivatization–based C-terminome methods provide powerful technologies to systematically study the C-terminal peptides, most of them requires at least three steps of chemical reaction on protein and peptide levels; thus, they are labor- and time-consuming. The amidation on the carboxyl group used in these methods are largely low-specific, which influences the C-terminal peptides enrichment efficiency (Zhang et al., 2015). In addition, the LysargiNase used in the LAACTer method is expensive and not conventionally used in the proteomics study. So far, a high-efficiency and high-throughput analysis of C-terminal peptides is still challenging. In order to make the chemical derivatization–based strategies more practical and efficient, we developed a new method, namely, 2-pyridinecarboxaldehyde (2-PCA)- and biotin labeling–based C-terminomics (PBC), for high-throughput and highly efficient enrichment of C-terminal peptides. In this study, we found that the peptide length and hydrophilic properties could significantly affect the 2-PCA labeling efficiency on the α-amine group. Through the PBC

method, we obtained the highest number of C-terminal peptides and highest enrichment efficiency in chemical derivatization–based C-terminomics up to date. We also systematically revealed the PTMs on C-terminal peptides and C-terminal truncated proteins in 293 T cells by combined usage of the PBC strategy and an open-search method. In this regard, our newly developed PBC method provides a powerful tool to efficiently study the C-terminome from a low amount of samples.

Methods

Cell culture

The human embryonic kidney cell line HEK 293 T was cultured in Dulbecco's modified Eagle's medium (DMEM). After the cells grew to 80% density in a 10-cm petri dish, the medium was washed out and the cells were harvested by centrifugation at 1,000 *g* for 5 min under room temperature. Then the cell pellets were washed with cold PBS buffer twice.

Proteome sample preparation

HEK 293 T cell pellets were suspended in lysis buffer [6 M guanidine hydrochloride, 100 mM HEPES (pH 8.0), and 1% (v/v) protease inhibitor cocktail (Roche, Swiss)] on ice for 30 min, followed by sonication for 2 min with 2 s on and 5 s off at 30% power. Then the lysates were centrifuged at 20,000 g at 4°C for 10 min and the supernatant was collected. Protein concentration was measured by using the BCA assay (Beyotime, China). For PBC technical evaluation, the 900 μg extracted proteins were used and divided into three equal parts (300 μg protein/each part). The proteins were reduced by using 5 mM dithiothreitol (DTT) at 56°C for 30 min and alkylated by using 15 mM iodoacetamide (IAA) in darkness at 25°C for 30 min; 20 mM DTT was added to the protein solution to quench the excess IAA. Then the proteome sample was digested with LysC (Hualishi, China) with an enzyme/protein ratio of 1:50 (w/w) at 37°C overnight.

2-PCA labeling and sulfo NHS-biotin labeling

For 2-PCA labeling, 10 mM 2-PCA (J&K Scientific, China) dissolved in 100 mM HEPES buffer (pH 8.0) was added to the 300 μl peptide solution (300 μg peptide) at 37°C for 16 h (MacDonald et al., 2015) For sulfo NHS-biotin labeling, 2 mM sulfo NHS-biotin (APExBIO, United States) dissolved in 100 mM HEPES (pH 8.5) was added to the 2-PCA–labeled peptide sample at 37°C for 30 min. The labeled-peptide samples were dried in SpeedVac and then desalted by using Sep-Pak C_{18} cartridges (50 mg sorbent per cartridge, Waters, United States).

Enrichment of C-terminal peptides

A volume of 1 ml streptavidin beads (GE Healthcare, United States) were washed twice with 800 μl PBS. The 2-PCA- and biotin-labeled peptide was re-suspended to 800 μl PBS, pH was adjusted to 7.5, and then it was incubated with streptavidin beads at room temperature for 1 h. The supernatant was collected by centrifugation at 300 g for 2 min, and the beads were washed with 800 μl PBS twice. The supernatant and washing solution was combined and dried in SpeedVac.

C-terminal peptides fractionation and desalting

The C-terminal peptides were fractionated by using a home-made StageTip C_{18} column. The StageTip C_{18} column was made as follows. C_{18} disks (3M, United States) were cut by a hypodermic needle and pushed into P200 pipet tips. Then 2mg C_{18} resin (Durashell C_{18}, Agela, China) was re-suspended in 200 μl acetonitrile (ACN), loaded into prepared pipet tips, and then centrifuged at 400 g for 10 min. The StageTip column was equilibrated with 150 μl water of 0.1% ammonium hydroxide ($NH_3{\cdot}H_2O$) for three times, respectively; the centrifuge time was controlled in 10 min. Then the peptide sample was loaded onto the column, the column was washed by 150 μl water (0.1% $NH_3{\cdot}H_2O$), and the peptides were eluted with 2%, 9%, 15%, 20%, 24%, 30%, and 80% ACN in water (0.1% $NH_3{\cdot}H_2O$). The fractionations were dried in SpeedVac and desalted with ZipTip C_{18} (Millipore, United States).

LC-MS/MS analysis

The sample was analyzed by using an EASY-nLC 1200 HPLC tandem with the Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, United States). The peptide was resolved in buffer A (2% ACN in water and 0.1% formic acid) and separated by using a home-made C_{18} capillary column (25 cm × 75 μm, 1.9 μm particle size, and 100 Å pore size) (Li et al., 2020b), A column oven was used and the heating temperature was set at 60°C (Kyte and Doolittle, 1982).

For PCA- and biotin-labeled evaluation, the peptide sample before and after labeling were detected by using 1 h gradient LC-MS/MS. The LC gradient was set as follows: 8 %–13% buffer B (90% ACN in water and 0.1% formic acid) for 20 min; 26% buffer B for 31 min, with a raise to 45% in 5 min; and finally 80% buffer B for 60 min. The flow rate was set to 300 nL/min. Then ions were scanned over 350–1,300 m/Z at a resolution of 12,000 (200 m/Z)

with the automatic gain control (AGC) target of 5.0e5 and maximum injection time of 50 ms. The charge state included was 2–6 and dynamic exclusion was 60 s. The data-dependent mode was set up with a cycle time of 3 s, and MS2 data were acquired by higher-energy collisional dissociation (HCD) fragmentation and normalized collision energy (NCE) of 32%. The AGC target was set to 7.0e3 and maximum injection time was set to 35 ms.

The enriched C-terminal peptide was detected by 2 h gradient LC-MS/MS. The LC gradient was set as follows: 2%–5% buffer B for 3 min; to 16% buffer B for 52 min; to 35% buffer B for 50 min, with a raise to 47% for 10 min; and finally 80% buffer B for 120 min. The flow rate was set as 300 nL/min. Then peptides were scanned over 350–1,800 m/z at a resolution of 60,000 with the automatic gain control (AGC) target of 3e6 and maximum injection time of 45 ms. The charge state included was 1–5 and dynamic exclusion was 20 s. The data-dependent mode was set up with the top 10 most abundant precursors and subjected to MS/MS fragmentation, and MS2 data were acquired by higher-energy collisional dissociation (HCD) fragmentation and normalized collision energy (NCE) of 28%. The AGC target was set to 1e5 and the maximum injection time was set to 80 ms.

Database searching and bioinformatic analysis

The raw data were searched against the UniProt homo proteome database (version 201,812) through Proteome Discoverer (version 2.2, Thermo Fisher Scientific) and loaded into the Mascot search engine (version 2.3, Matrix Science). The enzyme type was Lys-C/P. Up to two maximum missed cleavage was used. The precursor mass tolerance was set as 10 ppm, and the fragment mass tolerance was set as 0.02 Da. For samples before enrichment, carbamidomethyl (C) was set as static modification, and acetyl (protein N-term) and oxidation (M) were set as dynamic modifications. In addition, PCA (N-term), biotinylation (K), and biotinylation (N-term) were set as dynamic modifications for analyzing sulfo NHS-biotin–labeled samples. For sample of enrichment, carbamidomethyl (C) and biotinylation (K) were set as static modifications, and PCA (N-term) and oxidation (M) were set as dynamic modifications. The minimal peptide length was filtered with six amino acids. The results were filtered by a 1% false discovery rate (FDR) at PSM, peptide, and protein levels (Zhang et al., 2018; Wang et al., 2021).

For the open-search, all raw data files were processed using pFind software (version 3.1.0) with an open-search mode (Kahl et al., 2018). The enzyme type was set as Lys-C/P, and the maximum missed cleavage was 2. The precursor mass tolerance was 10 ppm and fragment mass tolerance was 10 ppm. The results were filtered by a 1% false discovery rate (FDR) at both PSM and protein levels (Chi et al., 2018; Guangcan et al., 2021).

The determination of 2-PCA and biotin labeling efficiency analyses and other character results were carried out with

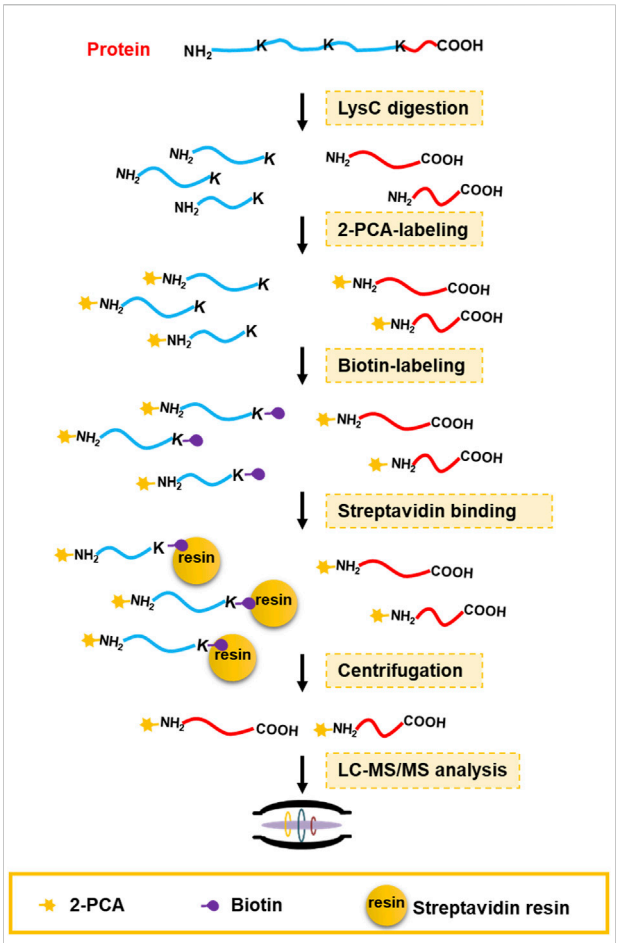


FIGURE 1
C-terminal peptide enrichment workflow through the PBC strategy. The extracted whole cell proteome is digested into peptides by LysC protease. Then the LysC-digested peptides are tandemly reacted with 2-PCA and NHS-biotin. Next, the peptides with biotin-labeled lysine were captured by streptavidin beads. Finally, the C-terminal peptides are enriched by filter-aided separation from streptavidin beads and detected by LC-MS/MS.

GraphPad (version 8.0). All statistical tests were analyzed using student *t*-tests. The GRAVY scores were calculated using the online tool (https://web.expasy.org/protparam/) (Kyte and Doolittle, 1982). The analysis of peptide sequences was conducted by iceLogo (Colaert et al., 2009), and a p-value < 0.05 was used. The bioinformatic analysis was performed using DAVID (version 6.8) with an adjusted p-value < 0.05 (Huang et al., 2007).

Data availability

All the original proteomics raw data and proteome database result files in this study have been deposited to the iProX Consortium with the subproject ID IPX0003710000.

Results and Discussion

Strategy design for 2-PCA- and biotin labeling–based C-terminomic

Chemical derivatization on protein or peptide is an enabling technology for a proteomics study. 2-PCA reported could selectively label the α-amino group on protein/peptide N-terminus over the ε-amino group of lysine residue through an N-terminal amine-specific cyclization reaction (MacDonald et al., 2015). By taking advantage of such reaction selectivity of 2-PCA, we designed a PBC strategy for high-throughput profiling of protein C-termini by bottom-up shotgun proteomics (Figure 1). The extracted whole cell proteome is digested into peptides by LysC protease, which cleaves peptide bond C-terminal to Lys residue. As a result, each non-C-terminal peptide ends with lysine and thus contains two amino groups on the peptide N-terminus and the lysine side chain. In contrast, C-terminal peptides lack lysine and only contain one α-amino group on peptide N-termini. Next, the peptides are tandemly reacted with 2-PCA and NHS-biotin. Ideally, all the α-amino groups on the peptide N-terminal are blocked with 2-PCA and ε-amino groups on the lysine side chain are labeled with biotin. The chemically derivatized peptides are then incubated with streptavidin beads. The original protein N-terminal peptides and internal peptides containing biotin-labeled lysine are all captured by streptavidin beads. Finally, the C-terminal peptides are enriched by filter-aided separation from streptavidin beads and detected by LC-MS/MS. Thus, our PBC strategy not only took advantage of LysC, which has higher digestion efficiency and lower cost than other commonly used proteases (LysargiNase and ArgC) (Giansanti et al., 2016) but also employed only two chemical derivatization steps at the peptide level, in contrast to other reported negative C-terminal peptide enrichment strategies requiring at least three chemical derivation steps at both protein and peptide levels (Schilling et al., 2010; Zhang et al., 2015; Hu et al., 2019). Therefore, the PBC strategy is less labor- and time-consuming for C-terminal peptide enrichment.

2-PCA could efficiently and highly selectively label the peptide N-terminal

In a previous study (MacDonald et al., 2015), only limited number of peptides (20 peptides) with similar sequence composition (only varying at N-terminal amino acid) were used to evaluate the characteristics of the reaction between 2-PCA and the peptide N-terminal α-amino group. Whether other factors could affect the selectivity and efficiency of this reaction needs to be deeply studied by using a large-scale and highly complex peptide sample. In this study, we first used the mass spectrometry–based shotgun proteomics approach to systemically study the characteristics of the selective reaction with the α-amino group on the peptide N-terminus. The highly complex peptides were generated by using LysC digested whole HEK 293 T cell proteome. The evaluation for 2-PCA labeling efficiency was performed in technical triplicates from the same proteome sample. The peptides were reacted with 2-PCA and then detected by LC-MS/MS. Compared to the unlabeled sample, the number of identified protein and peptide decreased in the 2-PCA–labeled sample under same LC-MS/MS conditions. We reasoned that the LC gradient for 2-PCA–labeled peptide sample was unsuitable and needed to be further optimized because the peptide retention time was changed after 2-PCA labeling (Supplementary Figures S1A,B; Supplementary Tables S1, S2). In order to evaluate whether 2-PCA could react with the ε-amino group on lysine, we both set the 2-PCA labeling at the peptide N-terminal and lysine as variable modification for database searching. The bioinformatic analysis results showed 2-PCA dominantly labeled at the peptide N-terminal, while only 2.3% lysines were labeled (Supplementary Figure S1C; Supplementary Table S3). Such results agreed with previously reported result that 2-PCA labeling could selectively react with the α-amino group on the peptide N-terminal. (MacDonald et al., 2015)

Then we evaluated the N-terminal labeling efficiency (the percentage of 2-PCA–labeled peptides in all peptides). Among 4,704 identified non-redundant peptides (peptides of the same sequence with different 2-PCA modification status were considered as two different peptides), 75% peptides (3,529 peptides) were fully labeled with 2-PCA, 8.1% peptides (380 peptides) were partially labeled, and 16.9% peptides (795 peptides) were completely unlabeled (Figure 2A, Supplementary Table S2). For the 795 unlabeled peptides, 25.1% peptides (199 peptides) were N-terminally acetylated and 48.0% (382 peptides) were identified to possess proline at the second amino acid position. We used spectral counting–based quantification to calculate the N-terminal labeling efficiency for the 380 partially 2-PCA–labeled peptides and found 122 peptides (occupied 32.1%) with labeling efficiency higher than 90% and 187 peptides (occupied 49.4%) with labeling efficiency between 50% and 90%. As mentioned before, by excluding the N-acetylated and proline-containing peptides which could not react with 2-PCA in theory, we found more than 90.7% peptides were fully labeled or with labeling efficiency higher than 80%. Such results showed that 2-PCA could react with peptide N-terminal with high efficiency and high selectivity.

We next analyzed factors that influenced the 2-PCA-labeling efficiency, such as peptide length, hydrophilic/hydrophobic properties, and amino acids composition. As shown in Figure 2B, the peptide length significantly affected the 2-PCA–labeling efficiency, with short peptides more preferred. The length of unlabeled peptides was significantly longer than the fully labeled or partially labeled peptides. We

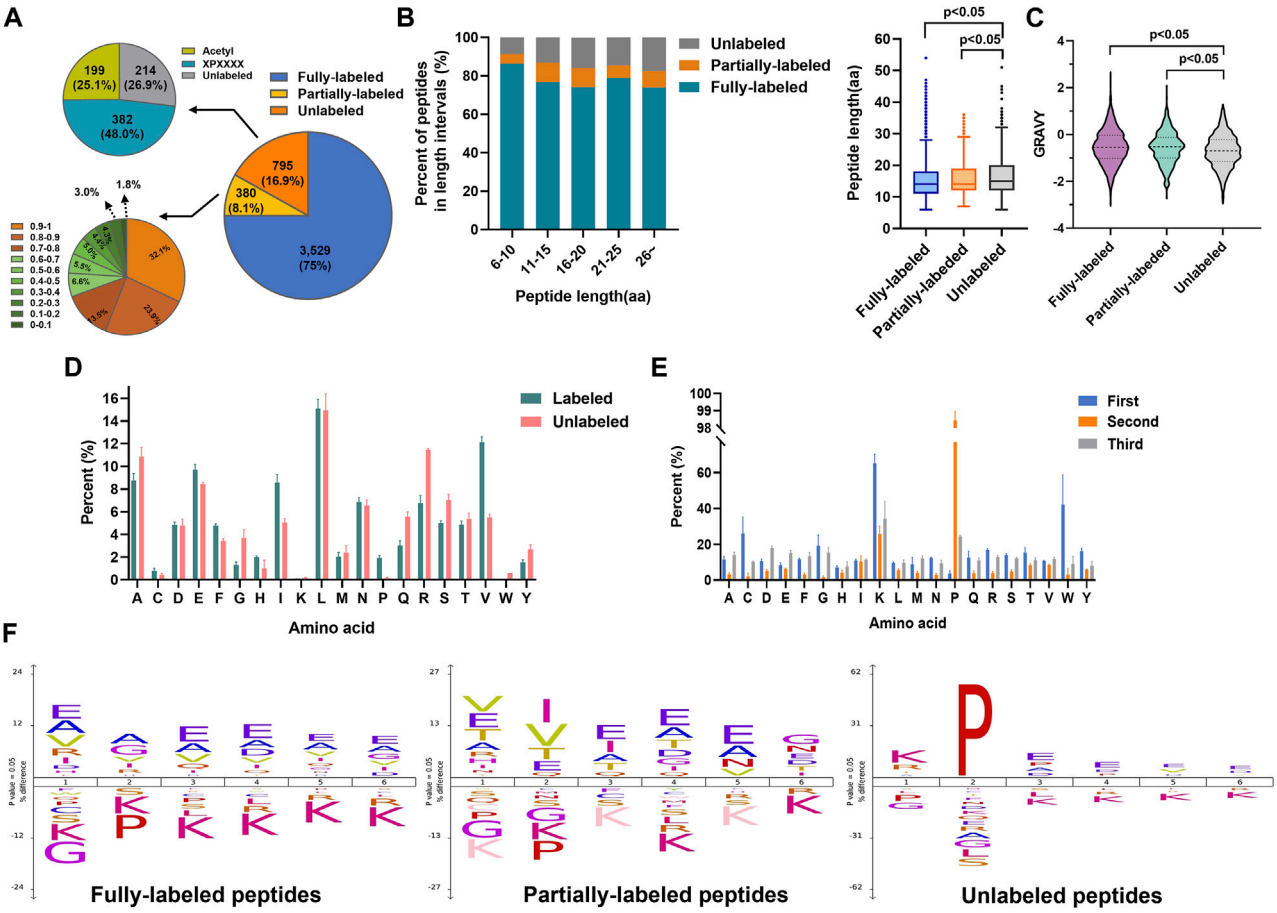


FIGURE 2
2-PCA displays efficiently and high selectively labeling peptide N-terminal. **(A)** Pie chart of the 2-PCA labeling efficiency distribution calculated by using the spectral counting-based quantification method. **(B,C)** 2-PCA labeling efficiency of different length **(B)** and hydrophilic/hydrophobic value of peptide **(C)**. The GRAVY scores were calculated using the online tool (https://web.expasy.org/protparam/) (Kyte and Doolittle, 1982). The *p*-value was calculated by paired Student *t*-test. **(D,E)** Full labeling **(D)** and un-labeling **(E)** peptide percent on different first three amino acids. **(F)** IceLogo analysis representation of the first six amino acids of full labeling, partial labeling, and un-labeling peptides.

also explored the relationship between peptide hydrophilicity (evaluated by GRAVY score) and labeling efficiency (Figure 2C). The data showed the peptides with a higher hydrophilic value achieved higher labeling efficiency. We reasoned the peptides with shorter and higher hydrophilic values could be more easily dissolved in aqueous buffer, which resulted in higher reaction efficiency.

The 2-PCA labeling efficiency was reported to be impaired when the second position of peptide was proline. However, whether different amino acids on other positions could influence the labeling efficiency is still unknown. In this study, we systemically analyzed the influence of first three amino acids in the peptide on reaction efficiency. We found the reaction efficiency was higher when the first amino acid as L/A/E than others (Figures 2D,E, Supplementary Figures S1D,E). Perhaps, less steric hindrance (A/L) and more hydrophilicity (E) of the first amino acid (A) benefited the chemical conversion of 2-PCA coupling. Interestingly, our results showed glycine as the first amino acid supplied less steric hindrance but dramatically inhibited the "reaction efficiency". Such lower reaction efficiency observed for peptides containing N-terminal glycine may be mainly due to the Thorpe–Ingold effect that led to kinetically slower cyclization (Kaneti et al., 2004). Our results showed proline as the second amino acid could inhibit the reaction, which was agreed with the previous study reported (MacDonald et al., 2015). In addition, we found lysine as the first amino acid could also affect the reaction efficiency. The iceLogo analysis indicated the peptides with first amino acid as A/E were overrepresented in fully labeled peptides (Figure 2F). These results indicate the 2-PCA labeling with high α-amino group labeling efficiency is applicable for the proteomic analysis.

10.3389/fcell.2022.995590

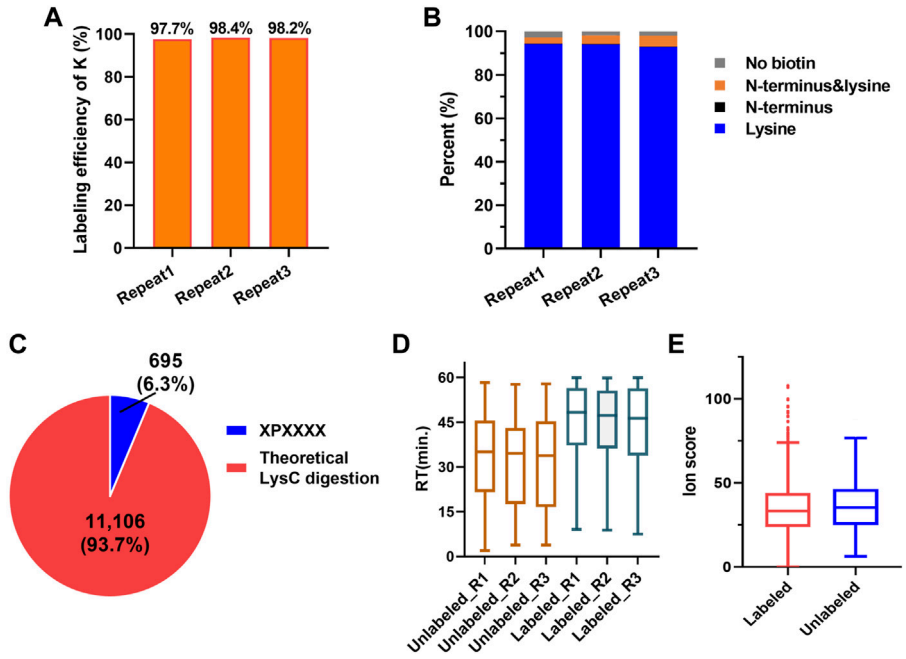


FIGURE 3
2-PCA and biotin labeling displays high-efficiency labeling for α-amino on peptide N-termini and ε-amino group on lysine. **(A)** Histogram distribution of biotin labeling efficiency on amino group in lysine in three technical replicates. More than 97.7% ε-amino group on K was blocked. **(B)** Histogram distribution of blocking efficiency on amino group in protein N-terminal in three technical replicates. More than 99.7% α-amino group on the peptide N-terminal were blocked. **(C)** Percentage of the theoretic "2-PCA-labeling" and "un-labeling" human protein C-terminal peptide by *in silico* digestion using LysC. Only 695 C-terminal peptides with second amino acid as proline could not be labeled by 2-PCA. **(D)** Peptide retention time comparison among before and after 2-PCA labeling. **(E)** Peptide ion score comparison among before and after 2-PCA labeling. All the experiments were performed in technical triplicates from the same started proteome sample.

2-PCA and biotin labeling efficiently blocks amino group on the peptide N-terminal and the lysine side chain

According to the PBC strategy design, the blocking efficiency of the amino group on the peptide N-terminal and the lysine side chain is the key to highly efficient C-terminal peptide enrichment. In order to evaluate the amino group blocking efficiency of the peptide tandem reaction with 2-PCA and biotin, we performed the C-terminal peptide blocking experiments in technical triplicates from same started proteome. We found more than 99.7% α-amino group on the peptide N-terminal and more than 97.7% ε-amino group on K were blocked (Figures 3A,B, Supplementary Table S4). Interestingly, the peptides with second amino acid as proline that could not react with PCA were labeled with biotin, which would help to remove such peptide by streptavidin beads and increase the C-terminal peptide enrichment efficiency. In addition, we analyzed the whole human proteome digested with LysC *in silico* and found only 695 C-terminal peptides with second amino acid as proline, which account for 6.3% in all 11,106 C-terminal peptides (comprised of 6–50 amino acid residues) (Figure 3C). So, we reasoned the C-terminal peptide with proline as the second position amino acid was minimal, which largely would not influence C-terminome peptide profiling using the PBC strategy. Since the chemical derivatization of peptides will influence the chromatography separation and peptide scoring (reflecting the peptide identification confidence), the retention time and the ion score of the same peptide with and without 2-PCA labeling were systematically compared. We found that the retention time was significantly increased after labeling. The retention time distribution clearly showed that the labeled peptides were dominantly eluted in the later stage of the LC gradient (Figure 3D, Supplementary Figure S1F). We reasoned the 2-PCA molecule contains a pyridine ring structure, which would increase the peptide hydrophobic value after labeling. The peptide matching scores showed no significance difference between peptide with and without 2-PCA labeling (Figure 3E), which indicates the 2-PCA labeling would not influence the C-terminal peptides identification by our PBC strategy.

PBC strategy represents the highest enrichment efficiency among the reported C-terminomic strategies

We used the PBC strategy to enrich C-terminal peptides from cell sample and systematically evaluated its performance. In total,

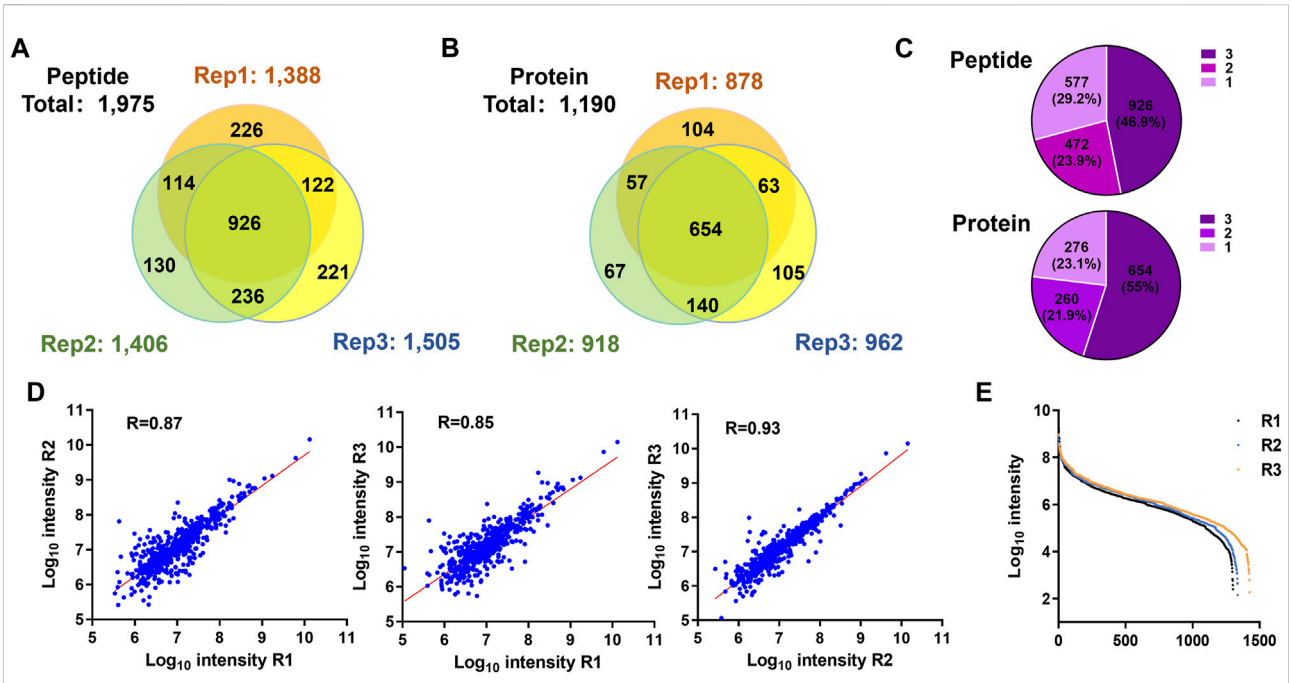


FIGURE 4
Highly efficient and high-throughput identification of protein C termini from 293T cells by using the PBC strategy. **(A,B)** Overlap analysis of C-terminal peptides **(A)** and proteins **(B)** from the triplicate experiment of C-terminal peptide enriched by using the PBC strategy. **(C)** Pie chart analysis of the frequency of C-terminal peptides identification from technical triplicates. A total of 46.9% peptides (926 peptides) were identified in all the triplicates, and 70.8% peptides (1,398 peptides) were identified in at least two replicates. **(D)** Correlation of intensity analysis of identified C-terminal peptides among technical triplicates. **(E)** Peptide abundance dynamic range of technical triplicates through the PBC strategy. The abundance dynamic range was larger than seven orders of magnitude.

300 μg proteins were used as the starting materiel, and three technical replicates were performed. Finally, we identified an average of 1,433 C-terminal peptides belonging to 919 proteins in each experiment (Figures 4A,B, Supplementary Table S5). It should be noted that our results showed a lot of C-terminal peptides were not blocked by 2-PCA at N-termini. We systematically evaluated the hydrophilic/hydrophobic values and peptide lengths of the enriched C-terminal peptides. We found the unlabeled C-terminal peptides displaying significantly higher hydrophobic values and peptide lengths than the 2-PCA–labeled or partially labeled C-terminal peptides (Supplementary Figure S2). These results were all consistent with our conclusions that the efficiency of 2-PCA labeling was related to the peptide hydrophilic/hydrophobic value in Figure 2C. A total of 1,975 C-terminal peptides belonging to 1,190 proteins were identified, which achieved 180% more C-terminal peptide identification than the highest reported C-terminal peptide number based on the chemical derivatization–based C-terminomics study. (Hu et al., 2019) We found that 46.9% peptides (926 peptides) were identified in all the triplicates, and 70.8% peptides (1,398 peptides) were identified in at least two replicates (Figure 4C). We also evaluated the quantification linear correlation among the replicates and found the Pearson correlation coefficients of each two pair among the three replicates were all remarkably high (>0.85) (Figure 4D). These results indicated that the PBC strategy provided a high-throughput and reproducibility for C-terminal peptide enrichment. The abundance dynamic range of the enriched C-terminal peptides from the PBC strategy was larger than seven orders of magnitude (Figure 4E). We compared our results with published C-terminome datasets of human cells (Table 1, Supplementary Figure S3A). Our PBC strategy achieved the highest identification number of C-terminal peptides among all reported chemical derivatization–based strategies. In addition, the C-terminal peptide enrichment efficiency of our PBC method is 68% (the percent of C-terminal peptides among all identified peptides), which is highest in the C-terminomics study up to date. Thus, although the enriched C-terminal peptides were not completely labeled with 2-PCA, the PBC provides a novel and simple strategy for high efficiency and reproducibility C-terminal peptides enrichment.

Overview of C-terminome on 293 T cells

The C-terminomic study could construct the landscape of C-termini in cell with high throughput and efficiency, which

TABLE 1 Comparison the results of human protein C-terminomics studies.

#	This study	Hu et al. (2019)	Li et al. (2020a)	Wang et al., (2021)
Cell type	293 T	293 T	293 T	HeLa
Sample amount	900 μg	1.5 mg	120 μg	1 mg
Enzyme	LysC	LysargiNase	LysargiNase	Trypsin
Number of C-terminal peptides	1,975	1,100	2,000	4,724
Selective enrichment efficiency (%)	68	39.1	22.4	2.7–34
Number of protein C-termini	1,190	924	1,812	2,219
Isolation method	Two chemical reaction steps	Three chemical reaction steps	Three chemical reaction steps + SCX	SCX

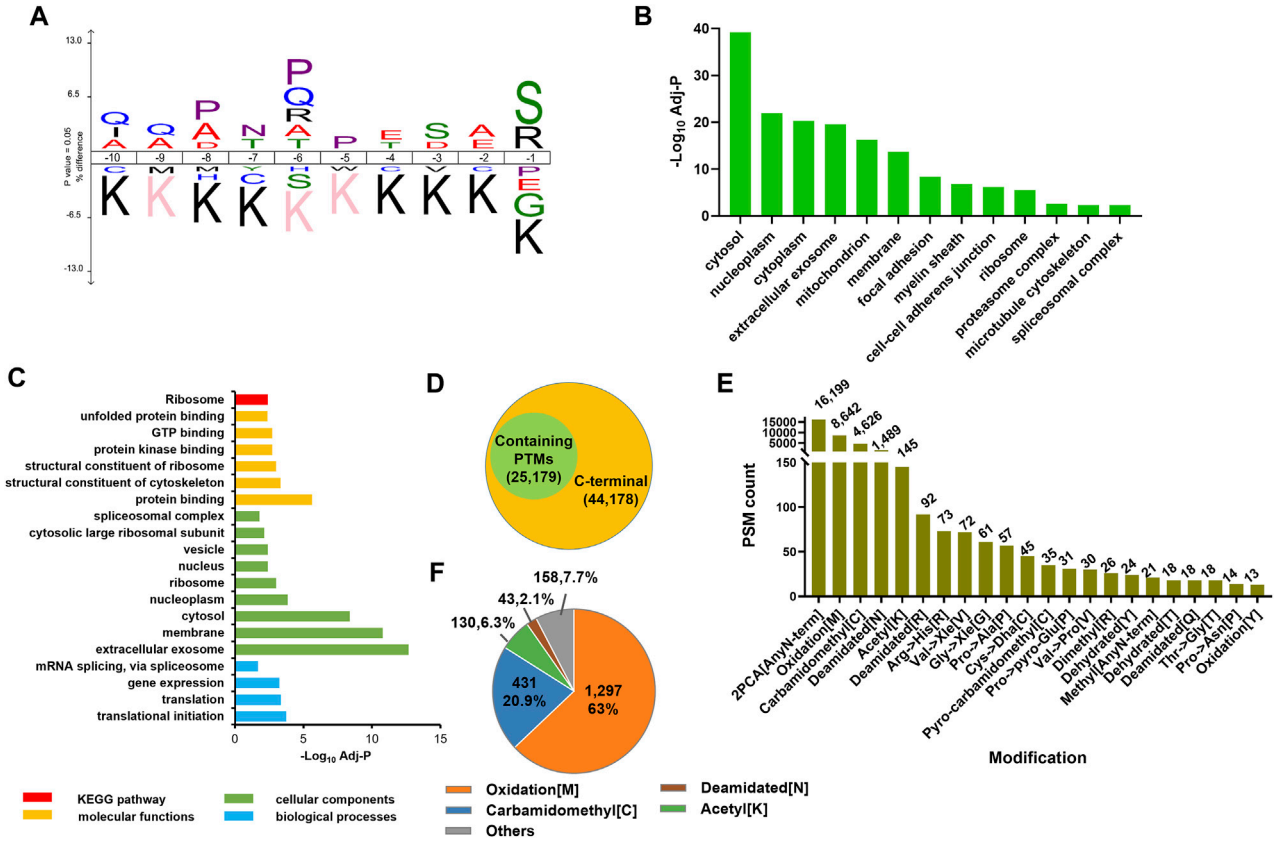


FIGURE 5
Function analysis of the enriched C-terminal truncated proteins and the PTM distribution on C-terminal peptide. **(A)** IceLogo analysis of the last 10 amino acids of original C-terminal identified by PBC strategy. **(B)** Cellular components enrichment analysis of C-terminal proteins identified by PBC strategy. **(C)** The gene ontology and KEGG pathway enrichment analysis of the proteins with C-terminal truncated. The truncated proteins significantly enriched in extracellular exosome, vesicle, and ribosome. **(D)** Pie chart analysis of identified PSM with and without containing PTMs. In total, more than 30% PSM were identified with PTM-labeled peptides. **(E)** Distribution of PTMs modified on C-terminal peptide. **(F)** Distribution of identified PTMs on last amino acid of protein C-termini.

could be used to study the C-end rules in cells with high accuracy. In our study, the iceLogo analysis of the C-terminal dataset showed that the P/Q/S is highly enriched in the protein C-terminal sequence (Figure 5A). This result was consistent with our previous study using the LAACter method. (Hu et al., 2019) In addition, the lysine was observed under-represented at −10 to −1 position, which could be explained by that the peptide sample used in the PBC strategy was lysC digestion generated. In order to study whether the C-terminal peptide enrichment bias existed in our PBC strategy, gene

TABLE 2 Identified Neo-N-termini for HSP90AB1 (P08238) and H4 (P62805).

Gene name	Protein accession number	Annotated sequence
HAP90AB1	P08238	[K].LGLGIDEDEVAA.[E]
		[K].LGLGIDEDEVAAEEPNAAVPD.[E]
		[K].LGLGIDEDEVAAEEPNAAVPDEIPPLEGD.[E]
		[K].LGLGIDEDEVAAEEPNAAVPDEIPPLEGDED.[A]
		[K].LGLGIDEDEVAAEEPNAAVPDEIPPLEGDEDASR.[M]
		[K].LGLGIDEDEVAAEEPNAAVPDEIPPLEGDEDASRMEEVD.[-]
H4	P62805	[K].VFLENVIR.[D]
		[K].VFLENVIRD.[A]
		[K].VFLENVIRDA.[V]
		[K].VFLENVIRDAV.[T]
		[K].VFLENVIRDAVTY.[T]
		[K].RQGRTLYGF.[G]
		[K].RQGRTLYGFG.[G]
		[K].RQGRTLYGFGG.[-]

ontology enrichment was used to systematically evaluate our C-terminal dataset. We found the enriched C-terminal peptides distributed among different cellular compositions (Figure 5B, Supplementary Table S6), such as cytosol, membrane, ribosome, and nuclear. "C-end rules" was reported as the key factor influencing protein functions in cell (Lin et al., 2018). In addition, we also systemically compared the distribution of first three amino acids in enriched C-terminal peptide and theoretical C-terminal peptide. The theoretical C-terminal peptides were generated from *in silico* digestion of whole human proteome by using LysC (Figure 3C). The results showed the similar distribution of first three N-terminal amino acids between enriched and theoretical C-terminal peptide (Supplementary Figures S3B–D), which also indicated the PBC strategy could enrich C-terminal peptides without bias.

Our C-terminomic study enabled us to examine C-terminal truncated proteins (Supplementary Table S7). Totally, 217 C-terminal truncation peptides corresponding to 50 proteins were identified. For example, six truncation peptides were identified from HSP90AB1 (Table 2). The annotated MS/MS spectra of these six truncation peptides are showed in Supplementary Figure S3E. In addition, we also found the truncated histone protein H4 in our data (Table 2). The annotated MS/MS spectra of the truncation peptides of H4 are showed in Supplementary Figures S4A,B. In order to further analyze the cellular composition and molecular functions of these 50 truncated proteins, the gene ontology (GO) and KEGG pathway enrichment were performed (Figure 5C). We found these proteins significantly enriched in extracellular exosome, vesicle, and ribosome, suggesting these proteins mainly partook in the biological processing of translation and translational initiation. The ribosome pathway was also significantly enriched through the KEGG pathway analysis. We inferred the truncated proteins were mainly generated from two origins. One was from immature proteins due to the incomplete translation step in ribosome; another might come from the proteolytic fragments by exopeptidases and metallocarboxypeptidase enzyme in exosome or vesicle.

PTMs on protein C-terminal region play important roles in protein functions (Marino et al., 2015). However, the reported PTMs studies on C-terminal were limited due to the lack of in-depth and large-scale C-terminome data. In this study, we used the open-search method to analyze the PTMs distribution in our C-terminomic dataset. All the raw data were reanalyzed by using pFind 3.1 search engine through open-search searching mode. Totally 44,178 PSMs were identified and more than 30% PSM were identified with containing PTMs (Figure 5D, Supplementary Table S8). The PTM types and their frequency distributions is showed in Figure 5E. The dominant PTM was 2-PCA modification on peptide N-terminus, which was introduced by chemical divinization *in vitro* at the peptide level. Deamidation on asparagine was frequently occurred in our data. Deamidation was reported could occur spontaneously on proteins both *in vivo* and *in vitro* (Hao et al., 2011; Brown et al., 2017). *In vitro*, the deamidation rate of protein or peptides closely related with pH and temperature of sample preparation buffer. In our PBC method, the peptides were reacted with 2-PCA and biotin all performed at pH 8.5 under 37°C. We reasoned that such high frequency of deamidation were mainly artefacts introduced in sample preparation. Amino acid substitutions were reported to alter physiological properties of protein (Huang and Gromiha, 2010; Gromiha et al., 2019), such as the enzymatic stability and

protein folding rate. We found the acetylation on lysine and amino acid substitution highly occurred in the protein C-terminal region (Figures 5E,F). Interestingly, we also found many lysines at protein C-termini with acetylation, which was also reported benefit for protein stability (Wang et al., 2017). In addition, we also found 176 PSMs were identified as the PTM-labeled histone peptides (Supplementary Figure S4C). The main PTM types were deamination, amino acid substitution, and dehydration. Thus, the PBC strategy provides a high-throughput analysis of the PTMs on the protein C-terminal region.

Conclusion

Due to the high complexity of peptidome sample, the signal of C-terminal peptide could easily be suppressed by other internal peptides in LC-MS detection. Protein C-terminal peptide profiling is still challenging. In this study, we developed 2-PCA- and biotin-labeling based C-terminomic (PBC) proteomic strategy to globally profile C-terminal peptides. We first systemically investigated the characters of 2-PCA labeling on complex peptide sample by using the proteomic method and found the peptide length, peptide hydrophilic value, and first amino acid composition significantly influenced the reaction efficiency. Importantly, the peptidome-wide labeling efficiency and selectivity of 2-PCA sufficiently high for labeling peptide N-terminal amines. The PBC strategy used LysC to digest proteome and only two chemical derivatization steps for C-terminal peptide profiling, which is less labor- and time-consuming than current reported C-terminal peptide profiling. According to the PBC-based C-terminomic results, a total of 1,975 C-terminal peptides belonging to 1,190 proteins were identified. This strategy showed the highest C-terminal enrichment efficiency among all the reported strategies. Our C-terminome results also revealed neo-C-terminal on proteins and new PTMs on C-terminal peptides, which would help in-depth study of the protein C-terminome and uncover more characters and biological functions of protein C-termini.

It should be noted that few limitations exist in the PBC technology. First, the 2-PCA reaction could not completely block all the α-amino groups on peptide N-termini, which lead to sub-optimal 2-PCA labeling on C-terminal peptides in PBC method. However, the technology evaluation results showed that 80% of the C-terminal proteins were identified in at least two technical replicates, and the correlation of the enriched C-terminal peptides was higher than 0.85 (Figures 4C,D), which suggest that this method is reliable and robust for the enrichment of C-terminal peptides. Second, 6.3% C-terminal peptides in the whole proteome contain proline as the second amino acid in theory, which could not react with 2-PCA and be enriched by PBC method. Nonetheless, the PBC method provides a novel idea for C-terminal peptide enrichment based on combination of α-amino group blocking, lysC digestion, and negative selection. In addition, since C-terminal peptides are dominant in the enrichment peptides and with low background interference, the PBC method can be combined the label free quantification technology or carboxyl group selectively isotope labeling technology to realize the quantitative analysis of C-terminal peptides. For example, combining with isotope labeling quantification technology by using the isotope labeling reagent d_0-/d_6-2,4-dimethoxy-6-piperazin-1-yl pyrimidine (DMPP), which could label the carboxyl group on peptide with high efficiency and selectivity (Leng et al., 2013). In addition, the PBC method could be also further simplified and optimized into one step chemical derivatization for C-terminal peptide enrichment by combining with commercially available hyperbranched polyaldehyde polymers (HPG-ALD polymers) (Supplemntary Figure S5) (Kleifeld et al., 2010; Kleifeld et al., 2011). After PCA labeling, the protein N-terminal peptides and internal LysC-digested peptides could be depleted through HPG-ALD polymers directly to enrich the C-terminal peptides. The simplified method by HPG-ALD polymers is devoid of further biotinylation on lysine and using streptavidin beads to deplete non-C-terminal peptides. To sum up, despite the enriched C-terminal peptides were not completely labeled with 2-PCA, the PBC provides a novel and simple strategy for high efficiency and reproducibility of C-terminal peptides enrichment.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

LZ and LW contributed equally to this work.

Funding

This work was supported by grants from the National Key R and D Program of China (2020YFE0202200), Natural Science Foundation of China (Nos 32071432, 21877062, 32171434), Basic research projects of Shanghai Science and Technology Commission “science and technology innovation action plan” (No. 19JC1416300), open fund of state key laboratory of Pharmaceutical Biotechnology, Nanjing University, China (Grant Nos. KF-202201), and NSFC-NRF (China–Korea) Joint Research Program (No. 82111540276).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.995590/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The evaluation for 2-PCA labeling efficiency in LysC-digested peptides. The number of protein **(A)** and peptide **(B)** identification before and after 2-PCA labeling. **(C)** The 2-PCA labeling efficiency analysis of α-amino group on peptide N-terminal and ε-amino group on lysine. **(D,E)** The identified number distribution of full labeling **(D)** and un-labeling **(E)** peptide on different first three amino acids. **(F)** The histogram distribution of peptide retention time before and after labeling with 2-PCA.

SUPPLEMENTARY FIGURE S2
The influences of 2-PCA labeling efficiency analysis of C-terminal peptides. The hydrophilic/hydrophobic value **(A)** and peptide length **(B)** comparison of 2-PCA fully-labeled, partially labeled, and unlabeled C-terminal peptides. The p-value was calculated by paired Student *t*-test.

SUPPLEMENTARY FIGURE S3
The analysis of C-terminome data by using the PBC method. **(A)** Overlap analysis of our C-terminome data with other published data. **(B,C,D)** The distribution comparison of the first amino acid **(B)**, second amino acid **(C)**, and third amino acid **(D)** between enriched C-terminal peptide and theoretical C-terminal peptide. The theoretical C-terminal peptide was generated from *in silico* digestion of whole human proteome by using LysC. **(E)** The MS/MS spectrum for truncated C-terminal peptides from HSP90AB1.

SUPPLEMENTARY FIGURE S4
The analysis of truncations and PTMs on histone.

SUPPLEMENTARY FIGURE S5
The further optimized PBC method. The PBC method could be simplified and optimized into one-step chemical derivatization for C-terminal peptide enrichment. After PCA labeling, the protein N-terminal peptides and internal LysC peptides could be depleted through commercially available hyperbranched polyaldehyde polymers (HPG-ALD polymers) directly and the C-terminal peptides were enriched.

References

Brown, C. W., Sridhara, V., Boutz, D. R., Person, M. D., Marcotte, E. M., Barrick, J. E., et al. (2017). Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. *BMC Genomics* 18 (1), 301. doi:10.1186/s12864-017-3676-8

Chen, L., and Kashina, A. (2021). Post-translational modifications of the protein termini. *Front. Cell Dev. Biol.* 9, 719590. doi:10.3389/fcell.2021.719590

Chi, H., Dong, M.-Q., Zhenlin, C., Chao, L., Shangtong, L., Hao, C., et al. (2021). How to use open-pFind in deep proteomics data analysis? a protocol for rigorous identification and quantitation of peptides and proteins from mass spectrometry data. *Biophys. Rep.* 9, 207–226. doi:10.52601/bpr.2021.210004

Chi, H., Liu, C., Yang, H., Zeng, W.-F., Wu, L., Zhou, W.-J., et al. (2018). Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* 36 (11), 1059–1061. doi:10.1038/nbt.4236

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6 (11), 786–787. doi:10.1038/nmeth1109-786

Dormeyer, W., Mohammed, S., Breukelen, B. v., Krijgsveld, J., and Heck, A. J. R. (2007). Targeted analysis of protein termini. *J. Proteome Res.* 6 (12), 4634–4645. doi:10.1021/pr070375k

Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nat. Protoc.* 11 (5), 993–1006. doi:10.1038/nprot.2016.057

Gromiha, M. M., Nagarajan, R., and Selvaraj, S. (2019). "Protein structural bioinformatics: an overview," in *Encyclopedia of bioinformatics and computational Biology*. Editors S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach (Oxford: Academic Press), 445

Guangcan, S., Yong, C., Zhenlin, C., Chao, L., Shangtong, L., Hao, C., et al. (2021). How to use open-pFind in deep proteomics data analysis?— a protocol for rigorous identification and quantitation of peptides and proteins from mass spectrometry data. *Biophys. Rep.* 7 (3), 207–226. doi:10.52601/bpr.2021.210004

Hao, P., Ren, Y., Alpert, A. J., and Sze, S. K. (2011). Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol. Cell. Proteomics.* 10 (10), O111.009381. doi:10.1074/mcp.O111.009381

Hu, H., Zhao, W., Zhu, M., Zhao, L., Zhai, L., Xu, J.-y., et al. (2019). LysargiNase and chemical derivatization based strategy for facilitating in-depth profiling of C-terminome. *Anal. Chem.* 91 (22), 14522–14529. doi:10.1021/acs.analchem.9b03543

Huang, D. W., Sherman, B. T., Tan, Q., Collins, J. R., Alvord, W. G., Roayaei, J., et al. (2007). The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8 (9), R183. doi:10.1186/gb-2007-8-9-r183

Huang, L.-T., and Gromiha, M. M. (2010). First insight into the prediction of protein folding rate change upon point mutation. *Bioinformatics* 26 (17), 2121–2127. doi:10.1093/bioinformatics/btq350

Huesgen, P. F., and Overall, C. M. (2012). N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. *Physiol. Plant.* 145 (1), 5–17. doi:10.1111/j.1399-3054.2011.01536.x

Kahl, L., Molloy, J., Patron, N., Matthewman, C., Haseloff, J., Grewal, D., et al. (2018). Opening options for material transfer. *Nat. Biotechnol.* 36 (10), 923–927. doi:10.1038/nbt.4263

Kaleja, P., Helbig, A. O., and Tholey, A. (2019). Combination of SCX fractionation and charge-reversal derivatization facilitates the identification of nontryptic peptides in C-terminomics. *J. Proteome Res.* 18 (7), 2954–2964. doi:10.1021/acs.jproteome.9b00264

Kaneti, J., Kirby, A. J., Koedjikov, A. H., and Pojarlieff, I. G. (2004). Thorpe–Ingold effects in cyclizations to five-membered and six-membered rings containing planar segments. the rearrangement of N(1)-alkyl-substituted dihydroorotic acids to hydantoinacetic acids in base. *Org. Biomol. Chem.* 2 (7), 1098–1103. doi:10.1039/b400248b

Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., et al. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* 28 (3), 281–288. doi:10.1038/nbt.1611

Kleifeld, O., Doucet, A., Prudova, A., Gioia, M., Kizhakkedathu, J. N., et al. (2011). Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* 6 (10), 1578–1611. doi:10.1038/nprot.2011.382

Klein, T., Eckhard, U., Dufour, A., Solis, N., and Overall, C. M. (2018). Proteolytic cleavage-mechanisms, function, and "omic" approaches for a near-ubiquitous posttranslational modification. *Chem. Rev.* 118 (3), 1137–1168. doi:10.1021/acs.chemrev.7b00120

Klein, T., Fung, S.-Y., Renner, F., Blank, M. A., Dufour, A., Kang, S., et al. (2015). The paracaspase MALT1 cleaves HOIL1 reducing linear ubiquitination by LUBAC to dampen lymphocyte NF-κB signalling. *Nat. Commun.* 6, 8777. doi:10.1038/ncomms9777

Koudelka, T., Winkels, K., Kaleja, P., and Tholey, A. (2021). Shedding light on both ends: an update on analytical approaches for N- and C-terminomics. *Biochim. Biophys. Acta. Mol. Cell Res.* 1869 (1), 119137. doi:10.1016/j.bbamcr.2021.119137

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157 (1), 105–132. doi:10.1016/0022-2836(82)90515-0

Leng, J., Wang, H., Zhang, L., Zhang, J., Wang, H., and Guo, Y. (2013). A highly sensitive isotope-coded derivatization method and its application for the mass spectrometric analysis of analytes containing the carboxyl group. *Anal. Chim. Acta* 758, 114–121. doi:10.1016/j.aca.2012.11.008

Li, B., Guo, F., Hu, H., Liu, P., Tan, M., Pan, J., et al. (2020). The characterization of column heating effect in nanoflow liquid chromatography mass spectrometry (nanoLC-MS)-based proteomics. *J. Mass Spectrom.* 55 (1), e4441. doi:10.1002/jms.4441

Li, Q., Zhang, Y., Huang, J., Wu, Z., Tang, L., Huang, L., et al. (2020). Basic strong cation exchange chromatography, BaSCX, a highly efficient approach for C-terminomic studies using LysargiNase digestion. *Anal. Chem.* 92 (7), 4742–4748. doi:10.1021/acs.analchem.9b05280

Lin, H.-C., Yeh, C.-W., Chen, Y.-F., Lee, T.-T., Hsieh, P.-Y., Rusnac, D. V., et al. (2018). C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol. Cell* 70 (4), 602–613.e3. doi:10.1016/j.molcel.2018.04.006

MacDonald, J. I., Munch, H. K., Moore, T., and Francis, M. B. (2015). One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* 11 (5), 326–331. doi:10.1038/nchembio.1792

Marino, G., Eckhard, U., and Overall, C. M. (2015). Protein termini and their modifications revealed by positional proteomics. *ACS Chem. Biol.* 10 (8), 1754–1764. doi:10.1021/acschembio.5b00189

Perrar, A., Dissmeyer, N., and Huesgen, P. F. (2019). New beginnings and new ends: methods for large-scale characterization of protein termini and their use in plant biology. *J. Exp. Bot.* 70 (7), 2021–2038. doi:10.1093/jxb/erz104

Rogers, L. D., and Overall, C. M. (2013). Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics.* 12 (12), 3532–3542. doi:10.1074/mcp.M113.031310

Schilling, O., Barré, O., Huesgen, P. F., and Overall, C. M. (2010). Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods* 7 (7), 508–511. doi:10.1038/nmeth.1467

Van Damme, P., Staes, A., Bronsoms, S., Helsens, K., Colaert, N., Timmerman, E., et al. (2010). Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat. Methods* 7 (7), 512–515. doi:10.1038/nmeth.1469

Wang, G., Li, S., Gilbert, J., Gritton, H. J., Wang, Z., Li, Z., et al. (2017). Crucial roles for SIRT2 and AMPA receptor acetylation in synaptic plasticity and memory. *Cell Rep.* 20 (6), 1335–1347. doi:10.1016/j.celrep.2017.07.030

Wang, Z., Zhang, L., Yuan, W., Zhang, Y., and Lu, H. (2021). SAPT, a fast and efficient approach for simultaneous profiling of protein N- and C-terminome. *Anal. Chem.* 93 (30), 10553–10560. doi:10.1021/acs.analchem.1c01598

Winter, N., Novatchkova, M., and Bachmair, A. (2021). Cellular control of protein turnover via the modification of the amino terminus. *Int. J. Mol. Sci.* 22 (7), 3545. doi:10.3390/ijms22073545

Zhang, Y., He, Q., Ye, J., Li, Y., Huang, L., Li, Q., et al. (2015). Systematic optimization of C-terminal amine-based isotope labeling of substrates approach for deep screening of C-terminome. *Anal. Chem.* 87 (20), 10354–10361. doi:10.1021/acs.analchem.5b02451

Zhang, Y., Li, Q., Huang, J., Wu, Z., Huang, J., Huang, L., et al. (2018). An approach to incorporate multi-enzyme digestion into C-tails for C-terminomics studies. *PROTEOMICS* 18 (1), 1700034. doi:10.1002/pmic.201700034

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fgene.2022.989459

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean University, China

REVIEWED BY
Jing Ji,
Jiangsu Ocean University, China
Linhui Zhai,
Shanghai Institute of Materia Medica (CAS), China

*CORRESPONDENCE
Changmin Wei,
wchangmin@163.com
Hao Chang,
changhao@hanyu-biomed.org,

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

RECEIVED 08 July 2022
ACCEPTED 25 July 2022
PUBLISHED 08 September 2022

CITATION
Xu C, Sun D, Wei C and Chang H (2022), Bioinformatic analysis and experimental validation identified DNA methylation–Related biomarkers and immune-cell infiltration of atherosclerosis.
Front. Genet. 13:989459.
doi: 10.3389/fgene.2022.989459

Bioinformatic analysis and experimental validation identified DNA methylation–Related biomarkers and immune-cell infiltration of atherosclerosis

Congjian Xu[1], Di Sun[1], Changmin Wei[1]* and Hao Chang [2]*

[1]Department of Cardiology, Shengli Oilfield Central Hospital, Dongying, Shandong, China, [2]Hanyu Biomed Center Beijing, Beijing, China

Background: DNA methylation is an important form of epigenetic regulation and is closely related to atherosclerosis (AS). The purpose of this study was to identify DNA methylation–related biomarkers and explore the immune-infiltrate characteristics of AS based on methylation data.

Methods: DNA methylation data of 15 atherosclerotic and paired healthy tissues were obtained from Gene Expression Omnibus database. Differential methylation positions (DMPs) and differential methylation regions (DMRs) were screened by the ChAMP R package. The methylation levels of DMPs located on CpG islands of gene promoter regions were averaged. The limma R package was used to screen differentially methylated genes in the CpG islands of the promoter regions. The diagnostic values of the methylation levels were evaluated using the pROC R package. The EpiDISH algorithm was applied to quantify the infiltration levels of seven types of immune cells. Subsequently, three pairs of clinical specimens of coronary atherosclerosis with Stary's pathological stage III were collected, and the methylation levels were detected by the methylation-specific PCR (MS-PCR) assay. Western blot was performed to detect the protein expression levels of monocyte markers.

Results: A total of 110, 695 DMPs, and 918 DMRs were screened in the whole genome. Also, six genes with significant methylation differences in the CpG islands of the promoter regions were identified, including 49 DMPs. In total, three genes (GRIK2, HOXA2, and HOXA3) had delta beta greater than 0.2. The infiltration level of monocytes was significantly upregulated in AS tissues. MS-PCR assay confirmed the methylation status of the aforementioned three genes in AS samples. The Western blot results showed that the expression levels of the monocyte marker CD14 and M1-type macrophage marker CD86 were significantly increased in AS while M2-type macrophage marker protein CD206 was significantly decreased.

Conclusion: This study identified potential DNA methylation–related biomarkers and revealed the role of monocytes in early AS.

KEYWORDS
DNA methylation, diagnosis, MS-PCR, immune-cell infiltration, atherosclerosis

Introduction

Atherosclerosis (AS) is a chronic disease with extensive studies (Libby et al., 2019). It is no doubt that AS is continuously developing worldwide, and there are emerging challenges. For example, AS has been found in the younger population and females. In addition, new risk factors have been found besides diets, such as physical inactivity, microbiome, and epigenetics (Xu et al., 2018; Libby, 2021). All of which guide the new strategies for the prevention, diagnosis, and treatment of AS.

Chronic AS may lead to acute cardiovascular events (Ahmadi et al., 2019). The timely diagnosis of AS is critical. Until now, the monitoring of AS plaque development and rupture has been still most effective, since it causes most cardiovascular diseases in clinical practice (Rader and Daugherty, 2008; Mushenkova et al., 2020). However, instead of plaque monitoring, the early diagnosis of AS has been more important in preventing the disease and subsequent outcomes (Herrington et al., 2016). Inflammation and immunity-related biomarkers have been reported to link with traditional and emerging risk factors, including pro-inflammatory cytokines, inflammatory signaling pathways factors, bioactive lipids, and adhesion molecules (Zhu et al., 2018; Libby, 2021). Recently, more evidence suggests that AS is an epigenetic disease and its development involves several epigenetic processes, including DNA methylation, histone modification, and non-coding RNAs (Xu et al., 2018).

As a chronic disease, the pathogenesis of AS can be classified into three processes (Hai and Zuo, 2016; Tabaei and Tabaee, 2019). First, the activation of endothelial cells, infiltration of monocytes and formation of foam cells. Second, the stimulation of smooth muscle cells (SMCs), followed by differentiation, migration, and phenotypic switching of immune cells, thus forming the plaque. Third, the rupture of AS plaque and the occurrence of thrombosis. Several mechanisms and regulatory pathways participate in the long and complex processes during AS development, which may involve various epigenetic regulations (Khyzha et al., 2017; Tang et al., 2021).

DNA methylation is the epigenetic event that covalently transfers a methyl group to the cytosine, mainly on the CpG dinucleotide site and CpG islands, regulating the expression of specific target genes at the transcriptional level (Moore L. D. et al., 2013). Genomic DNA methylations have been proved to play a crucial part in the early progression of AS (Dong et al., 2002; Zhang et al., 2021). For example, DNA methylation can regulate or partially regulate different AS-related genes, such as ERα/β, MMP9/2/7, EC-SOD, INF-γ, and eNOS. These genes were targets of inflammatory response and reaction, macrophages, apoptosis, cell proliferation, and differentiation (Hai and Zuo, 2016).

CpG islands are GC-rich and primarily located at the 5′ regulatory regions of all housekeeping genes (Deaton and Bird, 2011). Aberrant DNA methylation in CpG islands would upregulate AS-susceptible genes and downregulate AS-protective genes, contributing to the AS progression (Dong et al., 2002). Varied DNA methylation aberrations have been reported during the development of AS. The exploration of this information may be significant for better understanding AS pathogenesis. Furthermore, some epigenetic drugs with therapeutic potential may be screened (Xu et al., 2018; Zhang et al., 2021).

In this study, bioinformatic analysis has been performed on the DNA methylation data of AS patients. Differential methylation positions (DMPs) and differential methylation regions (DMRs) were systematically screened. Then differentially methylated genes in the CpG islands of the promoter regions were identified. Furthermore, the infiltration levels of different immune cells were quantified for analyzing the immune status of AS. Finally, the DNA methylation of target genes and variations of immune cells have been confirmed experimentally in samples of AS patients.

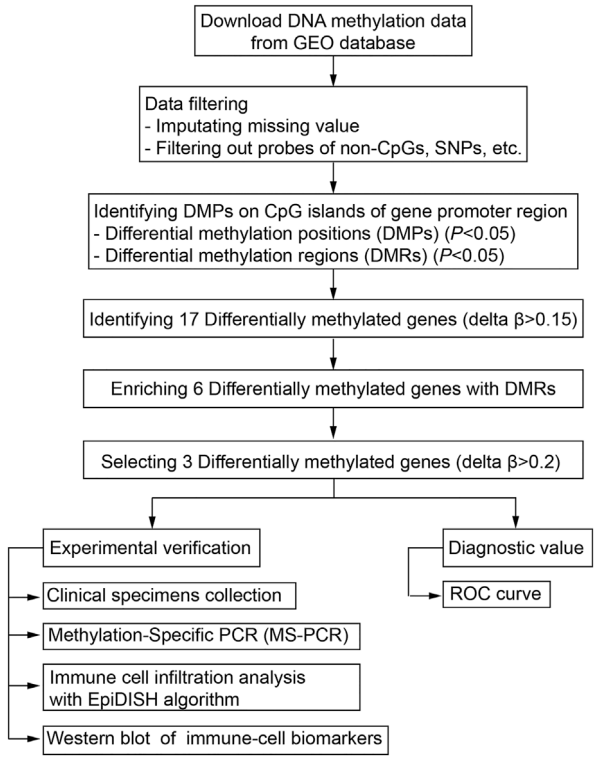


FIGURE 1
Workflow of the bioinformatic analysis for screening aberrant DNA methylation sites.

Materials and methods

Methylome dataset

The DNA methylation dataset GSE46394 including 15 atherosclerotic and paired healthy tissues was obtained from the Gene Expression Omnibus (GEO) database (https://ncbi.nlm.nih.gov/gds). The methylation levels were evaluated based on the Illumina Infinium Human Methylation 450 Beadchip platform and quantified as a β-value. The mean values of methylated (M) and non-methylated (U) signal intensity for each tissue and CpG sites (CpGs) were calculated using the formula [$\beta = M/(M + U)$]. The workflow for screening aberrant DNA methylation has been presented (Figure 1).

DNA methylation differential analyses

After the imputation of missing values with the function of impute.knn in the impute package of R, ChAMP (Tian et al., 2017) was used for the follow-up analyses. First, probes that belong to non-CpGs, including SNPs (Zhou et al., 2017), align to multiple locations, and located on X are Y chromosomes were filtered out with the champ.filter function. Then the filtered data were normalized with the BMIQ method for types I and II probe correction by the champ.norm function. Differential methylation positions (DMPs) with a BH adjusted *p*-value below 0.05 were screened using the champ.DMP function. Differential methylation regions (DMRs) were identified using the Bumphunter method with a *p*-value below 0.05 by using the champ.DMR function. The org.Hs.eg.db R package was then used to annotate the genes corresponding to these DMRs with UCSC.hg19 as the reference genome file. The function enrichments were performed with Metascape (www.metascape.org).

Methylation analysis of CpG islands in gene promoter regions

The β-value was defined as the value of the methylation expression profiles, with range between 0 (no methylation) and 1 (complete methylation). The average β-value of DMPs located on CpG islands of the gene promoter region was calculated. The limma R package (Ritchie et al., 2015) was used to screen differentially methylated genes in the CpG islands of the promoter regions. The diagnostic values of the methylation levels were evaluated with receiver operating characteristic (ROC) established using the pROC package in R (Robin et al., 2011).

Clinical specimens and ethical statement

A total of three pairs of clinical specimens of coronary atherosclerosis with Stary's pathological stage III were collected from the Central Hospital of Shengli Oil Field (Yutani et al., 1999). All patients signed the informed consent form, and the study was approved by the Scientific Research Ethics Committee of the Central Hospital of Shengli Oil Field.

Methylation-specific PCR

MS-PCR based on bisulfite conversion was conducted. Genomic DNA from three pairs of clinical specimens of coronary atherosclerosis with Stary's pathological stage III was isolated using the DNA extraction kit (Beyotime, China), and the premium bisulfite kit (Diagenode, Belgium) was applied for sodium bisulfite treatment of the genomic DNA, according to the manufacturer's protocol. The MS-PCR primers used in this study were designed using the Methyl Primer Express v1.0 and listed in Table 1.

Immune-cell infiltration analysis

Based on the methylation β-value of CpGs, the EpiDISH R package (Zheng et al., 2019) was applied to quantify the infiltration levels of seven types of immune cells, including B cells, NK cells, $CD4^+$ T cells, $CD8^+$ T cells, monocytes, neutrophils, and eosinophils. The difference analysis was performed using the Wilcoxon test.

Western blot

The expression levels of proteins were determined by Western blotting. The tissues were homogenized with RIPA lysis buffer (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% deoxycholic acid, and 0.1% SDS) for extracting proteins, and the proteins were separated using SDS-PAGE and transferred to PVDF membrane (Millipore, Germany). Then after blocking with 5% skim milk, the membrane was incubated with primary antibodies for overnight at 4°C. The applied primary antibodies were as follows: CD14 (1:1,000, ab106285, Abcam), CD86 (1:1,000, ab220188, Abcam), and CD206 (1:1,000, K006619P, Solarbio). The HRP-labeled secondary antibody was incubated for 2 h at room temperature. The ECL Western blotting detection system (Tanon, China) and ImageJ software were applied to visualize and analyze the results.

Statistics

Methylation data were analyzed in R 4.1.2 software, and WB data were analyzed using Prism 9.0 software. Student's *t*-test was used for two groups. $p < 0.05$ was indicated to be statistically significant.

TABLE 1 Sequences of MSP primers.

Gene	Primer sequence	
GRIK2	Methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	TCGCGTTTTTTTTTTTTC	ACTAATAATCCTCACACGCG
	No methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	TTATTGTGTTTTTTTTTTTTT	CTAACTAATAATCCTCACACACA
HOXA2	Methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	TATTTTTTTGGTTGGTCGTC	AAACGACTCTCGAAACTTCC
	No methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	GGTATTTTTTTGGTTGGTTGTT	AAAAACAACTCTCAAAACTTCC
HOXA3	Methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	GGATTAGACGTTGTTTCGC	CCCGAAAATAAACGCTAAT
	No methylation	
	Forward (5′–3′)	Reverse (5′–3′)
	GGATTAGATGTTGTTTTGT	CCCAAAAATAAACACTAAT

Results

Differential methylation positions

After the imputation of missing values, probes that belong to non-CpGs (n = 3,156), including SNPs (n = 59,901), align to multiple locations (n = 11), and located on X and Y chromosomes (n = 10,028) were filtered out using the champ.filter function. The raw data of 412,481 probes were density plotted (Figure 2A). Then the data were normalized using the BMIQ method for types I and II probe correction (Teschendorff et al., 2013). The principal component analysis (PCA) of the normalized data in the atherosclerotic and healthy groups was performed (Figure 2B). A total of 110695 significant DMPs with a BH adjusted p-value below 0.05 were identified using the champ.DMP function. Heat maps of DMPs with β-value variance in the top 1,000 were drawn using the pheatmap R package (Figure 2C). Most DMPs were hypermethylated.

Differential methylation regions

A total of 918 differential methylation regions (DMRs) were identified using the Bumphunter method with a p-value below 0.05 by using the champ.DMR function (Supplementary File S1). We use UpSetR package to plot the upset diagram of these DMRs distribution regions. Most DMRs are distributed in promoter, genic, exon, and 5′ UTR regions (Figure 3A). These DMRs were annotated to 854 genes, and a functional enrichment analysis revealed that the functions of these genes are mainly concentrated in skeletal system development, muscle structure development, cell morphogenesis, heart development, extracellular cell matrix organization, tube morphogenesis, and so on (Figure 3B).

Methylation analysis of CpG islands in gene promoter regions

Since the existing low-throughput methylation detection techniques, such as MS-PCR and bisulfite sequencing PCR (BS-PCR), and other classical methods are designed based on the CpG islands of gene promoter regions (Li, 2007); we further estimated the methylation levels of gene promoter regions to select markers with a practical application value. The average β-value of DMPs located on CpG islands of the gene promoter region was calculated. In total, 17 genes were identified with delta β over 0.15, and the heat map was shown (Figure 3C). The details of differential methylation data are shown in Table 2. In addition, six of these genes (HOXA2, GRIK2, HOXA3, TBC1D16, MYL9, and HNF1A) were enriched with DMRs (Figure 3D). Next, we selected three genes (HOXA2, GRIK2, and HOXA3) with delta β over 0.20 to conduct experimental verification.

Exploration of the diagnostic biomarkers in methylation levels with methylation-specific PCR

The diagnostic values of the identified three genes were evaluated with the receiver operating characteristic (ROC)

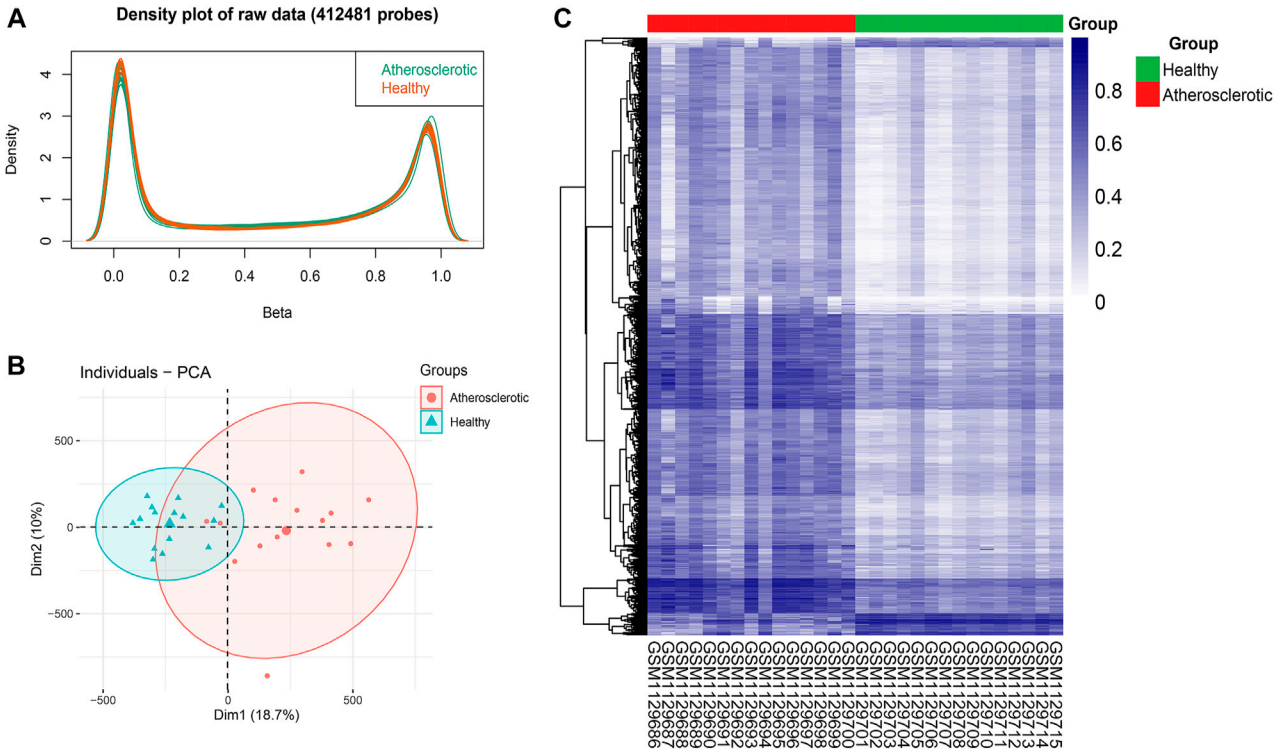


FIGURE 2
Identification of differential methylation positions (DMPs) in atherosclerosis. **(A)** Density plot of 412,481 probes' raw data. **(B)** Principal component analysis (PCA) of the normalized data in the atherosclerotic and healthy groups. **(C)** Heat map of DMPs with β-value variance in the top 1,000.

curves. The area under curve (AUC) values were 0.822 for GRIK2, 0.978 for HOXA2, and 1.0 for HOXA3 separately (Figure 4A). The β-value and distribution characteristics of CpG sites within the DMRs located in these genes were plotted (Figure 4B). The characteristics of CpG sites located on the islands of gene promoter regions are shown in Table 3. The MS-PCR results indicated that the promoters' methylation statuses of all the three genes were almost un-methylated in the healthy group while partially methylated in the atherosclerotic group (Figure 5). These data confirmed the differences in the promoter regions and their potential diagnostic values.

The infiltration of monocytes and M1-type macrophages were significantly increased in atherosclerotic tissues at the pre-atheroma stage

The EpiDISH algorithm was used to assess immune-cell infiltration (Supplementary File S2). The heat map and violin plots are shown in Figures 6A,B separately. The results of immune-cell infiltration analysis showed that the infiltration degree of monocytes was significantly upregulated in AS tissues, whereas the infiltration level of NK cells and CD4T cells showed a downward trend, but there were no statistical differences (Figure 6B). It has been reported that macrophage polarization plays a key role in the progression of atherosclerosis, so we tested the expression of marker proteins (CD14, CD86, and CD206) by Western blot. The expression levels of monocyte marker CD14 and M1-type macrophage marker CD86 were significantly increased in AS while that of M2-type macrophage marker protein CD206 was significantly decreased (Figures 7A,B). The CD86/CD206 ratios were also significantly increased in AS (Figure 7B).

Discussion

In this study, based on the DNA methylation data of AS patients downloaded from the database, a bioinformatic analysis has been performed for demonstrating the aberrant DNA methylations. A total of three differentially methylated genes (GRIK2, HOXA2, and HOXA3) in the CpG islands of the promoter regions have been screened. These genes can be applied as biomarkers for the early diagnosis of AS. The MS-PCR data proved that all the three genes were almost un-

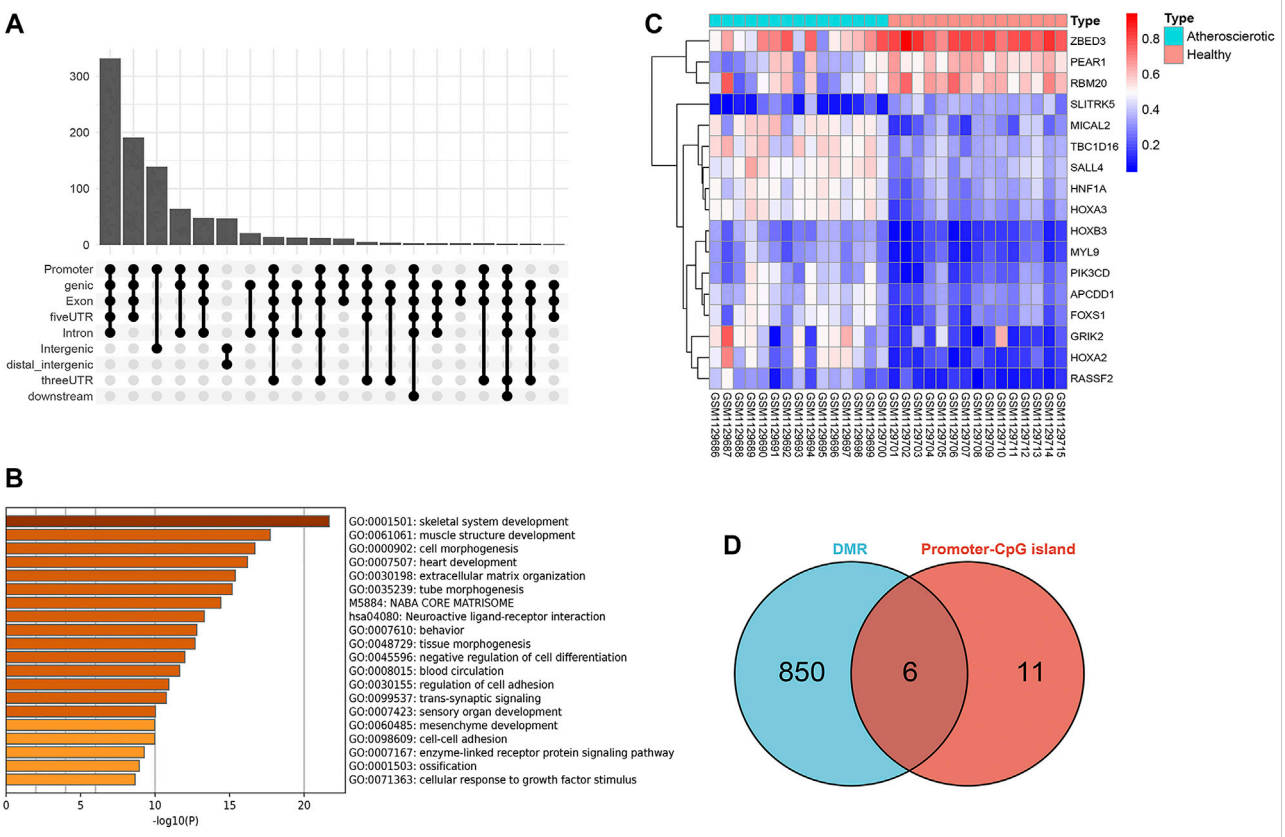


FIGURE 3
Screening of gene promoter regions enriched with differential methylation regions (DMRs). **(A)** Upset diagram of DMRs' distribution. **(B)** Functional enrichment analysis with Metascape. **(C)** Heat map of the 17 genes identified with delta β over 0.15. **(D)** Venn diagram identified six genes which were enriched with DMRs.

methylated in the healthy group while partially methylated in AS group. The immune-cell infiltration in AS patients was also investigated. The infiltration level of monocytes was significantly upregulated in AS tissues. The Western blot data indicated similar results; the expression levels of monocyte marker CD14 and M1-type macrophage marker CD86 were significantly increased while that of M2-type macrophage marker CD206 was significantly decreased, with significantly increased CD86/CD206 ratios. Both the diagnosis biomarkers and potential immune cell–relevant mechanisms were theoretically and experimentally revealed.

Recent studies have found that DNA methylations were related to the whole development process of AS (Borghini et al., 2013; Zaina, 2014). The potential roles of DNA methylation in AS have been reported to link with oxidative stress (Kalea et al., 2018), inflammation, SMCs, and homocysteine (Hcy) (Ma et al., 2017; Tabaei and Tabaee, 2019). Numerous studies have shown that oxidative stress affected DNA methylation during AS. At the presence of oxidative stress such as H_2O_2 and reactive oxygen species, DNA methyltransferase1 (DNMT1) may be relocalized from non-GC-rich to GC-rich areas. The uprising methylation level in CpG island may promote AS development (O'Hagan et al., 2011). Inflammation has been found in all stages of AS; thus, inflammatory molecules have been identified as markers of the AS progression (Zhu et al., 2018; Bäck et al., 2019). DNMT may regulate the expression of inflammatory factors by varying their DNA methylation, thus accelerating the progression of AS (Zhu et al., 2018). A bioinformatic study reported an imprinted gene PLA2G7, which encoded lipoprotein-associated phospholipase A2 (Lp-PLA2). The hypomethylation of PLA2G7 increased expression upon inflammation in AS, indicating the effects of DNA methylation modification on atheroprogression and destabilization *via* inflammatory processes (Li et al., 2021). SMCs migrate from the media to the intima and form the AS plaques. The DNA methylation of SMCs regulated their differentiation, migration, and phenotypic switching, thus influencing plaque formation (Tabaei and Tabaee, 2019). Furthermore, Hcy has been considered an independent risk factor for AS. Hcy can promote global DNA hypomethylation,

TABLE 2 Characteristics of genes with significantly difference methylation levels in promoters.

Gene	AS mean	Healthy mean	Delta β	CpGs hits	adj *p* value
HOXA2	0.429	0.156	0.273	7	2.43E-05
GRIK2	0.435	0.216	0.219	10	2.49E-03
MICAL2	0.495	0.278	0.217	2	3.24E-05
HOXA3	0.498	0.295	0.203	19	7.86E-07
RASSF2	0.307	0.124	0.182	1	8.51E-05
FOXS1	0.423	0.241	0.182	2	4.51E-05
TBC1D16	0.519	0.348	0.171	3	8.87E-06
HOXB3	0.305	0.147	0.159	7	6.86E-06
MYL9	0.323	0.165	0.158	3	2.43E-05
HNF1A	0.480	0.323	0.157	7	8.29E-06
PIK3CD	0.375	0.220	0.155	2	5.45E-04
SALL4	0.513	0.361	0.153	1	6.63E-06
APCDD1	0.422	0.272	0.150	1	2.61E-05
PEAR1	0.414	0.581	−0.167	1	4.24E-04
RBM20	0.450	0.623	−0.173	2	2.07E-03
SLITRK5	0.156	0.347	−0.191	3	4.51E-05
ZBED3	0.595	0.795	−0.200	1	4.11E-04

which was reported to participate in endothelial dysfunction and SMC proliferation induced by DNA methylations (Ehrlich, 2019). However, another study performed on AS patients reported that the upregulated level of Hcy was positively correlated with more hypermethylation of CpG islands in the ER-α gene promoter region, as well as the severity of AS lesion. It has been further verified with *in vitro* experiments (Huang et al., 2009).

Since the DNA methylation and CpG islands methylation/demethylation matters in AS progress, several studies have tried to explore more potential DNA methylation sites and regions. There are generally three DNA methylation approaches for finding DNA methylation aberrations: candidate gene, global methylation, and epigenome-wide association studies (Fernández-Sanlés et al., 2017). However, there is still challenging to find critical DNA methylation aberrations effectively. The approach based on candidate gene seems to be low in throughput. For improving the throughput, the genome-wide DNA methylation aberrations have been obtained based on the methylation microarray chip. More than 400,000 methylated CpGs could be obtained with a typical Illumina Human Methylation 450 BeadChip. Even after setting criteria involving absolute β-value, there were still 1,458 differently methylated CpGs covering 971 genes were extracted (Chen et al., 2020). The high-throughput bioinformatic analysis seems to be more significant for pathway and mechanism analysis. In addition, in most of the recent studies involving global methylation, the DNA methylation has been explored based on the promoter regions of target genes while there are still abundant but unrevealed CpGs located on the non-promoter region. Thus, some researchers have tried to conduct synthetic analysis to improve the efficiency of screening significant candidate genes and corresponding DNA methylation. For example, in a study on DNA methylation and coronary heart disease including AS, a total of 51 individual articles were comprehensively analyzed (Fernández-Sanlés et al., 2017). Consistent results reported in at least two articles were identified. In addition, the probability of finding the same gene or CpG in two studies was calculated. With this strategy, some candidate genes were revealed, including hypermethylation in ESRα, ABCG1 and FOXP3, and hypomethylation in IL-6. These target genes were associated with several diseases and functions, including inflammatory, metabolic, and cardiovascular diseases (Zhu et al., 2018). In our study, the methylation levels of gene promoter regions were innovatively introduced for selecting markers with practical application value. The average β-value of DMPs located on CpG islands of the gene promoter region was calculated. A total of 17 genes identified with delta β over 0.15 and 6 were enriched. In total, three genes (HOXA2, GRIK2, and HOXA3) with delta β over 0.20 were further selected to conduct experimental verification. Since existing strategies for screening target genes and corresponding DNA methylation aberration have been still limited, more studies can be performed for providing new and effective theoretical or experimental tools.

Some studies have been performed to investigate the three genes (HOXA2, GRIK2, and HOXA3) identified as biomarker genes for AS. HOXA2 and HOXA3 were both homeobox genes,

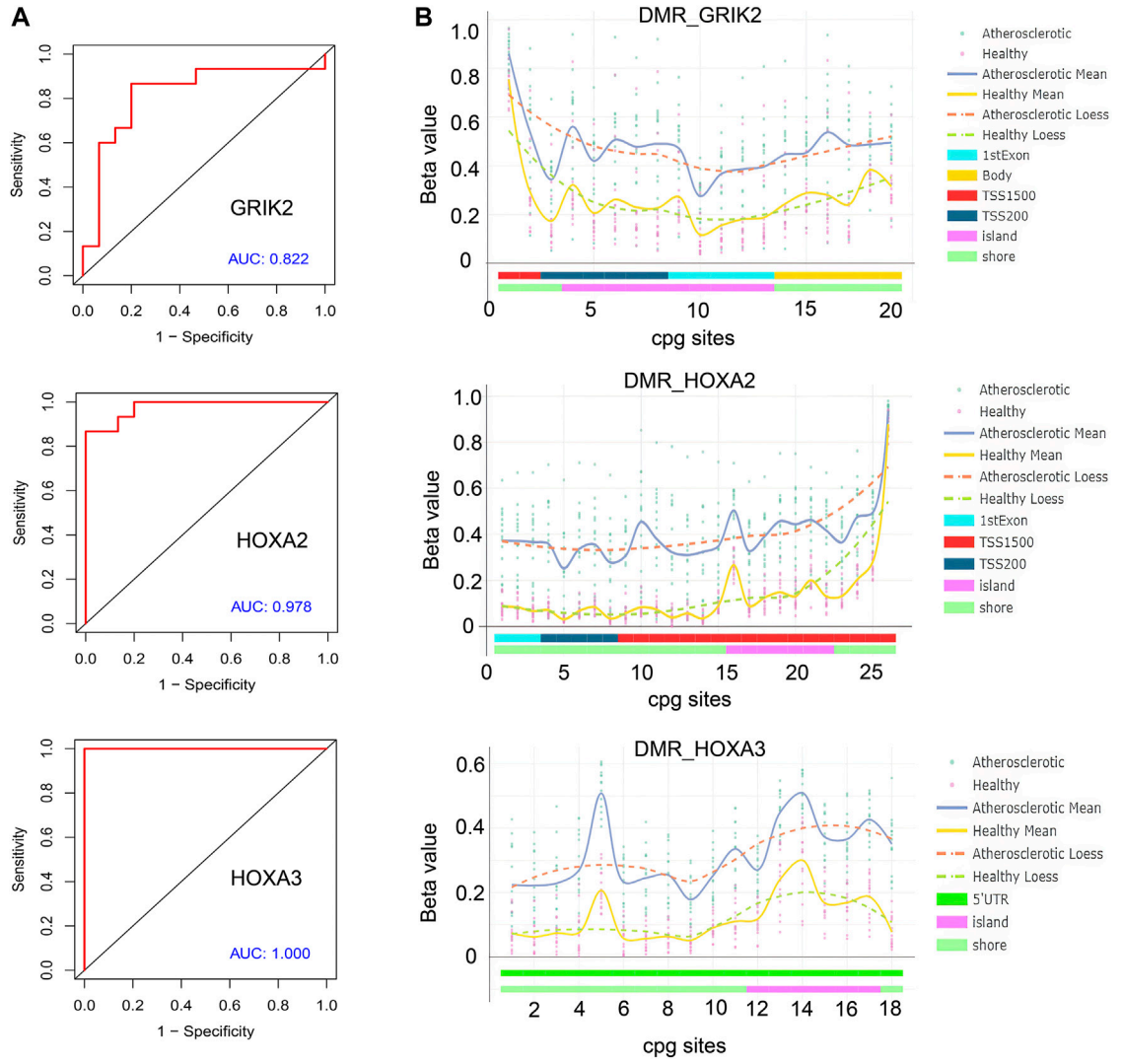


FIGURE 4
Bioinformatic analysis of the identified genes' diagnostic values. **(A)** Receiver operating characteristic (ROC) curves of GRIK2, HOXA2, and HOXA3. **(B)** β-value and distribution characteristics of CpG sites within the DMRs located in GRIK2, HOXA2, and HOXA3.

which were associated with the regulation of normal differentiation and development of cells. The homeobox genes have been reported to link with various cancers. Especially, hypermethylated CpG islands for HOXA gene promoters were associated with pathways in cancers. In colorectal cancer, the percentage of methylation of three HOXA genes (HOXA5, HOXA2, and HOXA6) were up to 67.62%, 58.36%, and 31.32%, respectively. The results demonstrated that colorectal cancer tissues and cells had a stronger methylation status around these three HOXA gene promoter regions, compared with adjacent controls. The epigenetic silencing of these three HOXA genes may be an important event in the progression of colorectal cancer (Li et al., 2019). The low-methylation epigenotype of HOXA2 and HOXA9 in squamous cell carcinoma was associated with idiopathic pulmonary fibrosis and poorer prognosis (Hata et al., 2020). In another study on prostate cancer, HOXA2, HOXA9, and HOXA10 were identified as critical genes, which were both abnormally expressed and associated with clinical outcomes of patients with prostate cancer (Song et al., 2022). In addition to cancers, the CpG methylation of HOXA2 was related to severe fibrosis and the progression in hepatitis B-related chronic liver disease (Zeybel et al., 2016). An early Russia study has also reported that HOXA2 exhibited most pronounced difference in the methylation level for it's CpG sites. HOXA2 was found to be hypomethylated in the carotid atherosclerotic plaques compared to its methylation patterns in normal control veins (Nazarenko et al., 2013). Until now, few studies have reported the association between GRIK2 and

TABLE 3 Characteristics of CpG sites located on the islands of gene promoter regions.

Gene	CpG sites	AS	Healthy	Delta β	adj *p* value	Feat cgi	UCSC CpG islands name
GRIK2	cg18193094	0.275	0.117	0.158	5.31E-04	1stExon island	chr6:101846766-101847135
	cg10591607	0.368	0.155	0.212	1.55E-03	1stExon island	chr6:101846766-101847135
	cg22541254	0.420	0.206	0.214	3.48E-03	TSS200 island	chr6:101846766-101847135
	cg24301620	0.473	0.273	0.199	5.16E-03	1stExon island	chr6:101846766-101847135
	cg13080565	0.478	0.230	0.248	5.38E-03	TSS200 island	chr6:101846766-101847135
	cg26316946	0.387	0.181	0.206	8.24E-03	1stExon island	chr6:101846766-101847135
	cg06247406	0.507	0.262	0.245	9.88E-03	TSS200 island	chr6:101846766-101847135
	cg05942459	0.491	0.227	0.264	1.12E-02	TSS200 island	chr6:101846766-101847135
	cg21635870	0.396	0.188	0.208	1.24E-02	1stExon island	chr6:101846766-101847135
	cg24753760	0.561	0.322	0.239	1.27E-02	TSS200 island	chr6:101846766-101847135
HOXA2	cg06166490	0.458	0.149	0.309	6.94E-07	TSS1500 island	chr7:27143181-27143479
	cg19432993	0.391	0.126	0.266	6.99E-07	TSS1500 island	chr7:27143181-27143479
	cg01217984	0.504	0.267	0.236	2.05E-06	TSS1500 island	chr7:27143181-27143479
	cg00445443	0.417	0.131	0.286	3.78E-06	TSS1500 island	chr7:27143181-27143479
	cg04027736	0.463	0.201	0.261	4.15E-06	TSS1500 island	chr7:27143181-27143479
	cg10319053	0.444	0.131	0.313	5.60E-06	TSS1500 island	chr7:27143181-27143479
	cg02225599	0.327	0.089	0.239	2.19E-05	TSS1500 island	chr7:27143181-27143479
HOXA3	cg09591524	0.652	0.384	0.268	2.87E-08	5′UTR island	chr7:27150030-27150418
	cg02439266	0.612	0.344	0.268	6.02E-08	5′UTR island	chr7:27150030-27150418
	cg26297005	0.562	0.246	0.316	7.07E-08	5′UTR island	chr7:27162087-27162426
	cg19999161	0.427	0.187	0.240	9.75E-08	5′UTR island	chr7:27154999-27155426
	cg18430152	0.713	0.349	0.365	9.75E-08	5′UTR island	chr7:27162087-27162426
	cg04778178	0.713	0.363	0.350	1.18E-07	5′UTR island	chr7:27162087-27162426
	cg04351734	0.373	0.168	0.205	1.04E-06	5′UTR island	chr7:27154999-27155426
	cg16748008	0.270	0.118	0.153	1.72E-06	5′UTR island	chr7:27154999-27155426
	cg22798849	0.366	0.169	0.197	1.84E-06	5′UTR island	chr7:27154999-27155426
	cg16406967	0.449	0.241	0.208	2.78E-06	5′UTR island	chr7:27154999-27155426
	cg18680977	0.510	0.301	0.209	9.42E-06	5′UTR island	chr7:27154999-27155426
	cg07522913	0.351	0.220	0.132	2.72E-05	5′UTR island	chr7:27150030-27150418
	cg01301319	0.451	0.308	0.143	7.76E-05	5′UTR island	chr7:27153187-27153647
	cg13172549	0.446	0.252	0.193	8.64E-05	5′UTR island	chr7:27153187-27153647
	cg22962123	0.493	0.335	0.158	1.54E-04	5′UTR island	chr7:27153187-27153647
	cg09144964	0.398	0.286	0.112	5.72E-04	5′UTR island	chr7:27150030-27150418
	cg05851442	0.446	0.347	0.099	1.39E-03	5′UTR island	chr7:27153187-27153647
	cg24360871	0.733	0.579	0.154	1.44E-03	5′UTR island	chr7:27163819-27164098
	cg03536885	0.503	0.411	0.092	8.60E-03	5′UTR island	chr7:27163819-27164098

atherosclerosis. GRIK2 gene encoded proteins belong to the kainate family of glutamate receptors. GRIK2 was involved in various normal neurophysiologic processes. The aberrant expression of GRIK2 has been reported to be involved in regions, functional genes, biological function, and pathways that mediate depression disorder. GRIK2 also showed abnormal methylation pattern specific to astrocytic dysfunction associated with depressive psychopathology (Nagy et al., 2015; Wang W. et al., 2021). It is interesting to explore the potential association of AS with the nervous system.

Immune infiltration is closely related to the progression and prognosis of AS (Wang L. et al., 2021; Tan et al., 2021). Several studies have tried to analyze the immune-cell infiltration profiles at different AS progresses. A review has suggested the importance of metabolic and functional reprogramming in monocytes and macrophages for AS progression. Furthermore, macrophages and monocyte contributed to pro- or anti-inflammatory mechanisms (Groh et al., 2018; Kim et al., 2020). Monocytes seemed to matter in early AS (Moroni et al., 2019). Blood monocytes expressed receptors for vascular endothelial growth

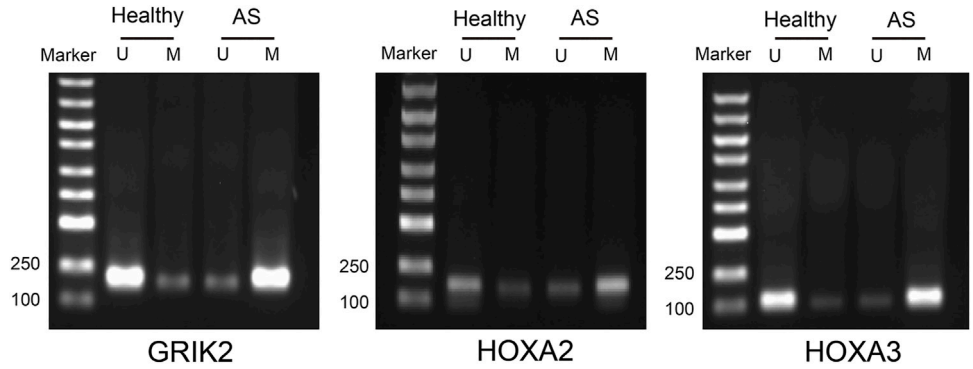


FIGURE 5
Methylation-specific PCR detection of the methylation statuses in the CpG islands of the promoter regions in atherosclerosis and healthy samples. U, un-methylated; M, methylated.

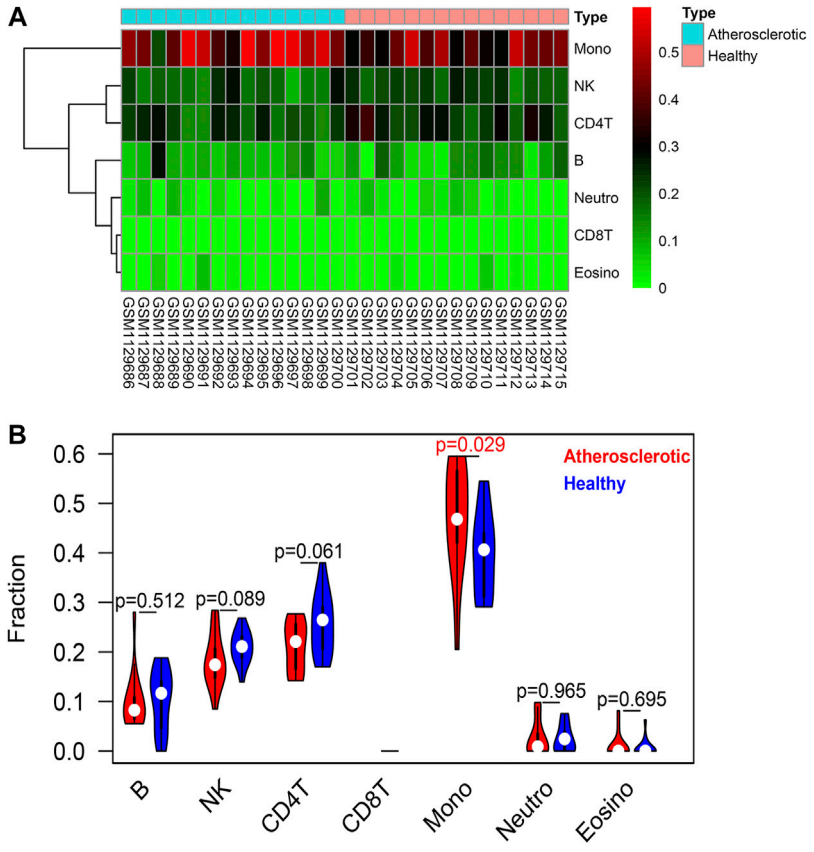


FIGURE 6
Immune-cell infiltration analysis based on the EpiDISH algorithm. **(A)** Heat map of the infiltration levels of seven types of immune cells (B cells, NK cells, CD4$^+$ T cells, CD8$^+$ T cells, monocytes, neutrophils, and eosinophils). **(B)** Violin plot and difference analysis between the atherosclerotic and healthy groups.

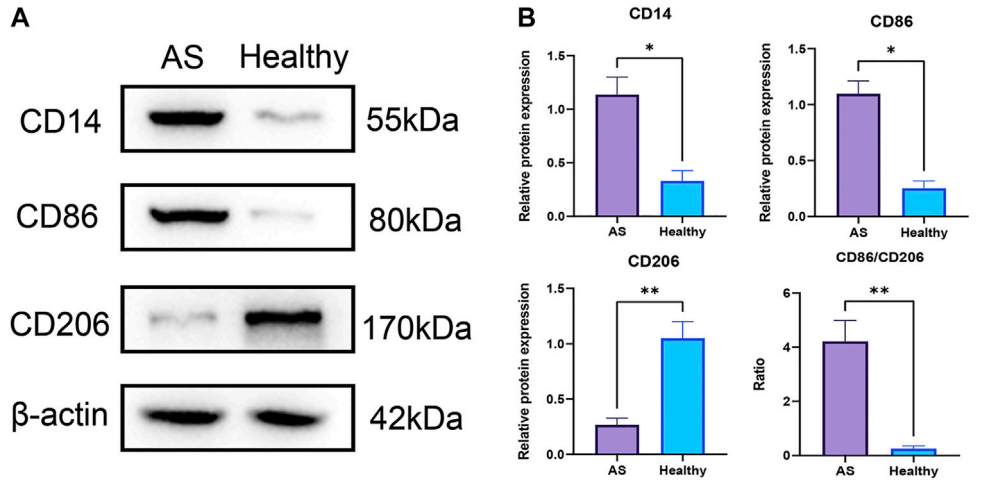


FIGURE 7
Infiltrations of monocytes and M1-type macrophages were significantly increased in atherosclerotic tissues at pre-atheroma stage. **(A)** Western blots of the monocyte marker CD14, M1-type macrophage marker CD86, and M2-type macrophage marker CD206. **(B)** Differential analyses of the relative protein expression levels. **: $p < 0.01$; *: $p < 0.05$.

factors for endothelial cells. After activation, monocytes were associated with inflammation by producing inflammatory molecules. Notably, monocytes consisted of distinct subsets with varied cell surface markers and functional characteristics, which may be relevant to angiogenic processes in AS (Jaipersad et al., 2014). The macrophages with four subtypes (M2a, M2b, M2c, and M2d) have different impacts on AS (Moore K. J. et al., 2013). M1 and M2 are strongly related to vascular calcification. Both the M1 and M2 phenotypes are found in the early and advanced lesions of AS. M1 macrophages predominated in unstable plaques, as AS developed, the number of M2 macrophages decreased (Bisgaard et al., 2016). However, the proportion of M2 macrophages in stable plaques was relatively higher and more M2 macrophages were required in plaque regression (Rahman et al., 2017; Yang et al., 2020). Our results have shown similarities with these conclusions, in AS tissues at the pre-atheroma stage, the expression levels of monocyte marker CD14 and M1-type macrophage marker CD86 were significantly increased, indicating higher immune filtration of monocyte and M1-type macrophage, while that of M2-type macrophage marker protein CD206 was significantly decreased, indicating a reduced ratio of M2-type macrophage. These results suggested the early formation of plaque.

Some studies have tried to explore the relationship between DNA methylation and immune infiltrate, which may assist in understanding the immune-related mechanism of certain diseases. In one study on hepatocellular carcinoma, a comprehensive analysis has been performed to explore the cell division cycle-associated family genes (CDCAs) methylation and immune infiltrates. The biological enrichment analysis of CDCAs demonstrated that they were significantly associated with the immune function regulation of infiltrating immune cells. Also, the methylation analysis of CDCAs indicated an association with the tumor immunogenicity, i.e., low-methylation of CDCA1, CDCA2, and CDCA8 dramatically reduced the immune infiltrate levels of T cells and cytotoxic lymphocytes. In addition, CDCA1-6 and CDCA8 with low-methylation levels significantly deteriorated the overall survival of HCC patients. It concluded that the methylation levels of CDCAs were related to the prognostic value and infiltrating immune differences, which could be a convincing biomarker for predicting the response of immunotherapy (Wang et al., 2020). In another study on hepatocellular carcinoma, LOXL3 was first reported to link with immune infiltrates. A statistical analysis has been applied to explore the relationship between the LOXL3 expression and the infiltration of multiple immune cells (Wang N. et al., 2021; Triki et al., 2022). The association between target gene and immune-cell infiltration can be further verified experimentally, with quantitative multiplex immunohistochemistry or immunohistochemistry staining (Wang Y. et al., 2021; Gatti et al., 2021).

There are still limitations in this study. One limitation is the sample resource. For the early diagnosis and prevention of AS, the expression profile of biomarkers during early AS should be focused on. Thus, the samples should be collected from the population without AS but developed into AS in the future, such as those with familial history of AS (Wright et al., 2021). However, most of the existing data have been derived from AS patients. Considering the complex progress during AS development, bias is inevitable. An early and lasting monitoring plan can be designed and conducted for obtaining more valuable data from people in the pre-AS stages. Another

limitation is that the potential relationship between methylation of target genes and immune infiltration can be further explored. The statistical analysis should be preliminarily performed, and then, the relationship can be experimentally verified with immunohistochemistry. It may assist in understanding the roles of target genes in immuno-microenvironment during disease progression.

Conclusion

Both the biomarkers and potential immune cell–relevant mechanisms were demonstrated with both bioinformatic analysis and experimental results. The bioinformatic analysis revealed three differentially methylated genes (GRIK2, HOXA2, and HOXA3) in the CpG islands of the promoter regions between healthy and AS groups. These genes can be applied as biomarkers for the early diagnosis of AS. For the AS tissues at the pre-atheroma stage, the expression levels of monocyte marker CD14 and M1-type macrophage marker CD86 were significantly increased while that of M2-type macrophage marker CD206 was significantly decreased. The specific immune-cell filtration conditions may also assist in understanding the progression of AS and finding the treatment targeting specific immune cells.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by the Scientific Research Ethics Committee of the Central Hospital of Shengli Oil Field. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CW and HC designed this study. CX and HC performed the bioinformatic analyses. CX, DS, and CW conducted the MS-PCR and Western blotting experiments and collected the original data. CW and HC performed the statistics and wrote the article. All authors read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.989459/full#supplementary-material

References

Ahmadi, A., Argulian, E., Leipsic, J., Newby, D. E., and Narula, J. (2019). From subclinical atherosclerosis to plaque progression and acute coronary events: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* 74 (12), 1608–1617. doi:10.1016/j.jacc.2019.08.012

Bäck, M., Yurdagul, A., Jr., Tabas, I., Öörni, K., and Kovanen, P. T. (2019). Inflammation and its resolution in atherosclerosis: mediators and therapeutic opportunities. *Nat. Rev. Cardiol.* 16 (7), 389–406. doi:10.1038/s41569-019-0169-2

Bisgaard, L. S., Mogensen, C. K., Rosendahl, A., Cucak, H., Nielsen, L. B., Rasmussen, S. E., et al. (2016). Bone marrow-derived and peritoneal macrophages have different inflammatory response to oxLDL and M1/M2 marker expression - implications for atherosclerosis research. *Sci. Rep.* 6, 35234. doi:10.1038/srep35234

Borghini, A., Cervelli, T., Galli, A., and Andreassi, M. G. (2013). DNA modifications in atherosclerosis: From the past to the future. *Atherosclerosis* 230 (2), 202–209. doi:10.1016/j.atherosclerosis.2013.07.038

Chen, W. D., Song, T., Cao, Q. H., Li, R., Wang, H., Chen, X. B., et al. (2020). Atherosclerosis prediction by microarray-based DNA methylation analysis. *Exp. Ther. Med.* 20 (3), 2863–2869. doi:10.3892/etm.2020.9025

Deaton, A. M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes. Dev.* 25 (10), 1010–1022. doi:10.1101/gad.2037511

Dong, C., Yoon, W., and Goldschmidt-Clermont, P. J. (2002). DNA methylation and atherosclerosis. *J. Nutr.* 132 (8), 2406s–2409s. doi:10.1093/jn/132.8.2406S

Ehrlich, M. (2019). DNA hypermethylation in disease: Mechanisms and clinical relevance. *Epigenetics* 14 (12), 1141–1163. doi:10.1080/15592294.2019.1638701

Fernández-Sanlés, A., Sayols-Baixeras, S., Subirana, I., Degano, I. R., and Elosua, R. (2017). Association between DNA methylation and coronary heart disease or other atherosclerotic events: a systematic review. *Atherosclerosis* 263, 325–333. doi:10.1016/j.atherosclerosis.2017.05.022

Gatti, G., Betts, C., Rocha, D., Nicola, M., Grupe, V., Ditada, C., et al. (2021). High IRF8 expression correlates with CD8 T cell infiltration and is a predictive biomarker of therapy response in ER-negative breast cancer. *Breast Cancer Res.* 23 (1), 40. doi:10.1186/s13058-021-01418-7

Groh, L., Keating, S. T., Joosten, L. A. B., Netea, M. G., and Riksen, N. P. (2018). Monocyte and macrophage immunometabolism in atherosclerosis. *Semin. Immunopathol.* 40 (2), 203–214. doi:10.1007/s00281-017-0656-7

Hai, Z., and Zuo, W. (2016). Aberrant DNA methylation in the pathogenesis of atherosclerosis. *Clin. Chim. Acta.* 456, 69–74. doi:10.1016/j.cca.2016.02.026

Hata, A., Nakajima, T., Matsusaka, K., Fukuyo, M., Morimoto, J., Yamamoto, T., et al. (2020). A low DNA methylation epigenotype in lung squamous cell carcinoma and its association with idiopathic pulmonary fibrosis and poorer prognosis. *Int. J. Cancer* 146 (2), 388–399. doi:10.1002/ijc.32532

Herrington, W., Lacey, B., Sherliker, P., Armitage, J., and Lewington, S. (2016). Epidemiology of atherosclerosis and the potential to reduce the global burden of atherothrombotic disease. *Circ. Res.* 118 (4), 535–546. doi:10.1161/circresaha.115.307611

Huang, Y. S., Zhi, Y. F., and Wang, S. R. (2009). Hypermethylation of estrogen receptor-alpha gene in atheromatosis patients and its correlation with homocysteine. *Pathophysiology* 16 (4), 259–265. doi:10.1016/j.pathophys.2009.02.010

Jaipersad, A. S., Lip, G. Y., Silverman, S., and Shantsila, E. (2014). The role of monocytes in angiogenesis and atherosclerosis. *J. Am. Coll. Cardiol.* 63 (1), 1–11. doi:10.1016/j.jacc.2013.09.019

Kalea, A. Z., Drosatos, K., and Buxton, J. L. (2018). Nutriepigenetics and cardiovascular disease. *Curr. Opin. Clin. Nutr. Metab. Care* 21 (4), 252–259. doi:10.1097/mco.0000000000000477

Khyzha, N., Alizada, A., Wilson, M. D., and Fish, J. E. (2017). Epigenetics of atherosclerosis: emerging mechanisms and methods. *Trends Mol. Med.* 23 (4), 332–347. doi:10.1016/j.molmed.2017.02.004

Kim, K. W., Ivanov, S., and Williams, J. W. (2020). Monocyte recruitment, specification, and function in atherosclerosis. *Cells* 10 (1), E15. doi:10.3390/cells10010015

Li, D., Bai, Y., Feng, Z., Li, W., Yang, C., Guo, Y., et al. (2019). Study of promoter methylation patterns of HOXA2, HOXA5, and HOXA6 and its clinicopathological characteristics in colorectal cancer. *Front. Oncol.* 9, 394. doi:10.3389/fonc.2019.00394

Li, J., Zhang, X., Yang, M., Yang, H., Xu, N., Fan, X., et al. (2021). DNA methylome profiling reveals epigenetic regulation of lipoprotein-associated phospholipase A(2) in human vulnerable atherosclerotic plaque. *Clin. Epigenetics* 13 (1), 161. doi:10.1186/s13148-021-01152-z

Li, L. C. (2007). Designing PCR primer for DNA methylation mapping. *Methods Mol. Biol.* 402, 371–384. doi:10.1007/978-1-59745-528-2_19

Libby, P., Buring, J. E., Badimon, L., Hansson, G. K., Deanfield, J., Bittencourt, M. S., et al. (2019). Atherosclerosis. *Nat. Rev. Dis. Prim.* 5 (1), 56. doi:10.1038/s41572-019-0106-z

Libby, P. (2021). The changing landscape of atherosclerosis. *Nature* 592 (7855), 524–533. doi:10.1038/s41586-021-03392-8

Ma, S. C., Hao, Y. J., Jiao, Y., Wang, Y. H., Xu, L. B., Mao, C. Y., et al. (2017). Homocysteine-induced oxidative stress through TLR4/NF-κB/DNMT1-mediated LOX-1 DNA methylation in endothelial cells. *Mol. Med. Rep.* 16 (6), 9181–9188. doi:10.3892/mmr.2017.7753

Moore, K. J., Sheedy, F. J., and Fisher, E. A. (2013a). Macrophages in atherosclerosis: a dynamic balance. *Nat. Rev. Immunol.* 13 (10), 709–721. doi:10.1038/nri3520

Moore, L. D., Le, T., and Fan, G. (2013b). DNA methylation and its basic function. *Neuropsychopharmacology* 38 (1), 23–38. doi:10.1038/npp.2012.112

Moroni, F., Ammirati, E., Norata, G. D., Magnoni, M., and Camici, P. G. (2019). The role of monocytes and macrophages in human atherosclerosis, plaque neoangiogenesis, and atherothrombosis. *Mediat. Inflamm.* 2019, 7434376. doi:10.1155/2019/7434376

Mushenkova, N. V., Summerhill, V. I., Zhang, D., Romanenko, E. B., Grechko, A. V., and Orekhov, A. N. (2020). Current advances in the diagnostic imaging of atherosclerosis: insights into the pathophysiology of vulnerable plaque. *Int. J. Mol. Sci.* 21 (8), E2992. doi:10.3390/ijms21082992

Nagy, C., Suderman, M., Yang, J., Szyf, M., Mechawar, N., Ernst, C., et al. (2015). Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Mol. Psychiatry* 20 (3), 320–328. doi:10.1038/mp.2014.21

Nazarenko, M. S., Markov, A. V., Lebedev, I. N., Sleptsov, A. A., Frolov, A. V., Barbash, O. L., et al. (2013). DNA methylation profiling of the vascular tissues in the setting of atherosclerosis. *Mol. Biol.* 47 (3), 398–404. doi:10.7868/s0026898413030099

O'Hagan, H. M., Wang, W., Sen, S., Destefano Shields, C., Lee, S. S., Zhang, Y. W., et al. (2011). Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* 20 (5), 606–619. doi:10.1016/j.ccr.2011.09.012

Rader, D. J., and Daugherty, A. (2008). Translating molecular discoveries into new therapies for atherosclerosis. *Nature* 451 (7181), 904–913. doi:10.1038/nature06796

Rahman, K., Vengrenyuk, Y., Ramsey, S. A., Vila, N. R., Girgis, N. M., Liu, J., et al. (2017). Inflammatory Ly6Chi monocytes and their conversion to M2 macrophages drive atherosclerosis regression. *J. Clin. Invest.* 127 (8), 2904–2915. doi:10.1172/jci75005

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 12, 77. doi:10.1186/1471-2105-12-77

Song, Y. P., Xian, P., Luo, H., Dai, J. Y., Bai, Y., Li, Y., et al. (2022). Comprehensive landscape of HOXA2, HOXA9, and HOXA10 as potential biomarkers for predicting progression and prognosis in prostate cancer. *J. Immunol. Res.* 2022, 5740971. doi:10.1155/2022/5740971

Tabaei, S., and Tabaee, S. S. (2019). DNA methylation abnormalities in atherosclerosis. *Artif. Cells Nanomed. Biotechnol.* 47 (1), 2031–2041. doi:10.1080/21691401.2019.1617724

Tan, L., Xu, Q., Shi, R., and Zhang, G. (2021). Bioinformatics analysis reveals the landscape of immune cell infiltration and immune-related pathways participating in the progression of carotid atherosclerotic plaques. *Artif. Cells Nanomed. Biotechnol.* 49 (1), 96–107. doi:10.1080/21691401.2021.1873798

Tang, H., Zeng, Z., Shang, C., Li, Q., and Liu, J. (2021). Epigenetic regulation in pathology of atherosclerosis: a novel perspective. *Front. Genet.* 12, 810689. doi:10.3389/fgene.2021.810689

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., et al. (2013). A beta-mixture quantile normalization method for correcting probe design bias in illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29 (2), 189–196. doi:10.1093/bioinformatics/bts680

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., et al. (2017). ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 33 (24), 3982–3984. doi:10.1093/bioinformatics/btx513

Triki, H., Declerck, K., Charfi, S., Ben Kridis, W., Chaabane, K., Ben Halima, S., et al. (2022). Immune checkpoint CD155 promoter methylation profiling reveals cancer-associated behaviors within breast neoplasia. *Cancer Immunol. Immunother.* 71 (5), 1139–1155. doi:10.1007/s00262-021-03064-6

Wang, L., Gao, B., Wu, M., Yuan, W., Liang, P., and Huang, J. (2021a). Profiles of immune cell infiltration in carotid artery atherosclerosis based on gene expression data. *Front. Immunol.* 12, 599512. doi:10.3389/fimmu.2021.599512

Wang, N., Zhou, X., Tang, F., Wang, X., and Zhu, X. (2021b). Identification of LOXL3-associating immune infiltration landscape and prognostic value in hepatocellular carcinoma. *Virchows Arch.* 479 (6), 1153–1165. doi:10.1007/s00428-021-03193-4

Wang, W., Li, W., Wu, Y., Tian, X., Duan, H., Li, S., et al. (2021c). Genome-wide DNA methylation and gene expression analyses in monozygotic twins identify potential biomarkers of depression. *Transl. Psychiatry* 11 (1), 416. doi:10.1038/s41398-021-01536-y

Wang, Y., Gu, W., Wen, W., and Zhang, X. (2021d). SERPINH1 is a potential prognostic biomarker and correlated with immune infiltration: a pan-cancer analysis. *Front. Genet.* 12, 756094. doi:10.3389/fgene.2021.756094

Wang, Y., Yang, Y., Gao, H., Ouyang, T., Zhang, L., Hu, J., et al. (2020). Comprehensive analysis of CDCAs methylation and immune infiltrates in hepatocellular carcinoma. *Front. Oncol.* 10, 566183. doi:10.3389/fonc.2020.566183

Wright, R. S., Ray, K. K., Raal, F. J., Kallend, D. G., Jaros, M., Koenig, W., et al. (2021). Pooled patient-level analysis of inclisiran trials in patients with familial hypercholesterolemia or atherosclerosis. *J. Am. Coll. Cardiol.* 77 (9), 1182–1193. doi:10.1016/j.jacc.2020.12.058

Xu, S., Pelisek, J., and Jin, Z. G. (2018). Atherosclerosis is an epigenetic disease. *Trends Endocrinol. Metab.* 29 (11), 739–742. doi:10.1016/j.tem.2018.04.007

Yang, S., Yuan, H. Q., Hao, Y. M., Ren, Z., Qu, S. L., Liu, L. S., et al. (2020). Macrophage polarization in atherosclerosis. *Clin. Chim. Acta.* 501, 142–146. doi:10.1016/j.cca.2019.10.034

Yutani, C., Imakita, M., Ishibashi-Ueda, H., Tsukamoto, Y., Nishida, N., and Ikeda, Y. (1999). Coronary atherosclerosis and interventions: pathological sequences and restenosis. *Pathol. Int.* 49 (4), 273–290. doi:10.1046/j.1440-1827.1999.00861.x

Zaina, S. (2014). Unraveling the DNA methylome of atherosclerosis. *Curr. Opin. Lipidol.* 25 (2), 148–153. doi:10.1097/mol.0000000000000059

Zeybel, M., Vatansever, S., Hardy, T., Sarı, A. A., Cakalağaoğlu, F., Avcı, A., et al. (2016). DNA methylation profiling identifies novel markers of progression in hepatitis B-related chronic liver disease. *Clin. Epigenetics* 8, 48. doi:10.1186/s13148-016-0218-1

Zhang, Y., Mei, J., Li, J., Zhang, Y., Zhou, Q., and Xu, F. (2021). DNA methylation in atherosclerosis: a new perspective. *Evid. Based. Complement. Altern. Med.* 2021, 6623657. doi:10.1155/2021/6623657

Zheng, S. C., Breeze, C. E., Beck, S., Dong, D., Zhu, T., Ma, L., et al. (2019). EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. *Bioinformatics* 36 (6), btz833–1951. doi:10.1093/bioinformatics/btz833

Zhou, W., Laird, P. W., and Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 45 (4), e22. doi:10.1093/nar/gkw967

Zhu, Y., Xian, X., Wang, Z., Bi, Y., Chen, Q., Han, X., et al. (2018). Research progress on the relationship between atherosclerosis and inflammation. *Biomolecules* 8 (3), E80. doi:10.3390/biom8030080

frontiers | Frontiers in Cell and Developmental Biology

TYPE Review
PUBLISHED 12 September 2022
DOI 10.3389/fcell.2022.1012193

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Junnan Xu,
China Medical University, China
Jing Yang,
Jinzhou Medical University, China

*CORRESPONDENCE
Hua-Chuan Zheng,
zheng_huachuan@hotmail.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 05 August 2022
ACCEPTED 26 August 2022
PUBLISHED 12 September 2022

CITATION
Zheng H-C, Xue H and Zhang C-Y (2022), REG4 promotes the proliferation and anti-apoptosis of cancer. *Front. Cell Dev. Biol.* 10:1012193. doi: 10.3389/fcell.2022.1012193

REG4 promotes the proliferation and anti-apoptosis of cancer

Hua-Chuan Zheng[1]*, Hang Xue[1] and Cong-Yu Zhang[2]

[1]Department of Oncology and Central Laboratory, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Cancer Center, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

Regenerating islet-derived 4 (REG4) gene was discovered by high-throughput sequencing of ulcerative colitis cDNA libraries. REG4 is involved in infection and inflammation by enhancing macrophage polarization to M2, via activation of epidermal growth factor receptor (EGFR)/Akt/cAMP-responsive element binding and the killing inflammatory *Escherichia coli*, and closely linked to tumorigenesis. Its expression was transcriptionally activated by caudal type homeobox 2, GATA binding protein 6, GLI family zinc finger 1, SRY-box transcription factor 9, CD44 intracytoplasmic domain, activating transcription factor 2, and specificity protein 1, and translationally activated by miR-24. REG4 can interact with transmembrane CD44, G protein-coupled receptor 37, mannan and heparin on cancer cells. Its overexpression was observed in gastric, colorectal, pancreatic, gallbladder, ovarian and urothelial cancers, and is closely linked to their aggressive behaviors and a poor prognosis. Additionally, REG4 expression and recombinant REG4 aggravated such cellular phenotypes as tumorigenesis, proliferation, anti-apoptosis, chemoradioresistance, migration, invasion, peritoneal dissemination, tumor growth, and cancer stemness via EGFR/Akt/activator protein-1 and Akt/glycogen synthase kinase three β/β-catenin/transcription factor 4 pathways. Sorted REG4-positive deep crypt secretory cells promote organoid formation of single Lgr5 (+) colon stem cells by Notch inhibition and Wnt activation. Histologically, REG4 protein is specifically expressed in neuroendocrine tumors and signet ring cell carcinomas of the gastrointestinal tract, pancreas, ovary, and lung. It might support the histogenesis of gastric intestinal–metaplasia–globoid dysplasia–signet ring cell carcinoma. In this review, we summarized the structure, biological functions, and effects of REG4 on inflammation and cancer. We conclude that REG4 may be employed as a biomarker of tumorigenesis, subsequent progression and poor prognosis of cancer, and may be a useful target for gene therapy.

KEYWORDS

cancer, REG4, tumor suppressor, tumor phenotype, transcriptional regulation

Introduction

In 1984, Yonemura et al. (Yonemura et al., 1984) discovered regenerating islet-derived (REG) proteins during the regeneration of pancreatic islets. The REG family belongs to the calcium-dependent lectin (C-type lectin) gene superfamily, which encodes four multi-functional and secreted small proteins. REG proteins serve as anti-apoptotic factors, acute phase reactants, lectins, and growth factors for neural cells, pancreatic β cells, and epithelial cells in the digestive system. To date, researchers have identified human *REG I* (*Iα and Iβ*), *REG III* (*III* and *HIP/PAP*), and *REG4*, which encode homologous 158–175aa proteins. *REG* genes are located on chromosomes 2p12 (*HIP/PAP*, *REG Iα*, *REG Iβ*, and *REG III*) and 1q12-q21 (*REG4*) (Nata et al., 2004).

The discovery and expression profile of REG4

REG4 was discovered by high-throughput sequencing of a cDNA library from an ulcerative colitis (UC) sample in 2001 (Hartupee et al., 2001). The *REG4* gene has 10 exons and encodes four types of variants by alternative splicing, which contain different open reading frames. Its longest cDNA has an open reading frame of 477 bp and encodes an 18-kDa peptide of 158aa. The REG4 protein is composed of a 22aa signal peptide and calcium-dependent lectin domain, within which are an N-glycosylation site and two carbohydrate binding sites (Figure 1) (Zhang et al., 2021). In the rat, *REG4* mRNA was detected in the brain cortex, stomach, pancreas, spleen, small intestine, colon, kidney, and urinary bladder, but not in the sciatic nerve, thymus, liver, cerebellum, suprarenal gland, heart, soleus muscle, lung, and esophagus by reverse transcriptase–PCR. Using western blot, REG4 protein expression was detectable in the pancreas, stomach, small intestine, colon, spleen, brain cortex, kidney, and urinary bladder, but not in the sciatic nerve, liver, thymus, cerebellum, suprarenal gland, soleus muscle, heart, esophagus, and lung. Immunohistochemically, positive REG4 staining was detected throughout the gastric mucosa and was mainly distributed in the basal portion of intestinal crypts. Pancreatic acinar cells appeared positive for REG4, but not pancreatic islet β cells. REG4 was expressed in large spleen cells with a large nucleus in the red pulp, but not the white pulp. Immunoreactivity for REG4 was found in large neurons of the brain cortex, and in glomerular and urinary bladder epithelial cells, but rarely in renal tubular cells (Azman et al., 2011). REG4 immunoreactivity was significantly higher in the ovary than the uterus. The expression of REG4 was strongly detectable in oocytes and granulosa cells of ovarian follicles, interstitial cells and corpus luteum, while only weak expression was found in the glandular and luminal epithelium of the rat endometrium (Du and Yao, 2013). The tissue-specific expression of REG4 is closely linked to its biological function in different organs.

The regulation of REG4 expression

In mammalian cells, caudal type homeobox 2 (CDX2) was found to induce *REG4* expression by binding to consensus

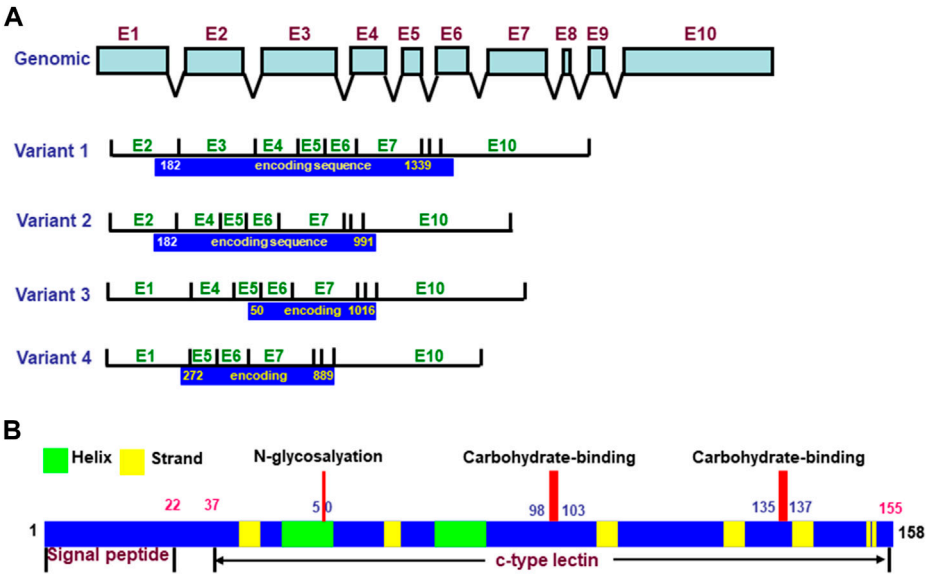


FIGURE 1
Structures of *REG4* gene and protein. The *REG4* gene has 10 exons and is alternatively spliced into four variants with different open reading frames **(A)**. The encoding protein of *REG4* variant one produces a 158aa protein with signal peptide and c-type lectin domain **(B)**.

CDX2-binding elements upstream of the *REG4* gene, supported by the positive relationship between CDX2 and REG4 expression in gastric cancer cells and tissues (Naito et al., 2012; Chai et al., 2021). MicroRNA (miR)-363 suppressed the translation of GATA binding protein 6 (GATA6), which functioned as a transcriptional factor to induce REG4 and leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5) expression essential for the growth of colon cancer cells under adherent conditions (Kawasaki et al., 2015). The key transcriptional factor in the Hedgehog signaling pathway, GLI family zinc finger 1 (GLI1) bound to *REG4* promoter regions (GATCATCCA) for its transcription and translation in pancreatic cancer cells, supported by the synergic expression of REG4 and GLI1 (Wang et al., 2011). Additionally, SRY-box transcription factor 9 (SOX9) knockdown upregulated REG4 protein expression in gastric cancer cells; a positive correlation of REG4 expression with SOX9 expression in gastric cancer was noted (Zhang et al., 2018). The activating transcription factor 2 (ATF2) targeted the *REG4* promoter to induce *REG4* expression during enteritis (Xiao et al., 2019). Duan et al. (Duan et al., 2014) found that the tumor suppressor, miR-24, translationally restrained the progression of gastric cancer by down-regulating REG4. In short, *REG4* expression was transcriptionally activated by CDX2, GATA6, GLI1, ATF2, and SOX9, and translationally activated by miR-24. However, further miRNAs may be discovered in future that are associated with the translational regulation of *REG4*.

REG4-related signal pathways

As for the cellular signaling pathway, recombinant human REG4 (rhREG4) treatment resulted in anti-apoptosis of colorectal cancer cells with the overexpression of B cell lymphoma-extra large (Bcl-xL), B cell lymphoma 2 (Bcl-2), survivin, and matrix metalloproteinase-7 (MMP-7), and the phosphorylation of epidermal growth factor receptor (EGFR) at Tyr992 and Tyr1068, and Akt at Thr308 and Ser473. It also strengthened the transcriptional activity of activator protein-1 (AP-1) by interaction with JunB, JunD, and FosB (Bishnupuri et al., 2006b). rhREG4 treatment also protected normal intestinal crypt cells from irradiation-induced apoptosis by enhancing the expression of Bcl-2, Bcl-xL, and survivin, in agreement with data from human colorectal cancer cells (Bishnupuri et al., 2010). rhREG4 treatment promoted G_2 progression for the mitogenesis of colorectal cancer cells by Akt/glycogen synthase kinase three β (GSK3β)/β-catenin/transcription factor 4 (TCF-4) signaling (Bishnupuri et al., 2014). Meanwhile, REG4 might protect acinar cells against necrosis in experimental pancreatitis by enhancing the expression of Bcl-2 and Bcl-xL *via* activation of the EGFR/Akt pathway (Hu et al., 2011). Li et al. (Li et al., 2011) found anti-tumor effects of proteoglycan from *Phellinus linteus* on colorectal cancer cells *via* inactivation of the REG4/EGFR/Akt pathway. Taken together, these findings suggested that REG4 is a potent activator of the EGFR/Akt pathway for the proliferation and anti-apoptosis of colorectal cancer cells.

In addition, Ho et al. (2010) demonstrated that REG4 bound to mannan and heparin in the absence of calcium. In addition, REG4 was found to interact with CD44 to activate its regulated intramembrane proteolysis. This resulted in the γ-secretase–mediated cleavage and release of the CD44 intracytoplasmic domain (CD44ICD) that functions as a transcriptional activator of D-type cyclins involving in cell proliferation, and Kruppel-like factor 4 and SRY-box transcription factor 2 (SOX2) expression involved in the pluripotency of cancer stem cells (Figure 2) (Bishnupuri et al., 2022). A significant correlation between REG4 and CD44 or CD44ICD supported the above-mentioned hypothesis (Sninsky et al., 2021). Liu et al. (2013) found that REG4 down-regulation also resulted in the hypoexpression of p21 and p27, which negatively regulated cyclin D1 and blocked the G_1/S transition of prostate cancer cells. Wang et al. (2016) showed that transforming growth factor (TGF)-α stimulated specificity protein 1 (SP1) to transcriptionally promote REG4 expression, while G protein-coupled receptor 37 (GPR37) complexed with REG4, which mediated EGFR signal transduction by REG4 and promoted peritoneal metastasis of gastric cancer cells. Therefore, TGF-α/EGFR/SP1 was responsible for the transcriptional activation of REG4 in a positive feedback loop.

Infection and inflammation

REG4 gene was screened from UC samples (Hartupee et al., 2001). *REG4* mRNA was found to be up-regulated in Crohn's disease and UC samples (Takasawa et al., 2018). Nanakin et al. (2007) found that *REG4* mRNA was strongly expressed in inflammatory epithelium, and dysplastic and cancerous lesions, and positively correlated with the expression of basic fibroblast growth factor (bFGF) and hepatocyte growth factor (HGF) mRNA expression in UC. In pediatric patients with intestinal failure, serum REG4 was positively correlated with serum interleukin (IL)-6 and tumor necrosis factor (TNF)-α, and REG4 protein was increased and highly expressed toward the luminal face of inflamed intestine. In intestinal conditional REG4 knockout mice, REG4 abrogation altered the colonic bacterial composition, and weakened bacterial adhesion to the colonic mucosa, finally ameliorating dextran sodium sulfate-induced colitis (Xiao et al., 2019). Further study indicated that REG4 stimulated complement-mediated attack complexes to eliminate intestinal dominant *Escherichia coli* in order to maintain homeostasis. These results supported the protective effects of REG4 on the epithelia of UC, possibly by being anti-inflammatory and anti-infection (Qi et al., 2020).

Hu et al. (Hu et al., 2011) also found that REG4 expression was significantly up-regulated during acute pancreatitis. The

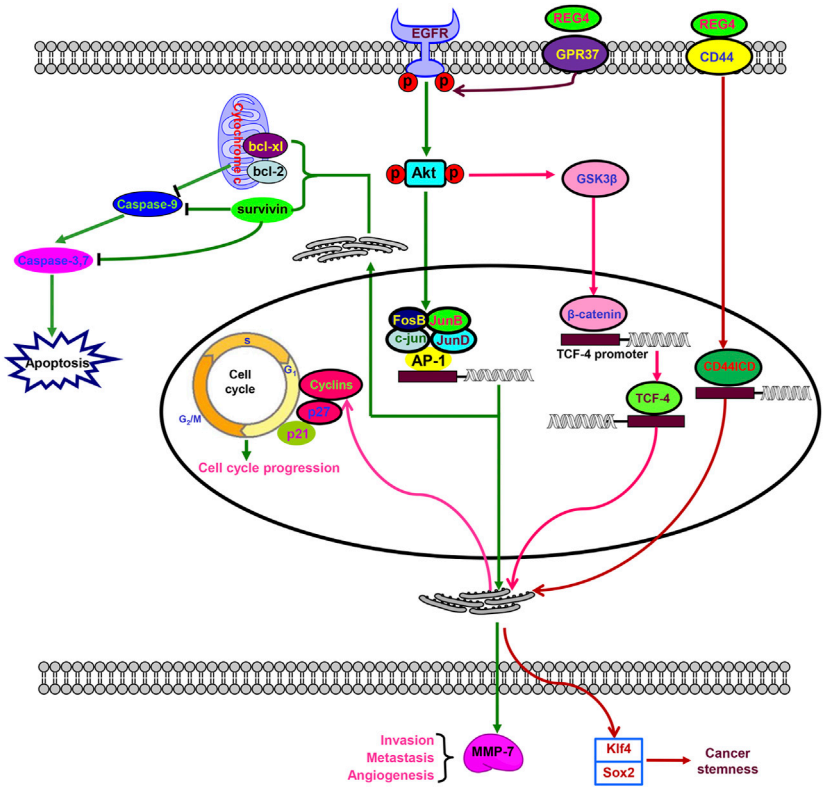


FIGURE 2
Biological functions of REG4. REG4 indirectly interacts with epidermal growth factor receptor (EGFR) to phosphorylate and activate Akt, which induces activator protein-1 (AP-1)–mediated transcriptional initiation by an AP-1/c-Jun/JunA/JunB complex and glycogen synthase kinase three β (GSK3β)–β-catenin–transcription factor 4 (TCF-4) signaling. Additionally, REG4 can bind to G protein-coupled receptor 37 (GPR37) to activate EGFR. REG4 interacts with CD44 to activate regulated transmembrane proteolysis of CD44 resulting in γ-secretase–mediated cleavage and release of the CD44 intracytoplasmic domain (CD44ICD). TCF-4 and CD44ICD serve as transcriptional activators that up-regulate the expression of cyclins for cell-cycle progression. CD44ICD-mediated Kruppel-like factor 4 (Klf4) and SRY-box transcription factor 2 (SOX2) expression is involved in cancer stemness. Activator protein-1–induced overexpression of B cell lymphoma-extra large (Bcl-xL), B cell lymphoma 2 (Bcl-2), survivin, and matrix metalloproteinase-7 (MMP7) plays an important role in apoptosis, angiogenesis, invasion, and metastasis.

REG4 secreted by pancreatic cancer cells promoted macrophage polarization to M2 via EGFR/Akt/cAMP-responsive element binding activation, finally promoting tumor growth and distant metastasis (Ma et al., 2016). Li et al. (2021) found that rhREG4 attenuated the severity of rat osteoarthritis by facilitating the proliferation of articular chondrocytes. In a rat model of acutely-injured liver, treatment with recombinant interleukin 22 (IL-22) lentivirus reduced serum total bilirubin, alanine and aspartate transaminases, and enhanced REG4 expression, suggesting that REG4 might be involved in the protective effects of IL-22 on hepatic injury (Zhang et al., 2015).

Gastric cancer

At the mRNA level, *REG4* gene was significantly up-regulated in gastric cancer compared with normal mucosa (Tao et al., 2011), while our group showed higher *REG4* expression in intestinal metaplasia than in gastritis and gastric cancer (Zheng et al., 2010). *REG4* mRNA was found to positively correlate with the wall penetration (Miyagawa et al., 2008), depth of invasion, and clinicopathological stages (Ying et al., 2013) of gastric cancer. These results indicate that *REG4* mRNA expression might reflect gastric carcinogenesis and subsequent progression.

In the stomach, foveolar epithelium was negative for REG4, whereas goblet and neuroendocrine cells of intestinal metaplasia were positive for REG4 (Oue et al., 2005). Meanwhile, REG4 expression was significantly associated with both the intestinal mucin phenotype (mucin 2 [MUC2] and CDX2) and neuroendocrine differentiation of gastric cancer (Oue et al., 2005; Yamagishi et al., 2009). Zheng et al. (2010) reported that REG4 immunostaining was gradually decreased from intestinal metaplasia, adenoma, cancer to gastritis, positively correlated with mucin 5AC (MUC-5AC) and MUC-2 expression, and was most frequently expressed in

signet ring cell carcinoma (SRCCs). The expression of REG4 was found to be significantly correlated with advanced T stage, N stage, M stage, TNM stage, frequent peritoneal recurrence and dissemination, diffuse-type carcinoma, and dedifferentiation of gastric cancer (Yamagishi et al., 2009; Tao et al., 2011; Moon et al., 2012). REG4 was detected in peritoneal lavage fluids of gastric cancer patients as well (Miyagawa et al., 2008; Kuniyasu et al., 2009; Yamagishi et al., 2009). The serum REG4 level was higher in patients with gastric cancer than in healthy individuals, in advanced than early gastric cancer patients, and in pre-surgical than post-surgical gastric cancer patients respectively (Mitani et al., 2007; Miyagawa et al., 2008; Kobayashi et al., 2010; Zheng et al., 2010). As for prognosis, REG4 immunoexpression was considered as an independent prognostic factor for both worse peritoneal recurrence-free and overall survival (Miyagawa et al., 2008; Tao et al., 2011; Moon et al., 2012). These data demonstrated that REG4 expression is involved in the histogenesis of gastric SRCC, neuroendocrine differentiation, and gastric carcinogenesis, and is closely linked to aggressive behaviors and adverse prognosis in gastric cancers.

With regard to drug resistance, Ying et al. (2013) found that up-regulation of *REG4* mRNA was closely linked to the intrinsic drug resistance of gastric cancer cells to fluorouracil (5-FU) or its combination therapy. All 14 REG4-positive patients with gastric cancer showed no change or disease progression when treated with a combination of low-dose 5-FU and cisplatin (Mitani et al., 2007). In gastric cancer cells, REG4 enhanced the resistance of gastric cancer cells to 5-FU through the mitogen-activated protein kinase/extracellular-signal-regulated kinase/Bim pathway (Jin et al., 2017). REG4 antibody significantly inhibited proliferation and chemosensitivity of gastric cancer cells to 5-FU (Zhang et al., 2019) and REG4 silencing caused the loss of stemness properties (Zhou et al., 2013). It was suggested that REG4 overexpression predicted chemoresistance in gastric cancer cells, possibly by promoting proliferation and stemness.

As for molecular mechanisms, REG4 expression facilitated invasion and migration of gastric cancer cells by up-regulating SOX9 expression, in contrast to REG4 knockdown (Zhang et al., 2018). Another report described how REG4 promoted proliferation, tumor growth, and migration of gastric cancer cells through the protein kinase B pathway (Huang et al., 2014a). Katsuno et al. (2012) found that coexpression of aldehyde dehydrogenase one and REG4 was involved in the tumorigenesis of diffuse-type gastric carcinoma, which was blocked by TGF-β. Kuniyasu et al. (2009) found that REG4 overexpression increased levels of Bcl-xL, Bcl-2, survivin, phosphorylated Akt, and EGFR, and decreased nitric oxide-induced apoptosis in gastric cancer cells, in contrast to REG4 silencing. In mice models of gastric cancer, REG4 expression enhanced peritoneal metastasis, weakened apoptosis, and shortened survival time (Miyagawa et al., 2008; Kuniyasu et al., 2009). These results demonstrated that REG4 aggravated the proliferation, anti-apoptosis, tumor growth, and peritoneal metastasis of gastric cancer cells.

Colorectal cancer

At the genetic level, Lu et al. (2013) found that a single nucleotide polymorphism in *REG4* might be a genetic marker for the progression of colorectal cancer. In colorectal tissues, *REG4* mRNA–positive cells are mostly enteroendocrine and goblet cells. Adenomatous and cancer cells positive for *REG4* mRNA exhibited enterocyte-like, mucus-secreting, or undifferentiated features (Violette et al., 2003), in agreement with observations in the stomach (Oue et al., 2005; Yamagishi et al., 2009; Zheng et al., 2010). *REG4* mRNA was found to be highly expressed in all adenoma samples, with or without concurrent colorectal carcinoma, compared to normal mucosa samples (Zhang et al., 2003a; Zhang et al., 2003b). Statistically, *REG4* mRNA was more expressed in colorectal cancers (especially mucinous carcinomas) than in normal colorectal mucosa (Violette et al., 2003). Combining these results, we conclude that up-regulated *REG4* mRNA expression is markedly observed in colorectal adenoma and adenocarcinoma.

At the protein level, REG4 expression was observed in both the middle and outer parts of crypts and superficial epithelium, especially goblets (Granlund et al., 2011). Statistically, REG4 expression was significantly lower in colorectal cancer than in normal mucosa or adenomas, and inversely correlated with poor differentiation, venous invasion, low expression of MUC2 and EGFR phosphorylated at Tyr1068 (Li et al., 2010). REG4 expression was less frequently observed in colorectal cancer than in adjacent non-neoplastic mucosa, in well- and moderately-differentiated adenomas than in mucinous carcinoma (Zheng et al., 2011), and in the cancers of the right colon than in the left colon and rectum, respectively (Kang et al., 2021). Further study showed that REG4 expression was associated with lymph node metastasis, distant metastasis, metastatic recurrence in the liver, advanced TNM stage, histologic grade, and MMP-7 expression in colorectal cancer (Oue et al., 2007; Zhu et al., 2015). Oue et al. (2007) found that the preoperative serum REG4 concentration was not elevated in patients with colorectal cancer at stages 0–III, but was significantly elevated in those at stage IV. Additionally, REG4 expression, as an independent predictor, was significantly linked to a worse prognosis in patients with colorectal cancer (Oue et al., 2007; Numata et al., 2011; He et al., 2014). However, Kaprio et al. (2014) reported that REG4 expression was an independent marker of a lower risk of death for patients with non-mucinous colorectal cancer, 65 years and younger, within 5 years. These findings demonstrated that aberrant REG4 expression is involved in the colorectal adenoma–adenocarcinoma sequence, and can be used to

indicate the aggressive behaviors and prognosis of colorectal cancers.

A body of evidence has shown that REG4 was markedly related to chemoresistance, migration, and invasion of cancer cells. Violette et al. (Zhang et al., 2003b) discovered that REG4 protein was strongly expressed in drug-resistant rectal cancer cells, but expressed weakly in drug-sensitive rectal cancer cells. REG4 expression was found to correlate with γ-radiation sensitivity in rectal cancer patients receiving radiotherapy (Kobunai et al., 2011). In radiochemotherapy (RCT)-sensitive colorectal cancer cells, REG4 expression was down-regulated, while it was increased in radiochemoresistant cells (Gao et al., 2021). REG4–overexpressing cells had a high survival rate and showed few DNA breaks after irradiation (He et al., 2014). rhREG4 significantly induced resistance to ionizing radiation in colon adenocarcinoma cells by promoting anti-apoptotic Bcl-xL and Bcl-2 expression (Numata et al., 2011). Additionally, rhREG4 stimulated cell growth in a paracrine manner. Notably, REG4 promoted migration and invasion of colorectal cancer cells via its carbohydrate-recognition domain in both autocrine and paracrine manners, which was significantly decreased by anti-REG4 antibody (Guo et al., 2010; Rafa et al., 2010). Nanakin et al. (2007) found that REG4 expression was stimulated by TNFα, epidermal growth factor (EGF), bFGF, and HGF in colon cancer cells, and then promoted cell proliferation and resistance to H_2O_2-induced apoptosis. These data indicated that REG4 might be identified as a potential marker for RCT resistance.

In an animal model, *REG4* mRNA was elevated in the intestine of $APC^{min/+}$ mice carrying APC mutation at codon 850 for transcription stop before a spontaneous second mutation of APC. Adenomas from 14-week-old $APC^{min/+}$ mice showed significantly up-regulated expression of Bcl-2 and REG4 (Bishnupuri et al., 2006a). Sorted REG4-positive deep crypt secretory (DCS) cells facilitated organoid formation of single Lgr5 (+) stem cells, and DCS cells overwhelmingly originated from Lgr5 (+) stem cells by both Notch inactivation and Wnt activation (Sasaki et al., 2016). In an organoid model, mutant *KRAS*-induced REG4 promoted colorectal cancer stemness via a Wnt/β-catenin pathway (Hwang et al., 2020). This was also evidenced by the positive correlation of cancer stem markers with REG4 in intestinal tumors from APC^{min+}/KrasG12D LA2 mice. These results indicated that REG4 might be involved in the colorectal adenoma–adenocarcinoma sequence by the regulation of local stem cells.

Pancreatic cancer

In pancreatic tumors, *REG4* mRNA expression was significantly higher in intestinal-type rather than in gastric-type intraductal papillary mucinous neoplasms and normal pancreatic ductal epithelium. REG4 expression was higher in borderline lesions and carcinoma than in adenoma, and in colloid carcinoma than that in tubular carcinoma, respectively, and positively correlated with CDX2 expression (Nakata et al., 2009). A high serum REG4 level could be used to discriminate chronic pancreatitis and pancreatic ductal adenocarcinoma, as well as predict worse survival (Takehara et al., 2006; Takayama et al., 2010; Saukkonen et al., 2018). The pancreatic cancer patients with a higher serum REG4 concentration had an unfavorable response to RCT and frequently experienced local recurrence after surgery (Eguchi et al., 2009). The knockdown of REG4 or an anti-REG4 antibody both attenuated the cell viability of pancreatic cancer cells, while rhREG4 exposure showed the opposite effect in a dose-dependent manner (Eguchi et al., 2009). He et al. (2012) found that REG4 promoted not only tumor growth but also invasion of pancreatic cancer cells by up-regulating MMP-7 and MMP-9. REG4–overexpressing pancreatic cancer cells were resistant to gemcitabine and γ-radiation (Eguchi et al., 2009). These data suggested that REG4 overexpression contributed to pancreatic carcinogenesis and subsequent progression by promoting proliferation, invasion, and RCT resistance.

Gallbladder cancer

REG4 mRNA expression was significantly higher in gallbladder adenocarcinoma than peritumoral normal tissues, adenoma, and cholecystitis (Yang et al., 2016), in line with findings by Tamura et al. (Tamura et al., 2009). Immunohistochemically, REG4 was negative in all normal gallbladders and cholelithiasis, but positive in 50% of intestinal metaplasia with adenomyomatosis, and positive in 56% of gallbladder carcinomas (Tamura et al., 2009). REG4 expression was positively correlated with dedifferentiation, local invasiveness, and lymph node metastasis of gallbladder cancer. REG4 expression was independently associated with a poor prognosis in patients with advanced gallbladder cancer (Yang et al., 2016). A high serum REG4 level was evident preoperatively in four (33%) of 12 patients with gallbladder cancer, but not in benign diseases, and was postoperatively reduced (Tamura et al., 2009). These findings indicated that REG4 is involved in the carcinogenesis and subsequent progression of gallbladder adenocarcinoma.

Ovarian cancer

In ovarian carcinogenesis, *REG4* mRNA and protein levels were higher in ovarian tumors than in normal ovaries, in mucinous carcinomas than in serous carcinomas, and in well- and moderately-differentiated carcinomas than in poorly-differentiated carcinomas, respectively (Chen et al., 2015), in line with another report (Xiang et al., 2022). REG4 protein

expression was significantly higher in ovarian mucinous borderline tumors and mucinous carcinomas than in mucinous cystadenomas, and was more closely associated with ovarian borderline, intestinal-type, mucinous tumors rather than in endocervical-like type tumors. A significant positive correlation existed between CDX2 and REG4 expression in primary ovarian mucinous tumors (Huang et al., 2014b). High *REG4* mRNA expression was inversely associated with inferior overall, progression-free and post-progression survival in patients with ovarian cancer receiving platinum chemotherapy (Xiang et al., 2022). As an independent factor, the expression of REG4 was an overall or relapse-free poor prognostic factor for patients with ovarian cancer (Chen et al., 2015). In ovarian cancer cells, REG4 expression or 5-FU chemoresistance was enhanced by CDX2 transfection, in contrast to CDX2 knockdown (Koh et al., 2019). Either REG4 overexpression or rhREG4 treatment promoted proliferation, G_2/S progression, anti-apoptosis, migration, invasion, and cisplatin and paclitaxel resistance in ovarian cancer cells (Chen et al., 2015; Xiang et al., 2022). Taken together, REG4 overexpression might play an important role in ovarian carcinogenesis and subsequent aggressiveness.

Urothelial cancer

In renal clear cell carcinoma, the immunoexpression of REG4 was not detectable due to a lack of neuroendocrine and intestinal differentiation (Hayashi et al., 2009). In prostate cancer, 14 (14%) of 98 cases had tissues positive for REG4 staining, which was associated with MUC2 and chromogranin A expression. The expression of REG4 was a significant prognostic factor and independent predictor of the relapse-free survival of patients with prostate cancer. In prostate cancer, the serum REG4 concentration was significantly higher in patients with prostate cancer than in control individuals (Ohara et al., 2008). The overexpression of REG4 was also observed in hormone refractory xenografts and refractory metastatic prostate cancer (Gu et al., 2005). rhREG4 treatment enhanced EGFR phosphorylation in prostate cancer cells (Ohara et al., 2008). Taking these findings together, we speculated that REG4 was involved in prostate carcinogenesis and subsequent progression, especially in those patients showing intestinal mucin and neuroendocrine differentiation.

Neuroendocrine tumors and SRCC

REG4 is physiologically found in selected enteroendocrine cells (EEC). REG4-positive gastric EECs were associated with serotonin, gastrin, somatostatin, and pancreatic polypeptide (Sentani et al., 2010). REG4 was differentially expressed in ECCs of the small intestine and colon (Heiskala and Andersson, 2013). REG4 showed a cellular co-distribution with serotonin, substance P or chromogranin A in the gastrointestinal tract. Subpopulations of REG4-positive cells overlapped with EECs containing GLP-1, GLP-2, peptide YY, secretin, and ghrelin, relying on the anatomical sites of the tissues. Therefore, REG4 is thought to be involved in intestinal and neuroendocrine differentiation. Oue et al. (2005) found that insulin-secreting β cells of the pancreas were positive for REG4. Of 21 gastric SRCCs, 16 colorectal SRCCs, 10 breast SRCCs, and 47 lung SRCCs, all gastric and colorectal SRCCs showed REG4 immunopositivity, but the others did not indicate that REG4 might be a biomarker specific for gastrointestinal SRCCs (Sentani et al., 2008).

Others

In the esophagus, REG4 staining was not detected in squamous cell carcinoma (SCC) and small cell carcinomas, whereas REG4 staining was found in four of 10 (40%) adenocarcinoma samples. The serum REG4 level was significantly higher in patients with SCC than in control participants, and correlated with the control participants' age (Oue et al., 2011). In the lung, REG4 was highly expressed in *KRAS*-mutant adenocarcinoma with thyroid transcription factor-1 (TTF-1) hypoexpression. Silencing significantly REG4 reduced proliferation and tumor growth, and arrested the cell cycle by regulating E2F targets and the G_2/M checkpoint (Sun et al., 2019). In glioma, REG4 expression was significantly higher in tumor than normal brain tissues. REG4 immunoreactivity was significantly associated with advanced pathological grade and a low Karnofsky performance score, and short survival as an independent prognostic factor (Wang et al., 2012). Sasahira et al. (2008) found that REG4 was expressed in salivary duct epithelia and acinus myoepithelia, but not in squamous epithelia. REG4 expression was found in 41% (17/41) of adenoid cystic carcinomas (ACCs), but not in SCCs, and was associated with lymph node involvement and a poor prognosis in ACC. These findings indicated that REG4 expression was not detectable in SCC, but in glioma, and ACC.

Conclusion and perspectives

In conclusion, altered REG4 expression is involved in infection and inflammation, and in gastric, colorectal, pancreatic, gallbladder, ovarian and urothelial cancers, where it is closely linked to aggressive behaviors and a poor prognosis. REG4 expression and recombinant

REG4 aggravated tumorigenesis, proliferation, anti-apoptosis, chemoradioresistance, migration, invasion, peritoneal dissemination, tumor growth, and cancer stemness by EGFR/Akt/AP-1 and Akt/GSK3β/β-catenin/TCF-4 pathways. Histologically, REG4 protein is highly expressed in neuroendocrine tumors and SRCCs. In accordance with recent findings about REG4, we believe that REG4 should be used as a biomarker for SRCC and neuroendocrine tumors, and contributes to the histogenesis of gastric intestinal–metaplasia–globoid dysplasia–SRCC. Aberrant REG4 expression should be employed to predict the tumorigenesis, aggressive behaviors and poor prognosis of malignancies, and the lavage REG4 level should be determined to guard against peritoneal dissemination.

Author contributions

HZ and HX were mainly responsible for literature review and manuscript writing. HZ and C-YZ completed the construction pictures. HZ designed the ideas of this paper and modified the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by: Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (21377772D) and National Natural Scientific Foundation of China (81672700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JY declared a shared parent affiliation with the author CZ to the handling editor at the time of review.

Publisher's note

References

Azman, J., Starcevic Klasan, G., Ivanac, D., Picard, A., Jurisic-Erzen, D., Nikolic, M., et al. (2011). Reg IV protein and mRNA expression in different rat organs. *Acta Histochem.* 113 (8), 793–797. doi:10.1016/j.acthis.2010.11.008

Bishnupuri, K. S., Luo, Q., Korzenik, J. R., Henderson, J. O., Houchen, C. W., Anant, S., et al. (2006a). Dysregulation of Reg gene expression occurs early in gastrointestinal tumorigenesis and regulates anti-apoptotic genes. *Cancer Biol. Ther.* 5 (12), 1714–1720. doi:10.4161/cbt.5.12.3469

Bishnupuri, K. S., Luo, Q., Murmu, N., Houchen, C. W., Anant, S., and Dieckgraefe, B. K. (2006b). Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. *Gastroenterology* 130 (1), 137–149. doi:10.1053/j.gastro.2005.10.001

Bishnupuri, K. S., Luo, Q., Sainathan, S. K., Kikuchi, K., Sureban, S. M., Sabarinathan, M., et al. (2010). Reg IV regulates normal intestinal and colorectal cancer cell susceptibility to radiation-induced apoptosis. *Gastroenterology* 138 (2), 616–626. doi:10.1053/j.gastro.2009.10.050

Bishnupuri, K. S., Sainathan, S. K., Bishnupuri, K., Leahy, D. R., Luo, Q., Anant, S., et al. (2014). Reg4-induced mitogenesis involves Akt-GSK3β-β-Catenin-TCF-4 signaling in human colorectal cancer. *Mol. Carcinog.* 53 (101), E169–E180. doi:10.1002/mc.22088

Bishnupuri, K. S., Sainathan, S. K., Ciorba, M. A., Houchen, C. W., and Dieckgraefe, B. K. (2022). Reg4 interacts with CD44 to regulate proliferation and stemness of colorectal and pancreatic cancer cells. *Mol. Cancer Res.* 20 (3), 387–399. doi:10.1158/1541-7786.MCR-21-0224

Chai, D., Du, H., Li, K., Zhang, X., Li, X., Zhao, X., et al. (2021). CDX2 and Reg IV expression and correlation in gastric cancer. *BMC Gastroenterol.* 21 (1), 92. doi:10.1186/s12876-021-01678-9

Chen, S., Gou, W. F., Zhao, S., Niu, Z. F., Zhao, Y., Takano, Y., et al. (2015). The role of the REG4 gene and its encoding product in ovarian epithelial carcinoma. *BMC Cancer* 15, 471. doi:10.1186/s12885-015-1435-2

Du, F., and Yao, Z. W. (2013). The expression patterns of Reg IV gene in normal rat reproduction system. *J. Exp. Zool. A Ecol. Genet. Physiol.* 319 (1), 32–38. doi:10.1002/jez.1771

Duan, Y., Hu, L., Liu, B., Yu, B., Li, J., Yan, M., et al. (2014). Tumor suppressor miR-24 restrains gastric cancer progression by downregulating RegIV. *Mol. Cancer* 13, 127. doi:10.1186/1476-4598-13-127

Eguchi, H., Ishikawa, O., Ohigashi, H., Takahashi, H., Yano, M., Nishiyama, K., et al. (2009). Serum REG4 level is a predictive biomarker for the response to preoperative chemoradiotherapy in patients with pancreatic cancer. *Pancreas* 38 (7), 791–798. doi:10.1097/MPA.0b013e3181ac5337

Gao, L., Wu, X., Zhang, L., Dai, Y., Zhu, Z., Zhi, Y., et al. (2021). REG4 is a Potential biomarker for radiochemotherapy sensitivity in colorectal cancer. *Onco. Targets. Ther.* 14, 1605–1611. doi:10.2147/OTT.S296031

Granlund, A. V., Beisvag, V., Torp, S. H., Flatberg, A., Kleveland, P. M., Ostvik, A. E., et al. (2011). Activation of REG family proteins in colitis. *Scand. J. Gastroenterol.* 46 (11), 1316–1323. doi:10.3109/00365521.2011.605463

Gu, Z., Rubin, M. A., Yang, Y., Deprimo, S. E., Zhao, H., Horvath, S., et al. (2005). Reg IV: A promising marker of hormone refractory metastatic prostate cancer. *Clin. Cancer Res.* 11 (6), 2237–2243. doi:10.1158/1078-0432.CCR-04-0356

Guo, Y., Xu, J., Li, N., Gao, F., and Huang, P. (2010). RegIV potentiates colorectal carcinoma cell migration and invasion via its CRD domain. *Cancer Genet. cytogenet.* 199 (1), 38–44. doi:10.1016/j.cancergencyto.2010.01.011

Hartupee, J. C., Zhang, H., Bonaldo, M. F., Soares, M. B., and Dieckgraefe, B. K. (2001). Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV. *Biochim. Biophys. Acta* 1518 (3), 287–293. doi:10.1016/s0167-4781(00)00284-0

Hayashi, T., Matsubara, A., Ohara, S., Mita, K., Hasegawa, Y., Usui, T., et al. (2009). Immunohistochemical analysis of Reg IV in urogenital organs: Frequent expression of Reg IV in prostate cancer and potential utility as serum tumor marker. *Oncol. Rep.* 21 (1), 95–100.

He, H. L., Lee, Y. E., Shiue, Y. L., Lee, S. W., Lin, L. C., Chen, T. J., et al. (2014). Overexpression of REG4 confers an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *J. Surg. Oncol.* 110 (8), 1002–1010. doi:10.1002/jso.23764

He, X. J., Jiang, X. T., Ma, Y. Y., Xia, Y. J., Wang, H. J., Guan, T. P., et al. (2012). REG4 contributes to the invasiveness of pancreatic cancer by upregulating MMP-7 and MMP-9. *Cancer Sci.* 103 (12), 2082–2091. doi:10.1111/cas.12018

Heiskala, K., and Andersson, L. C. (2013). Reg IV is differently expressed in enteroendocrine cells of human small intestine and colon. *Regul. Pept.* 183, 27–34. doi:10.1016/j.regpep.2013.03.007

Ho, M. R., Lou, Y. C., Wei, S. Y., Luo, S. C., Lin, W. C., Lyu, P. C., et al. (2010). Human RegIV protein adopts a typical C-type lectin fold but binds mannan with two calcium-independent sites. *J. Mol. Biol.* 402 (4), 682–695. doi:10.1016/j.jmb.2010.07.061

Hu, G., Shen, J., Cheng, L., Guo, C., Xu, X., Wang, F., et al. (2011). Reg4 protects against acinar cell necrosis in experimental pancreatitis. *Gut* 60 (6), 820–828. doi:10.1136/gut.2010.215178

Huang, J., Yang, Y., Yang, J., and Li, X. (2014a). Regenerating gene family member 4 promotes growth and migration of gastric cancer through protein kinase B pathway. *Int. J. Clin. Exp. Med.* 7 (9), 3037–3044.

Huang, Q., Chen, X., Lu, W., Lai, M., and Lu, B. (2014b). Expression of REG4 in ovarian mucinous tumors. *Appl. Immunohistochem. Mol. Morphol.* 22 (4), 295–301. doi:10.1097/PAI.0b013e3182936d8e

Hwang, J. H., Yoon, J., Cho, Y. H., Cha, P. H., Park, J. C., and Choi, K. Y. (2020). A mutant KRAS-induced factor REG4 promotes cancer stem cell properties via Wnt/β-catenin signaling. *Int. J. Cancer* 146 (10), 2877–2890. doi:10.1002/ijc.32728

Jin, J., Lv, H., Wu, J., Li, D., Chen, K., Zhang, F., et al. (2017). Regenerating family member 4 (Reg4) enhances 5-Fluorouracil resistance of gastric cancer through activating MAPK/Erk/Bim signaling pathway. *Med. Sci. Monit.* 23, 3715–3721. doi:10.12659/msm.903134

Kang, G., Oh, I., Pyo, J., Kang, D., and Son, B. (2021). Clinicopathological significance and prognostic implications of REG4 immunohistochemical expression in colorectal cancer. *Med. Kaunas.* 57 (9), 938. doi:10.3390/medicina57090938

Kaprio, T., Hagström, J., Mustonen, H., Koskensalo, S., Andersson, L. C., and Haglund, C. (2014). REG4 independently predicts better prognosis in non-mucinous colorectal cancer. *PLoS One* 9 (10), e109600. doi:10.1371/journal.pone.0109600

Katsuno, Y., Ehata, S., Yashiro, M., Yanagihara, K., Hirakawa, K., and Miyazono, K. (2012). Coordinated expression of REG4 and aldehyde dehydrogenase 1 regulating tumourigenic capacity of diffuse-type gastric carcinoma-initiating cells is inhibited by TGF-β. *J. Pathol.* 228 (3), 391–404. doi:10.1002/path.4020

Kawasaki, Y., Matsumura, K., Miyamoto, M., Tsuji, S., Okuno, M., Suda, S., et al. (2015). REG4 is a transcriptional target of GATA6 and is essential for colorectal tumorigenesis. *Sci. Rep.* 5, 14291. doi:10.1038/srep14291

Kobayashi, Y., Niwa, Y., Tajika, M., Kawai, H., Kondo, S., Hara, K., et al. (2010). Serum tumor antigen REG4 as a useful diagnostic biomarker in gastric cancer. *Hepatogastroenterology.* 57 (104), 1631–1634.

Kobunai, T., Watanabe, T., and Fukusato, T. (2011). REG4, NEIL2, and BIRC5 gene expression correlates with gamma-radiation sensitivity in patients with rectal cancer receiving radiotherapy. *Anticancer Res.* 31 (12), 4147–4153.

Koh, I., Nosaka, S., Sekine, M., Sugimoto, J., Hirata, E., and Kudo, Y. (2019). Regulation of REG4 expression and prediction of 5-Fluorouracil sensitivity by CDX2 in ovarian mucinous carcinoma. *Cancer Genomics Proteomics* 16 (6), 481–490. doi:10.21873/cgp.20151

Kuniyasu, H., Oue, N., Sasahira, T., Yi, L., Moriwaka, Y., Shimomoto, T., et al. (2009). Reg IV enhances peritoneal metastasis in gastric carcinomas. *Cell Prolif.* 42 (1), 110–121. doi:10.1111/j.1365-2184.2008.00577.x

Li, X. H., Zheng, Y., Zheng, H. C., Takahashi, H., Yang, X. H., Masuda, S., et al. (2010). REG IV overexpression in an early stage of colorectal carcinogenesis: An immunohistochemical study. *Histol. Histopathol.* 25 (4), 473–484. doi:10.14670/HH-25.473

Li, X. J., Zhu, F., Li, B., Zhang, D., and Liang, C. W. (2021). Recombinant human regenerating gene 4 attenuates the severity of osteoarthritis by promoting the proliferation of articular chondrocyte in an animal model. *Curr. Mol. Pharmacol.* 15, 693–699. doi:10.2174/1874467214666210901163144

Li, Y. G., Ji, D. F., Zhong, S., Zhu, J. X., Chen, S., and Hu, G. Y. (2011). Anti-tumor effects of proteoglycan from Phellinus linteus by immunomodulating and inhibiting Reg IV/EGFR/Akt signaling pathway in colorectal carcinoma. *Int. J. Biol. Macromol.* 48 (3), 511–517. doi:10.1016/j.ijbiomac.2011.01.014

Liu, C. M., Hsieh, C. L., He, Y. C., Lo, S. J., Liang, J. A., Hsieh, T. F., et al. (2013). *In vivo* targeting of ADAM9 gene expression using lentivirus-delivered shRNA suppresses prostate cancer growth by regulating REG4 dependent cell cycle progression. *PLoS One* 8 (1), e53795. doi:10.1371/journal.pone.0053795

Lu, S., Bevier, M., Huhn, S., Sainz, J., Lascorz, J., Pardini, B., et al. (2013). Genetic variants in C-type lectin genes are associated with colorectal cancer susceptibility and clinical outcome. *Int. J. Cancer* 133 (10), 2325–2333. doi:10.1002/ijc.28251

Ma, X., Wu, D., Zhou, S., Wan, F., Liu, H., Xu, X., et al. (2016). The pancreatic cancer secreted REG4 promotes macrophage polarization to M2 through EGFR/AKT/CREB pathway. *Oncol. Rep.* 35 (1), 189–196. doi:10.3892/or.2015.4357

Mitani, Y., Oue, N., Matsumura, S., Yoshida, K., Noguchi, T., Ito, M., et al. (2007). Reg IV is a serum biomarker for gastric cancer patients and predicts response to 5-fluorouracil-based chemotherapy. *Oncogene* 26 (30), 4383–4393. doi:10.1038/sj.onc.1210215

Miyagawa, K., Sakakura, C., Nakashima, S., Yoshikawa, T., Fukuda, K., Kin, S., et al. (2008). Overexpression of RegIV in peritoneal dissemination of gastric cancer and its potential as A novel marker for the detection of peritoneal micrometastasis. *Anticancer Res.* 28 (2B), 1169–1179.

Moon, J. H., Fujiwara, Y., Nakamura, Y., Okada, K., Hanada, H., Sakakura, C., et al. (2012). REGIV as a potential biomarker for peritoneal dissemination in gastric adenocarcinoma. *J. Surg. Oncol.* 105 (2), 189–194. doi:10.1002/jso.22021

Naito, Y., Oue, N., Hinoi, T., Sakamoto, N., Sentani, K., Ohdan, H., et al. (2012). Reg IV is a direct target of intestinal transcriptional factor CDX2 in gastric cancer. *PLoS One* 7 (11), e47545. doi:10.1371/journal.pone.0047545

Nakata, K., Nagai, E., Ohuchida, K., Aishima, S., Hayashi, A., Miyasaka, Y., et al. (2009). REG4 is associated with carcinogenesis in the 'intestinal' pathway of intraductal papillary mucinous neoplasms. *Mod. Pathol.* 22 (3), 460–468. doi:10.1038/modpathol.2008.205

Nanakin, A., Fukui, H., Fujii, S., Sekikawa, A., Kanda, N., Hisatsune, H., et al. (2007). Expression of the REG IV gene in ulcerative colitis. *Lab. Invest.* 87 (3), 304–314. doi:10.1038/labinvest.3700507

Nata, K., Liu, Y., Xu, L., Ikeda, T., Akiyama, T., Noguchi, N., et al. (2004). Molecular cloning, expression and chromosomal localization of a novel human REG family gene, REG III. *Gene* 340 (1), 161–170. doi:10.1016/j.gene.2004.06.010

Numata, M., Oshima, T., Yoshihara, K., Watanabe, T., Tsuchida, K., Tamagawa, H., et al. (2011). Relationship between RegIV gene expression to outcomes in colorectal cancer. *J. Surg. Oncol.* 104 (2), 205–209. doi:10.1002/jso.21906

Ohara, S., Oue, N., Matsubara, A., Mita, K., Hasegawa, Y., Hayashi, T., et al. (2008). Reg IV is an independent prognostic factor for relapse in patients with clinically localized prostate cancer. *Cancer Sci.* 99 (8), 1570–1577. doi:10.1111/j.1349-7006.2008.00846.x

Oue, N., Kuniyasu, H., Noguchi, T., Sentani, K., Ito, M., Tanaka, S., et al. (2007). Serum concentration of reg IV in patients with colorectal cancer: Overexpression and high serum levels of reg IV are associated with liver metastasis. *Oncology* 72 (5-6), 371–380. doi:10.1159/000113147

Oue, N., Mitani, Y., Aung, P. P., Sakakura, C., Takeshima, Y., Kaneko, M., et al. (2005). Expression and localization of Reg IV in human neoplastic and non-neoplastic tissues: Reg IV expression is associated with intestinal and neuroendocrine differentiation in gastric adenocarcinoma. *J. Pathol.* 207 (2), 185–198. doi:10.1002/path.1827

Oue, N., Noguchi, T., Anami, K., Sentani, K., Sakamoto, N., Uraoka, N., et al. (2011). Serum concentration and expression of Reg IV in patients with esophageal cancer: Age-related elevation of serum Reg IV concentration. *Oncol. Lett.* 2 (2), 235–239. doi:10.3892/ol.2011.239

Qi, H., Wei, J., Gao, Y., Yang, Y., Li, Y., Zhu, H., et al. (2020). Reg4 and complement factor D prevent the overgrowth of *E. coli* in the mouse gut. *Commun. Biol.* 3 (1), 483. doi:10.1038/s42003-020-01219-2

Rafa, L., Dessein, A. F., Devisme, L., Buob, D., Truant, S., Porchet, N., et al. (2010). REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int. J. Oncol.* 36 (3), 689–698. doi:10.3892/ijo_00000544

Sasahira, T., Oue, N., Kirita, T., Luo, Y., Bhawal, U. K., Fujii, K., et al. (2008). Reg IV expression is associated with cell growth and prognosis of adenoid cystic carcinoma in the salivary gland. *Histopathology* 53 (6), 667–675. doi:10.1111/j.1365-2559.2008.03188.x

Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S. I., Fumagalli, A., Lyubimova, A., et al. (2016). Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc. Natl. Acad. Sci. U. S. A.* 113 (37), E5399–E5407. doi:10.1073/pnas.1607327113

Saukkonen, K., Hagström, J., Mustonen, H., Lehtinen, L., Carpen, O., Andersson, L. C., et al. (2018). Prognostic and diagnostic value of REG4 serum and tissue expression in pancreatic ductal adenocarcinoma. *Tumour Biol.* 40 (3), 1010428318761494. doi:10.1177/1010428318761494

Sentani, K., Oue, N., Noguchi, T., Sakamoto, N., Matsusaki, K., and Yasui, W. (2010). Immunostaining of gastric cancer with neuroendocrine differentiation: Reg IV-positive neuroendocrine cells are associated with gastrin, serotonin, pancreatic polypeptide and somatostatin. *Pathol. Int.* 60 (4), 291–297. doi:10.1111/j.1440-1827.2010.02519.x

Sentani, K., Oue, N., Tashiro, T., Sakamoto, N., Nishisaka, T., Fukuhara, T., et al. (2008). Immunohistochemical staining of Reg IV and claudin-18 is useful in the diagnosis of gastrointestinal signet ring cell carcinoma. *Am. J. Surg. Pathol.* 32 (8), 1182–1189. doi:10.1097/PAS.0b013e318163a8f8

Sninsky, J. A., Bishnupuri, K. S., González, I., Trikalinos, N. A., Chen, L., and Dieckgraefe, B. K. (2021). Reg4 and its downstream transcriptional activator CD44ICD in stage II and III colorectal cancer. *Oncotarget* 12 (4), 278–291. doi:10.18632/oncotarget.27896

Sun, S., Hu, Z., Huang, S., Ye, X., Wang, J., Chang, J., et al. (2019). REG4 is an indicator for KRAS mutant lung adenocarcinoma with TTF-1 low expression. *J. Cancer Res. Clin. Oncol.* 145 (9), 2273–2283. doi:10.1007/s00432-019-02988-y

Takasawa, S., Tsuchida, C., Sakuramoto-Tsuchida, S., Takeda, M., Itaya-Hironaka, A., Yamauchi, A., et al. (2018). Expression of human REG family genes in inflammatory bowel disease and their molecular mechanism. *Immunol. Res.* 66 (6), 800–805. doi:10.1007/s12026-019-9067-2

Takayama, R., Nakagawa, H., Sawaki, A., Mizuno, N., Kawai, H., Tajika, M., et al. (2010). Serum tumor antigen REG4 as a diagnostic biomarker in pancreatic ductal adenocarcinoma. *J. Gastroenterol.* 45 (1), 52–59. doi:10.1007/s00535-009-0114-y

Takehara, A., Eguchi, H., Ohigashi, H., Ishikawa, O., Kasugai, T., Hosokawa, M., et al. (2006). Novel tumor marker REG4 detected in serum of patients with resectable pancreatic cancer and feasibility for antibody therapy targeting REG4. *Cancer Sci.* 97 (11), 1191–1197. doi:10.1111/j.1349-7006.2006.00297.x

Tamura, H., Ohtsuka, M., Washiro, M., Kimura, F., Shimizu, H., Yoshidome, H., et al. (2009). Reg IV expression and clinicopathologic features of gallbladder carcinoma. *Hum. Pathol.* 40 (12), 1686–1692. doi:10.1016/j.humpath.2009.06.001

Tao, H. Q., He, X. J., Ma, Y. Y., Wang, H. J., Xia, Y. J., Ye, Z. Y., et al. (2011). Evaluation of REG4 for early diagnosis and prognosis of gastric cancer. *Hum. Pathol.* 42 (10), 1401–1409. doi:10.1016/j.humpath.2010.08.023

Violette, S., Festor, E., Pandrea-Vasile, I., Mitchell, V., Adida, C., Dussaulx, E., et al. (2003). Reg IV, a new member of the regenerating gene family, is overexpressed in colorectal carcinomas. *Int. J. Cancer* 103 (2), 185–193. doi:10.1002/ijc.10788

Wang, F., Xu, L., Guo, C., Ke, A., Hu, G., Xu, X., et al. (2011). Identification of RegIV as a novel GLI1 target gene in human pancreatic cancer. *PLoS One* 6 (4), e18434. doi:10.1371/journal.pone.0018434

Wang, H., Hu, L., Zang, M., Zhang, B., Duan, Y., Fan, Z., et al. (2016). REG4 promotes peritoneal metastasis of gastric cancer through GPR37. *Oncotarget* 7 (19), 27874–27888. doi:10.18632/oncotarget.8442

Wang, Q., Deng, J., Yuan, J., Wang, L., Zhao, Z., He, S., et al. (2012). Oncogenic Reg IV is a novel prognostic marker for glioma patient survival. *Diagn. Pathol.* 7, 69. doi:10.1186/1746-1596-7-69

Xiang, L. W., Xue, H., Ha, M. W., Yu, D. Y., Xiao, L. J., and Zheng, H. (2022). The effects of REG4 expression on chemoresistance of ovarian cancer. *J. Obstet. Gynaecol. Tokyo. 1995.*, 1–9. doi:10.1080/01443615.2022.2106834

Xiao, Y., Lu, Y., Wang, Y., Yan, W., and Cai, W. (2019). Deficiency in intestinal epithelial Reg4 ameliorates intestinal inflammation and alters the colonic bacterial composition. *Mucosal Immunol.* 12 (4), 919–929. doi:10.1038/s41385-019-0161-5

Yamagishi, H., Fukui, H., Sekikawa, A., Kono, T., Fujii, S., Ichikawa, K., et al. (2009). Expression profile of REG family proteins REG ialpha and REG IV in advanced gastric cancer: Comparison with mucin phenotype and prognostic markers. *Mod. Pathol.* 22 (7), 906–913. doi:10.1038/modpathol.2009.41

Yang, L., Lan, S., Liu, J., and Yang, Z. (2016). Expression of MK-1 and RegIV and its clinicopathological significances in the benign and malignant lesions of gallbladder. *Diagn. Pathol.* 6, 100. doi:10.1186/1746-1596-6-100

Ying, L. S., Yu, J. L., Lu, X. X., and Ling, Z. Q. (2013). Enhanced RegIV expression predicts the intrinsic 5-fluorouracil (5-FU) resistance in advanced gastric cancer. *Dig. Dis. Sci.* 58 (2), 414–422. doi:10.1007/s10620-012-2381-3

Yonemura, Y., Takashima, T., Miwa, K., Miyazaki, I., Yamamoto, H., and Okamoto, H. (1984). Amelioration of diabetes mellitus in partially depancreatized rats by poly(ADP-ribose) synthetase inhibitors. Evidence of islet B-cell regeneration. *Diabetes* 33 (4), 401–404. doi:10.2337/diab.33.4.401

Zhang, H. B., Luo, H. C., Xin, X. J., and Zeng, A. Z. (2015). Up-regulated Reg proteins induced by Interleukin-22 treatment ameliorate acute liver injury in rat model. *Int. J. Clin. Exp. Med.* 8 (1), 1253–1258.

Zhang, J., Zhu, Z., Miao, Z., Huang, X., Sun, Z., Xu, H., et al. (2021). The clinical significance and mechanisms of REG4 in human cancers. *Front. Oncol.* 10, 559230. doi:10.3389/fonc.2020.559230

Zhang, N., Chai, D., Du, H., Li, K., Xie, W., Li, X., et al. (2018). Expression of Reg IV and SOX9 and their correlation in human gastric cancer. *BMC Cancer* 18, 344. doi:10.1186/s12885-018-4285-x

Zhang, X. Q., Yu, L. T., Du, P., Yin, T. Q., Zhang, Z. Y., Xu, Y., et al. (2019). Single-chain antibody against Reg4 suppresses gastric cancer cell growth and enhances 5-FU-induced cell death *in vitro*. *Anticancer. Agents Med. Chem.* 19 (5), 610–619. doi:10.2174/1871520619666181122104720

Zhang, Y., Lai, M., Gu, X., Luo, M., and Shao, L. (2003). Reg IV, a differentially expressed gene in colorectal adenoma. *Chin. Med. J.* 116 (6), 918–922.

Zhang, Y., Lai, M., Lv, B., Gu, X., Wang, H., Zhu, Y., et al. (2003). Overexpression of Reg IV in colorectal adenoma. *Cancer Lett.* 200 (1), 69–76. doi:10.1016/s0304-3835(03)00460-9

Zheng, H. C., Sugawara, A., Okamoto, H., Takasawa, S., Takahashi, H., Masuda, S., et al. (2011). Expression profile of the REG gene family in colorectal carcinoma. *J. Histochem. Cytochem.* 59 (1), 106–115. doi:10.1369/jhc.2010.956961

Zheng, H. C., Xu, X. Y., Yu, M., Takahashi, H., Masuda, S., and Takano, Y. (2010). The role of Reg IV gene and its encoding product in gastric carcinogenesis. *Hum. Pathol.* 41 (1), 59–69. doi:10.1016/j.humpath.2009.06.013

Zhou, W., Sun, M., Wang, D. L., Wang, Y., Jin, F., Zhang, Y. Y., et al. (2013). Silencing of RegIV by shRNA causes the loss of stemness properties of cancer stem cells in MKN45 gastric cancer cells. *Oncol. Rep.* 30 (6), 2685–2690. doi:10.3892/or.2013.2745

Zhu, X., Han, Y., Yuan, C., Tu, W., Qiu, G., Lu, S., et al. (2015). Overexpression of Reg4, alone or combined with MMP-7 overexpression, is predictive of poor prognosis in colorectal cancer. *Oncol. Rep.* 33 (1), 320–328. doi:10.3892/or.2014.3559

Glossary

ACCs adenoid cystic carcinomas
AP-1 activator protein-1
ATF2 activating transcription factor 2
Bcl-xl B cell lymphoma-extra large
Bcl-2 B cell lymphoma two
bFGF basic fibroblast growth factor
CDX2 caudal type homeobox two
CD44ICD CD44 intracytoplasmic domain
C-type lectin calcium-dependent lectin
DCS deep crypt secretory
EEC enteroendocrine cells
EGF epidermal growth factor
EGFR epidermal growth factor receptor
GATA6 GATA binding protein six
GLI1 GLI family zinc finger one;
GPR37 G protein-coupled receptor 37
GSK3β glycogen synthase kinase three β
HGF hepatocyte growth factor
IL interleukin
Klf4 Kruppel-like factor 4
Lgr5 leucine-rich repeat-containing G-protein coupled receptor five
miR MicroRNA
MMP-7 matrix metalloproteinase-7
MUC2 mucin two
RCT radiochemotherapy
REG regenerating islet-derived
REG4 Regenerating islet-derived four
rhREG4 recombinant human REG4
SCC squamous cell carcinoma
SOX2 SRY-box transcription factor 2
SOX9 SRY-box transcription factor 9
SP1 specificity protein one
SRCCs signet ring cell carcinomas
TCF-4 transcription factor 4
TGF transforming growth factor
TNF tumor necrosis factor
TTF-1 transcription factor-1;
UC ulcerative colitis
5-FU 5-fluorouracil

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fgene.2022.1006582

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Weiqiang Zhou,
Shenyang Medical College, China
Wang Hao,
Capital Medical University, China

*CORRESPONDENCE
Hua-Chuan Zheng,
Hua-Chuan.Zheng@hotmail.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

RECEIVED 29 July 2022
ACCEPTED 11 August 2022
PUBLISHED 16 September 2022

CITATION
Zheng H-C, Xue H, Zhang C-Y, Shi K-H and Zhang R (2022), The clinicopathological significances and related signal pathways of BTG3 mRNA expression in cancers: A bioinformatics analysis.
Front. Genet. 13:1006582.
doi: 10.3389/fgene.2022.1006582

The clinicopathological significances and related signal pathways of BTG3 mRNA expression in cancers: A bioinformatics analysis

Hua-Chuan Zheng[1]*, Hang Xue[1], Cong-Yu Zhang[2], Kai-Hang Shi[3] and Rui Zhang[4]

[1]Department of Oncology, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Cancer Center, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China, [3]Department of Dermatology, The Affiliated Hospital of Chengde Medical University, Chengde, China, [4]Department of Colorectal Surgery, Liaoning Cancer Hospital, Shenyang, China

B cell transposition gene 3 (BTG3) is reported to be a tumor suppressor and suppresses proliferation and cell cycle progression. This study aims to analyze the clinicopathological and prognostic significances, and signal pathways of *BTG3* mRNA expression in human beings through bioinformatics analysis. We analyzed *BTG3* expression using Oncomine, TCGA (the cancer genome atlas), Xiantao, UALCAN (The University of ALabama at Birmingham Cancer data analysis Portal) and Kaplan-Meier plotter databases. Down-regulated *BTG3* expression was observed in lung and breast cancers, compared with normal tissues ($p < 0.05$), but not for gastric and ovarian cancer ($p < 0.05$). The methylation of *BTG3* was shown to be adversely correlated with its mRNA expression ($p < 0.05$). *BTG3* expression was higher in gastric intestinal-type than diffuse-type carcinomas, G_1 than G_3 carcinomas ($p < 0.05$), in female than male cancer patients, $T_{1\text{-}2}$ than $T_{3\text{-}4}$, and adenocarcinoma than squamous cell carcinoma of lung cancer ($p < 0.05$), in invasive ductal than lobular carcinoma, N_0 than N_1 and N_3, TNBC (triple-negative breast cancer) than luminal and Her2+, and Her2+ than luminal cancer of breast cancer ($p < 0.05$), and G_3 than G_2 ovarian carcinoma ($p < 0.05$). *BTG3* expression was positively related to the survival rate of gastric and ovarian cancer patients ($p < 0.05$), but not for breast cancer ($p < 0.05$). KEGG and PPI (protein-protein interaction) analysis showed that the *BTG3* was involved in cell cycle and DNA replication, digestion and absorption of fat and protein, spliceosome and ribosome in cancer. *BTG3* expression was positively linked to carcinogenesis, histogenesis, and aggressive behaviors, and was employed to evaluate the prognosis of cancers by regulating cell cycle, metabolism, splicing and translation of RNA.

KEYWORDS

***BTG3*, bioinformatics analysis, carcinogenesis, aggressive behavior, prognosis**

Introduction

B cell transposition gene 3 (BTG3) belongs to a member of B-cell translocation gene family, and maps to human chromosome 21q21.1 (Guéhenneux et al., 1997). Its encoding protein has been reported to be a tumor suppressor in some malignancies, including gastric cancer, breast cancer, renal cell carcinoma, esophageal adenocarcinoma, hepatocellular carcinoma, lung cancer, ovarian cancer, and prostate cancer (Yu et al., 2008; Majid et al., 2009; Lin et al., 2012; Chen et al., 2013; Deng et al., 2013; Lv et al., 2013; Du et al., 2015; Gou et al., 2015; Ren et al., 2015). In the nuclear compartment, BTG3 protein can interact with E2F1 (E2F transcription factor 1), Smad8 receptor-regulated Smad transcription factor, and CCR4 (C-C motif chemokine receptor 4) transcription factor-associated protein Caf1 to suppress finally cell proliferation and cell cycle progression (Yoshida et al., 2001; Ou et al., 2007; Miyai et al., 2009). BTG3 maintains genomic stability by promoting Lys63-linked ubiquitination and CHK1 (checkpoint kinase 1) activation, whereas BTG3 can be phosphorylated and activated *via* its interaction with CHK1 as a positive feedback loop (Cheng et al., 2013). In the cytosolic compartment, BTG3 binds to and suppresses src, Akt and Ras/MAP kinase signaling (Rahmani, 2006; Cheng et al., 2015). Ma et al. (2020) demonstrated that the combination of hypoxia and BTG3 expression could induce radiation resistance, indicating an important role of BTG3 in hypoxia-induced radiation resistance of colorectal cancer cells. Cucurbitacin B inhibited cell proliferation and anti-apoptosis of colorectal cancer by the reactivation of BTG3 by promoter demethylation (Mao et al., 2019).

By postnatal 21 months, BTG3-deficiency might promote bone morphogenetic protein-induced ectopic bone formation and lung adenocarcinogenesis (Yoneda et al., 2009). BTG3 deficiency triggers acute cellular senescence *via* the Erk-AP-1-JMJD3-p16 pathway (Lin et al., 2012). MiR-142-5p strengthens cell growth and migration in renal cell carcinoma, while miR-93 desensitizes esophageal cancer to radiotherapy by targeting BTG3 (Cui et al., 2017; Liu et al., 2017). IASPP promotes miR-20a expression that restores cell invasion and cisplatin chemoresistance of cervical cancer cells by targeting BTG3 (Xiong et al., 2017). A body of evidence identified BTG3 as a direct downstream target of miR-519c-3p and miR-20b-5p, which promoted proliferation and migration in hepatocellular carcinoma and colorectal cancer cells, respectively (Peng et al., 2019; Wang et al., 2019). Enkhnaran et al. (2022) demonstrated that miR-106b-5p promoted cell proliferation and cell cycle and increased hepatocellular carcinoma cells' resistance to sorafenib through the BTG3/Bcl-xL/p27 signaling pathway.

In our previous work, BTG3 overexpression was demonstrated to reverse the aggressive phenotypes of gastric and colorectal cancer cells (Gou et al., 2015; Zheng et al., 2017). Here, we aimed to clarify the clinicopathological and prognostic significances, and related signal pathways of *BTG3* mRNA expression in cancers by a bioinformatics analysis.

Material and methods

Oncomine database analysis

The individual gene expression level of *BTG3* mRNA was analyzed using Oncomine (www.oncomine.org), a cancer microarray database and web-based data mining platform for a new discovery from genome-wide expression analyses. We compared the differences in *BTG3* mRNA levels between normal tissue and cancer. All data were log-transformed, the median centered per array, and the standard deviation normalized to one per array.

The cancer genome atlas (TCGA) database analysis

The expression data (RNA-seqV2) and clinicopathological data of gastric (n = 392), lung (n = 865), breast (n = 1,093), and ovarian (n = 304) cancer patients were downloaded from the TCGA database (https://cancergenome.nih.gov/abouttcga/overview) by TCGA-assembler in R software. We integrated the raw data, analyzed *BTG3* mRNA expression in the cancers, and compared it with clinicopathological and prognostic data of cancer patients. The means were compared with student t-test. Kaplan-Meier survival plots were generated with survival curves and compared by the log-rank statistic. Cox's proportional hazards model was employed for multivariate analysis. SPSS 17.0 software was employed to analyze all data. Two-sided $p < 0.05$ was considered statistically significant.

Kaplan-Meier (KM) plotter analysis

The prognostic significance of ***BTG3*** mRNA was also analyzed in gastric, lung, breast, and ovarian cancers using KM plotter (https://kmplot.com/analysis/).

The university of aLabama at birmingham cancer data analysis portal analysis

The expression and methylation of *BTG3* gene were analyzed using the UALCAN database (http://ualcan.path.uab.edu/). They were also compared with the clinicopathological and prognostic features of gastric, lung, breast, and ovarian cancers.

Xiantao analysis

The expression and methylation of *BTG3* gene were analyzed using the xiantao platform (https://www.xiantao.love/). Additionally, we discovered the differential and related genes

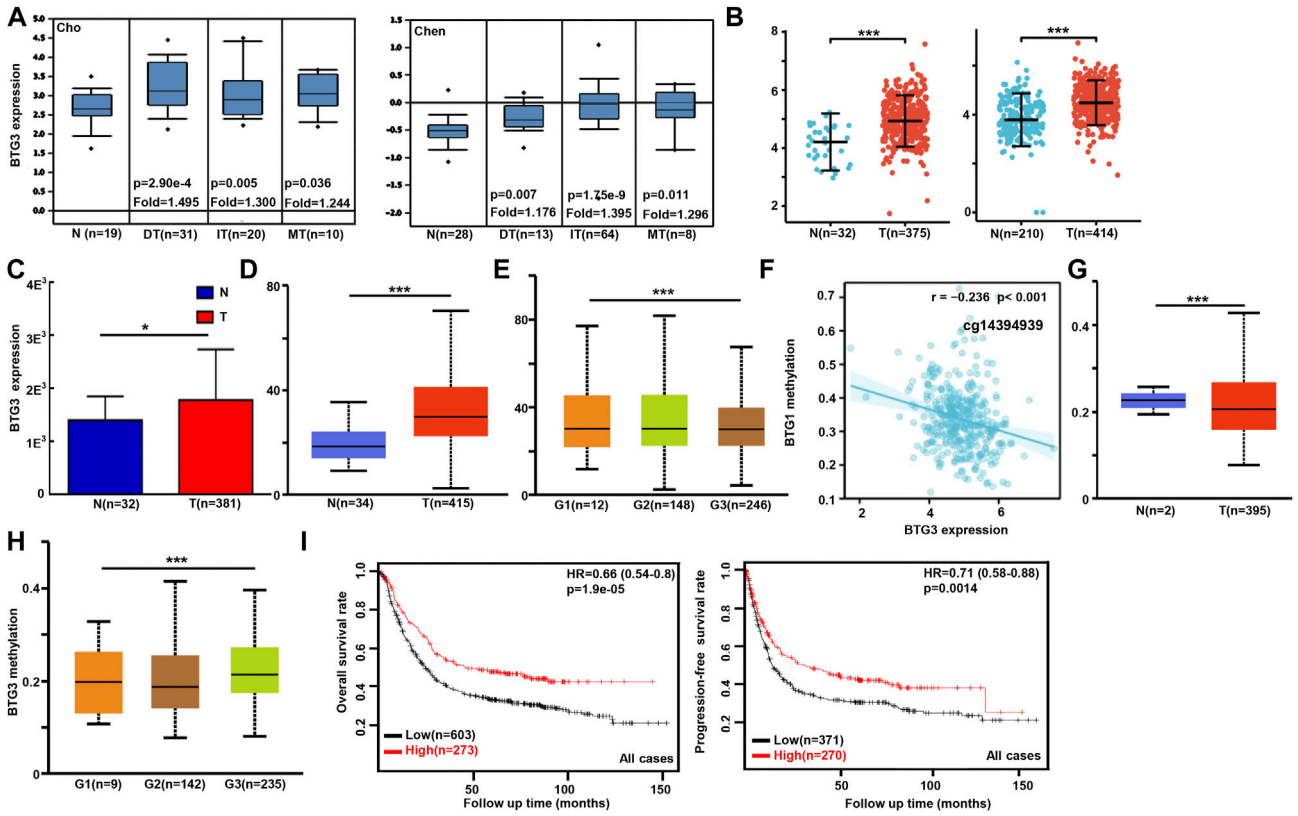


FIGURE 1
The clinicopathological and prognostic significances of *BTG3* mRNA expression in gastric cancer Cho's and Chen's datasets were used for bioinformatics analysis to explore *BTG3* expression in gastric cancer. A higher *BTG3* expression was detectable in gastric cancer than that in normal mucosa, even stratified into intestinal- (IT), diffuse- (DT), and mixed-type (MT) carcinomas by Lauren's classification [**(A)**, $p < 0.05$], in line with the findings from Xiantao **(B)**, TCGA **(C)** and UALCAN **(D)** databases ($p < 0.05$). UALCAN database showed that *BTG3* was more expressed in G_1 than G_3 carcinomas [**(E)**, $p < 0.05$]. The negative relationship between *BTG3* mRNA expression and methylation was found in gastric cancer using Xiantao database **(F)**. Its methylation was lower in gastric cancer than normal tissues in UALCAN **(G)**. We also compared *BTG3* methylation with histological grading of gastric cancer **(H)**. According to the data from Kaplan-Meier plotter, *BTG3* expression was positively related to both overall and progression-free survival rates of the patients with gastric cancer [**(I)**, $p < 0.05$)]. Note: N, normal; T, tumor; HR, hazard ratio.

using Xiantao. The differential genes were used to build the PPI (protein-protein interaction) network and identify the important hub genes. These genes were submitted to KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis in order to build signal pathways.

Results

The clinicopathological and prognostic significances of *BTG3* mRNA expression in gastric cancer

We used Cho's and Chen's datasets to perform bioinformatics analysis, and found that *BTG3* expression was higher in gastric cancer than in normal tissues, even stratified into intestinal-, diffuse-, and mixed-type carcinomas (Figure 1A, $p < 0.05$). In TCGA (Figure 1B), Xiantao (Figure 1C) and UALCAN (Figure 1D) data, it was the same for *BTG3* expression ($p < 0.05$). *BTG3* expression was higher in G_1 than G_3 carcinomas by UALCAN (Figure 1E, $p < 0.05$). There was a negative correlation between *BTG3* mRNA and methylation (cg23273752, cg12602426, cg01168851, cg04464940, cg14394939, and cg08875503) by Xiantao (Figure 1F, $p < 0.05$). *BTG3* methylation was lower in gastric cancer than in normal tissues (Figure 1G, $p < 0.05$), and G_1 than G_3 carcinoma (Figure 1H, $p < 0.05$) by UALCAN. According to Kaplan-Meier plotter, we found that a higher *BTG3* expression was positively correlated with overall and progression-free survival rates of all cancer patients, even stratified by gender and Her2+ expression (Figure 1I and Table 1, $p < 0.05$). The overall survival rate of the patients with intestinal-, diffuse-, or mixed-type carcinoma was higher in *BTG3*

TABLE 1 The prognostic significance of *BTG3* mRNA in gastric cancer.

Clinicopathological features	Overall survival		Progression-free survival	
	Hazard ratio	*p*	Hazard ratio	*p*
Sex				
Female	0.62 (0.41–0.93)	0.019	0.64 (0.41–0.99)	0.046
Male	0.64 (0.51–0.81)	0.00014	0.68 (0.53–0.87)	0.0018
T				
2	0.61 (0.38–0.97)	0.035	0.63 (0.38–1.03)	0.062
3	0.79 (0.53–1.17)	0.24	0.84 (0.59–1.2)	0.33
4	2.21 (0.94–5.17)	0.061	2.03 (0.93–4.41)	0.069
N				
0	2.22 (0.75–6.61)	0.14	2.23 (0.74–6.65)	0.14
1–3	0.71 (0.54–0.93)	0.012	0.8 (0.61–1.03)	0.081
1	0.69 (0.46–1.05)	0.079	0.68 (0.42–1.09)	0.1
2	0.7 (0.44–1.12)	0.14	0.76 (0.5–1.17)	0.21
3	2.07 (1.16–3.71)	0.013	1.58 (0.93–2.68)	0.088
M				
0	0.76 (0.57–1.02)	0.066	1.22 (0.91–1.64)	0.18
1	0.55 (0.29–1.06)	0.07		
TNM staging				
I	0.34 (0.11–1.08)	0.057	0.48 (0.16–1.45)	0.18
II	0.58 (0.28–1.17)	0.12	0.5 (0.23–1.09)	0.076
III	0.76 (0.56–1.03)	0.08	1.33 (0.88–2.01)	0.18
IV	0.71 (0.47–1.08)	0.11	1.23 (0.84–1.8)	0.29
Differentiation				
Well-differentiated	—	—	—	—
Moderately-differentiated	1.35 (0.68–2.69)	0.39	1.42 (0.75–2.67)	0.28
Poorly-differentiated	1.25 (0.81–1.92)	0.31	1.61 (1.02–2.54)	0.039
Lauren's classification				
Intestinal-type	0.66 (0.47–0.92)	0.013	0.73 (0.52–1.04)	0.084
Diffuse-type	0.62 (0.41–0.94)	0.022	0.7 (0.47–1.04)	0.073
Mixed-type	0.23 (0.06–0.81)	0.013	0.57 (0.2–1.59)	0.28
Her2 positivity				
–	0.63 (0.5–0.78)	4e–05	0.6 (0.45–0.81)	0.00058
+	1.31 (1.01–1.71)	0.043	1.51 (1.09–2.08)	0.012
Perforation				
—	1.38 (0.91–2.09)	0.12	1.42 (0.97–2.08)	0.07
Treatment				
Surgery alone	0.74 (0.54–1.02)	0.066	1.36 (1–1.87)	0.053
5-FU-based adjuvant	1.59 (1.08–2.35)	0.018	1.53 (1.08–2.17)	0.016
Other adjuvant	3.56 (0.82–15.42)	0.07	2.6 (1.17–5.78)	0.015

overexpression than in underexpression groups (Table 1, $p < 0.05$). There appeared to be a positive relationship between *BTG*3 expression and the overall survival rate of the patients with 5-FU-based adjuvant (Table 1, $p < 0.05$). It was the same for the progression-free survival in the patients with poorly-differentiated adenocarcinoma, 5-FU-based or other adjuvant treatment (Table 1, $p < 0.05$).

The clinicopathological and prognostic significances of *BTG3* mRNA expression in lung cancer

In Xiantao, *BTG3* expression was lower in lung cancer than in normal tissues (Figure 2A, $p < 0.05$). In TCGA database, BTG3 expression was higher in female than male cancer

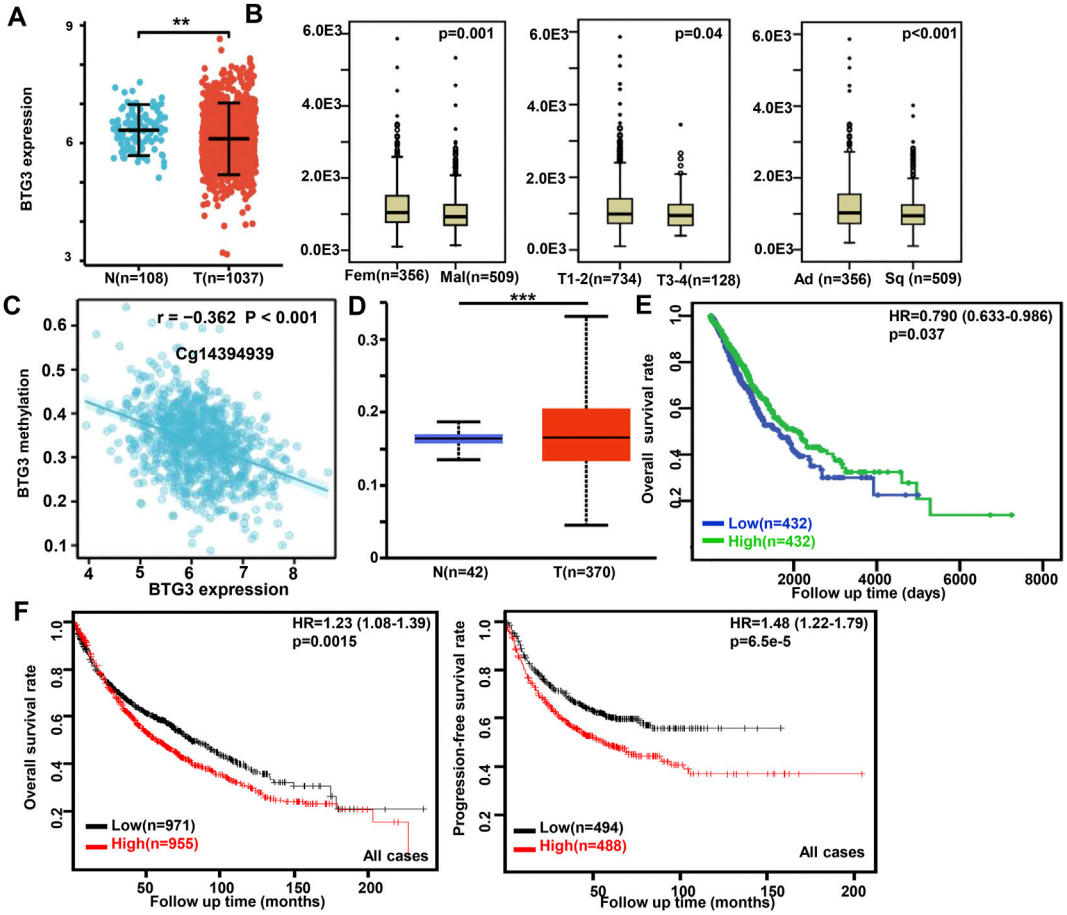


FIGURE 2
The clinicopathological and prognostic significances of *BTG3* mRNA expression in lung cancer Xiantao dataset was employed for bioinformatics analysis to analyze *BTG3* expression during lung carcinogenesis **(A)**. *BTG3* expression was compared with gender, histological subtyping, and T staging of the cancer patients by TCGA database **(B)**. The negative relationship between *BTG3* mRNA expression and methylation was analyzed in lung cancer using Xiantao database **(C)**. Its methylation was higher in lung cancer than normal tissues in UALCAN **(D)**. The correlation between *BTG3* expression and overall or post-progression survival rate of the patients with lung cancer was analyzed using TCGA database **(E)** and Kaplan-Meier plotter **(F)**. Note: N, normal; T, tumor; Ad, adenocarcinoma; Sq, squamous cell carcinoma; HR, hazard ratio.

patients, $T_{1\text{-}2}$ than $T_{3\text{-}4}$, and adenocarcinoma than squamous cell carcinoma patients (Figure 2B, $p < 0.05$). There was a negative correlation between *BTG3* mRNA and its methylation (cg14380517, cg10696191, cg03232933, cg05762769, cg14394939, and cg03232933) by xiantao (Figure 2C, $p < 0.05$). *BTG3* methylation was higher in lung cancer than in normal tissues (Figure 2D, $p < 0.05$). In TCGA, *BTG3* expression was positively associated with a high overall survival rate of cancer patients (Figure 2E, $p < 0.05$). Cox's risk proportional regression model indicated that T staging, lymph node status and *BTG3* expression were independent prognostic factors for lung cancer patients (Table 2, $p < 0.05$). According to Kaplan-Meier plotter, we found that a higher *BTG3* expression was negatively correlated with overall survival rates of all cancer patients, female or male patients, adenocarcinoma patients, N_1, M_0, Stage I and II cancer patients, smoking and non-smoking patients, or those with surgical margin negative (Figure 2F, $p < 0.05$). All, female T_1, T_2, or N_1 cancer patients with high *BTG3* expression showed a short progression-free survival time than those with its low expression ($p < 0.05$, data not shown). There appeared to be a negative relationship between *BTG3* expression and the progression-free survival rate of cancer patients with surgical margin negative or smoking cancer patients ($p < 0.05$, data not shown).

The clinicopathological and prognostic significances of *BTG3* mRNA expression in breast cancer

According to Xiantao (Figure 3A) and UALCAN (Figure 3B) databases, we found that *BTG3* expression was lower in breast

TABLE 2 Multivariate analysis of hazard factors of the prognosis of the patients with lung cancer.

Clinicopathological features	Hazard ratio (95% CI)	*p*
Gender (Female/male)	1.088 (0.853–1.387)	0.496
Stage T (T1-2/T3-4)	1.525 (1.089–2.137)	0.014
Lymph node status (−/+)	1.591 (1.223–2.070)	0.001
TNM staging (I-II/III-IV)	1.154 (0.817–1.631)	0.415
Histological classification (Ad/Sq)	0.996 (0.787–1.260)	0.971
BTG3 mRNA expression (low/high)	0.795 (0.634–0.997)	0.047

higher *BTG3* expression was negatively correlated with overall, progression-free, post-progression, and distant-metastasis-free survival rates of all cancer patients (Figure 3H, $p < 0.05$). There appeared to be a negative relationship between *BTG3* expression and the overall survival rate of patients with Luminal-B breast cancer ($p < 0.05$, data not shown). The relapse-free survival rate of the cancer patient with or without lymph node metastasis was lower in the groups of high *BTG3* expression than its low expression ($p < 0.05$, data not shown). A negative association between *BTG3* expression and relapse-free prognosis was

TABLE 3 The correlation between *BTG3* mRNA expression and clinicopathological characteristics of breast cancer.

Characteravbistic	Variables	Low expression	High expression	*p*
Age, *n* (%)	≤60	262 (24.2%)	339 (31.3%)	<0.001
	>60	279 (25.8%)	203 (18.7%)	
Race, *n* (%)	Asian	34 (3.4%)	26 (2.6%)	0.009
	Black or African American	72 (7.2%)	109 (11%)	
	White	389 (39.1%)	364 (36.6%)	
PR status, *n* (%)	Negative	104 (10.1%)	238 (23%)	<0.001
	Indeterminate	3 (0.3%)	1 (0.1%)	
	Positive	410 (39.7%)	278 (26.9%)	
ER status, *n* (%)	Negative	39 (3.8%)	201 (19.4%)	<0.001
	Indeterminate	1 (0.1%)	1 (0.1%)	
	Positive	477 (46.1%)	316 (30.5%)	
PAM50, *n* (%)	Normal	11 (1%)	29 (2.7%)	<0.001
	LumA	376 (34.7%)	186 (17.2%)	
	LumB	115 (10.6%)	89 (8.2%)	
	Her2	29 (2.7%)	53 (4.9%)	
	Basal	10 (0.9%)	185 (17.1%)	

ER, estrogen receptor; PR, progesterone receptor.

cancer than in normal tissues ($p < 0.05$). TCGA database showed a higher *BTG3* expression in invasive ductal than lobular carcinoma (Figure 3C, $p < 0.05$). It was higher in N_0 than N_1 and N_3, TNBC (triple-negative breast cancer) than luminal and Her2+, and Her2+ than Luminal cancer patients by UALCAN (Figure 3D, $p < 0.05$). As summarized in Table 3, BTG3 mRNA expression was negatively associated with elder age, non-Asian race, non-TNBC, ER positivity, and PR positivity ($p < 0.05$). There was a negative correlation between *BTG3* mRNA and its methylation (cg14380517, cg10696191, cg03232933, cg05762769, cg27075724, cg02652260, cg23273752, cg12602426, cg01168851, cg20227212, cg04464940, and cg14394939) by Xiantao (Figure 3E, $p < 0.05$). *BTG3* methylation was higher in breast cancer than normal tissues (Figure 3F, $p < 0.05$), N_3 than N_0, Luminal and Her2+ than TNBC, Her2+ than Luminal cancer patients by UALCAN (Figure 3G, $p < 0.05$). According to Kaplan-Meier plotter, we found that a observed in Luminal-B cancer patients ($p < 0.05$, data not shown). ER (estrogen receptor)- positive or Her2-negative cancer patients with high *BTG3* expression showed a shorter overall survival time than those with its low expression ($p < 0.05$, data not shown).

The clinicopathological and prognostic significances of *BTG3* mRNA expression in ovarian cancer

We performed bioinformatics analysis of *BTG3* expression in ovarian cancer using Bonome's, Hendrix's, Lu's, Welsh's, and TCGA's datasets. *BTG3* expression was higher in ovarian cancer than normal mucosa regardless of histological subtyping (Figure 4A, $p < 0.05$). It was the same for Xiantao data (Figure 4B, $p < 0.05$). *BTG3* expression was higher in G_3 than G_2 carcinoma patients by UALCAN (Figure 4C, $p < 0.05$). The

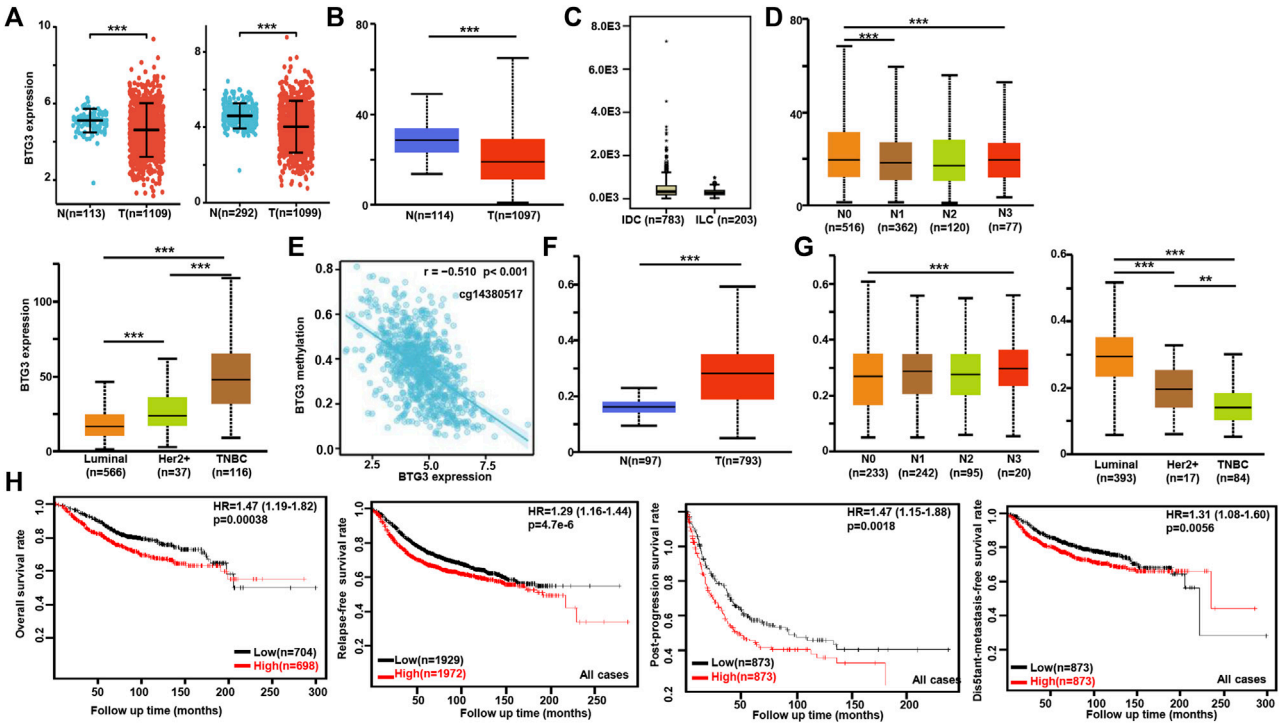


FIGURE 3
The clinicopathological and prognostic significances of *BTG3* mRNA expression in breast cancer. According to Xiantao **(A)** and UALCAN **(B)** databases, *BTG3* hypoexpression was detectable in breast cancer, compared with breast normal tissue ($p < 0.05$). TCGA database showed a higher *BTG3* expression in invasive ductal (IDC) than lobular (ILC) carcinoma [**(C)**, $p < 0.05$]. UALCAN database was also used to compare *BTG3* expression with N stage and molecular subtyping of breast cancer **(D)**. The negative relationship between *BTG3* mRNA expression and methylation was analyzed in breast cancer using Xiantao database **(E)**. Its methylation was higher in breast cancer than normal tissues **(F)**, and compared with N stage and molecular subtyping **(G)** in UALCAN. The correlation between *BTG3* expression and overall, post-progression, distant-metastasis-free or relapse-free survival rate was analyzed in the patients with breast cancer using Kaplan-Meier plotter **(H)**. Note: N, normal tissue; T, tumor; TNBC, triple-negative breast cancer; HR, hazard ratio.

data from Kaplan-Meier plotter showed a positive relationship between *BTG3* expression and the overall survival rate of the ovarian cancer patients with paclitaxel treatment (Figure 4D, $p < 0.05$). A positive correlation between BTG3 expression and the post-progression survival rate was observed in all ovarian cancer patients, or $G_{1\text{-}3}$, $G_{2\text{-}3}$, $Grade_3$ or suboptimal cancer patients (Figure 4D, $p < 0.05$).

The *BTG3*-related genes and pathways in cancers

On the Xiantao platform, we found the differential genes between low and high expression groups of *BTG3* mRNA in cancers. KEGG analysis showed that the top signal pathways of the differential genes included cell cycle, calcium and p53 signal pathway, pancreatic and insulin secretion, fat digestion and absorption, DNA replication, mismatch repair and homologous recombination in gastric cancer, platelet activation and coagulation, digestion and absorption of protein and fat, metabolism of arachidonic and linoleic acids in lung cancer, cell cycle, salivary and insulin secretion, DNA replication, and ovarian steroidogenesis in breast cancer, and PI3K/Akt signal pathway, focal adhesion and ECM-receptor interaction in ovarian cancer (Figure 5A). In addition, the STRING was used to identify the PPI pairs and the cytoscape to find out the top 10 nodes ranked by degree (Figure 5B). The top hub genes mainly contained replication protein, DNA replication helicase, replication factor, WRN RecQ like helicase and exonucleases in gastric cancer, Apolipoproteins, lipase C, phospholipase A2, and cholesteryl ester transfer protein in lung cancer, minichromosome maintenance protein in breast cancer, and ribosomal proteins in ovarian cancer.

According to the Xiantao database, the *BTG3*-correlated genes in cancers were analyzed and subjected to the KEGG analysis (Figure 6). The *BTG3*-correlated genes were involved in RNA transport, splicing and degradation, DNA replication and cell cycle, proteasomal degradation for gastric cancer, cell cycle, DNA replication, and mismatch repair, TNF and NF-κB signal pathways for lung cancer, ribosome and spliceosome,

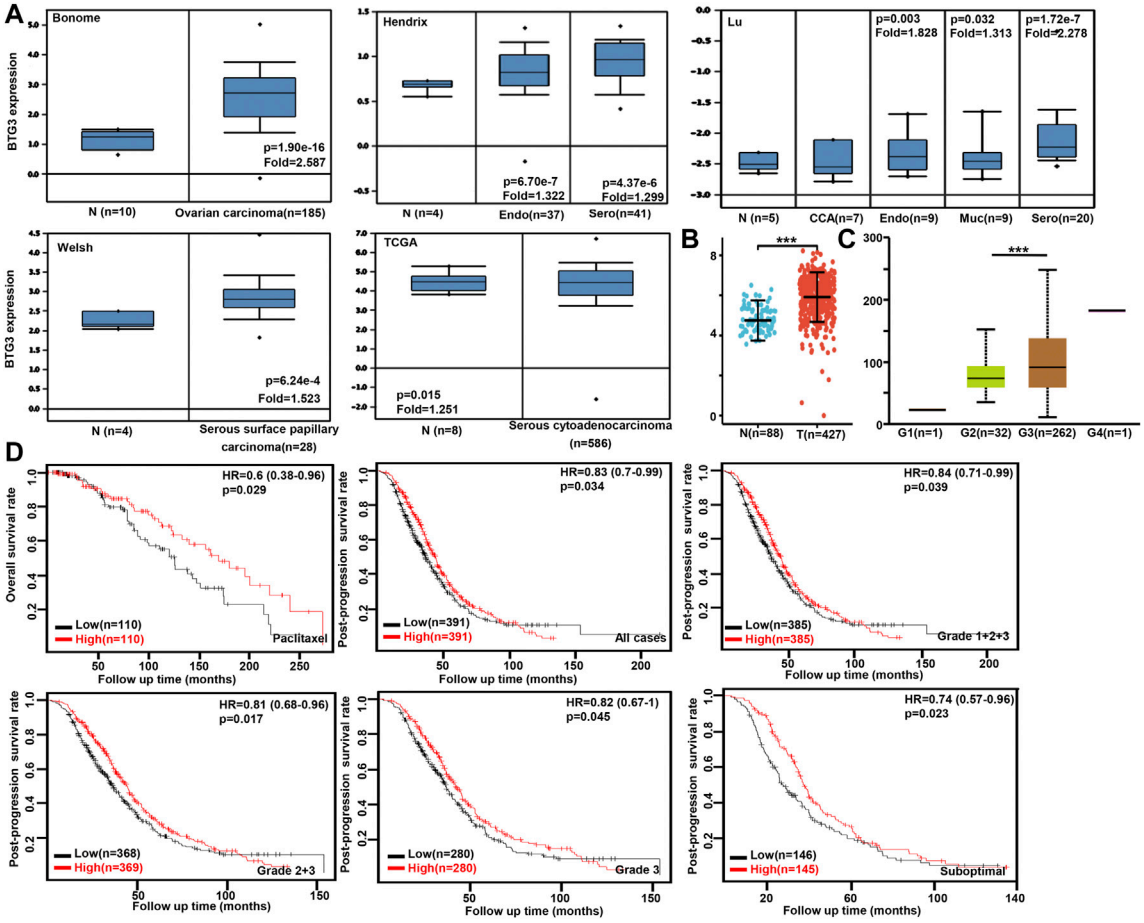


FIGURE 4
The clinicopathological and prognostic significances of *BTG3* mRNA expression in ovarian cancer Oncomine **(A)** and Xiantao **(B)** datasets were employed for bioinformatics analysis to observe BTG3 expression in ovarian cancer. A lower *BTG3* expression was detectable in ovary than that in ovarian carcinoma, clear cell adenocarcinoma (CCA), endometriod (Endo), mucinous (Muc) and serous (Sero) adenocarcinoma ($p < 0.05$). *BTG3* expression was compared with histological grading of ovarian cancer **(C)**. The correlation between *BTG3* expression and overall, or post-progression survival rate was analyzed in the patients with ovarian cancer using Kaplan-Meier plotter, even stratified by different clinicopathological parameters [**(D)**, $p < 0.05$]. Note: N, normal tissue; T, tumor; HR, hazard ratio.

and metabolism of amino acids for breast cancer, neural diseases, viral infection, oxidative phosphorylation for ovarian cancer.

Discussion

BTG3 overexpression suppressed the proliferation and invasion of epithelial ovarian and colorectal cancer cells by weakening Akt/GSK3β/β-catenin signaling (Mao et al., 2016; An et al., 2017). Lv et al. (2018) found that *BTG3* knockdown promoted cell proliferation, migration, invasion, relieved G_2 arrest, and inhibited apoptosis in colorectal cancer cells with PAK2 (p21 activated kinase 2), RPS6KA5 (ribosomal protein S6 kinase A5), YWHAB (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein beta), and STAT3 up-regulated and RAP1A (ras-related protein rap-1A), DUSP6 (dual specificity phosphatase 6), and STAT (signal transducer and activator of transcription) 1 down-regulated. Our group demonstrated that *BTG3* expression inhibited proliferation, tumor growth, migration and invasion, and induced autophagy, apoptosis, and chemosensitivity to cisplatin, MG132 (proteasome inhibitor), paclitaxel, and SAHA (histone deacetylase inhibitor) in gastric and colorectal cancer cells (Gou et al., 2015; Zheng et al., 2017). In esophageal adenocarcinoma cells, *BTG3* upregulation suppressed the proliferation and invasion (Du et al., 2015). Lv et al. (2013) found that *BTG3* suppressed proliferation, invasion and induced G_1/S cycle arrest of hepatocellular carcinoma cells. Reportedly, iASPP promoted epithelial-

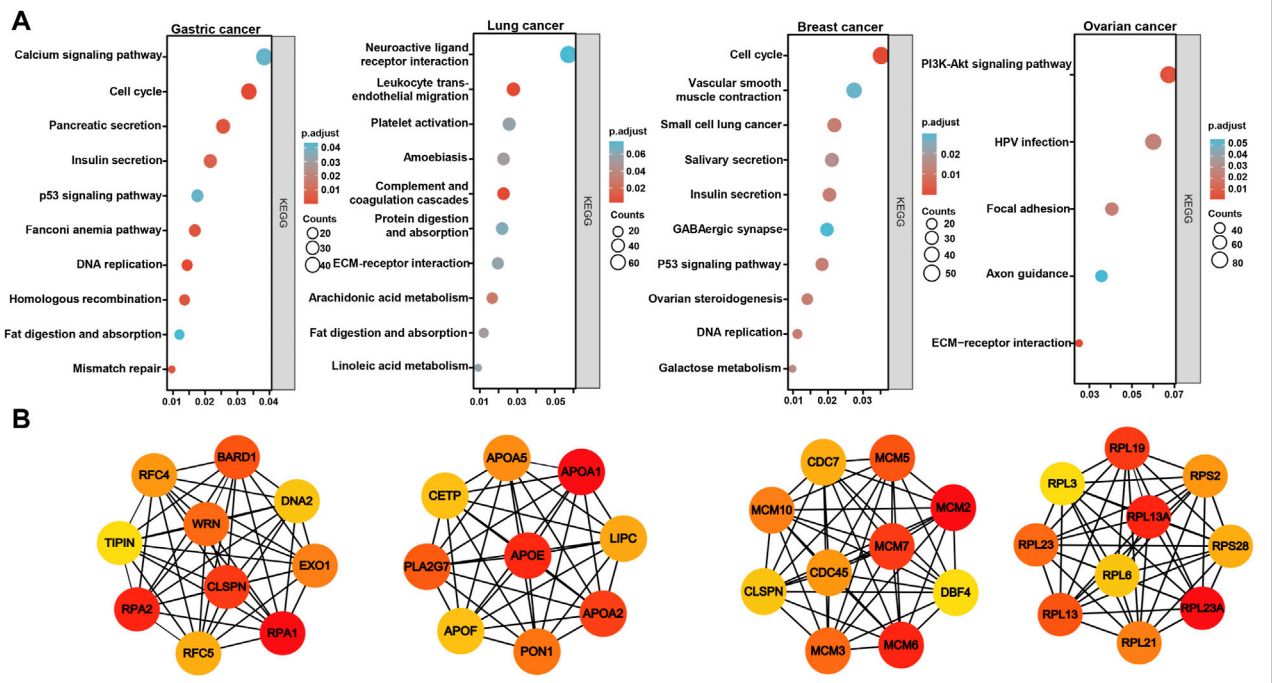


FIGURE 5
The differential genes and related signal pathways about *BTG3* in cancersn The differential genes of *BTG3* were subjected to the signal pathway analysis using KEGG **(A)**. STRING was used to identify the protein-protein interaction network of differential genes about *BTG3* in cancers, and Cytoscape was employed to find out the top 10 hub nodes ranked by degree **(B)**.

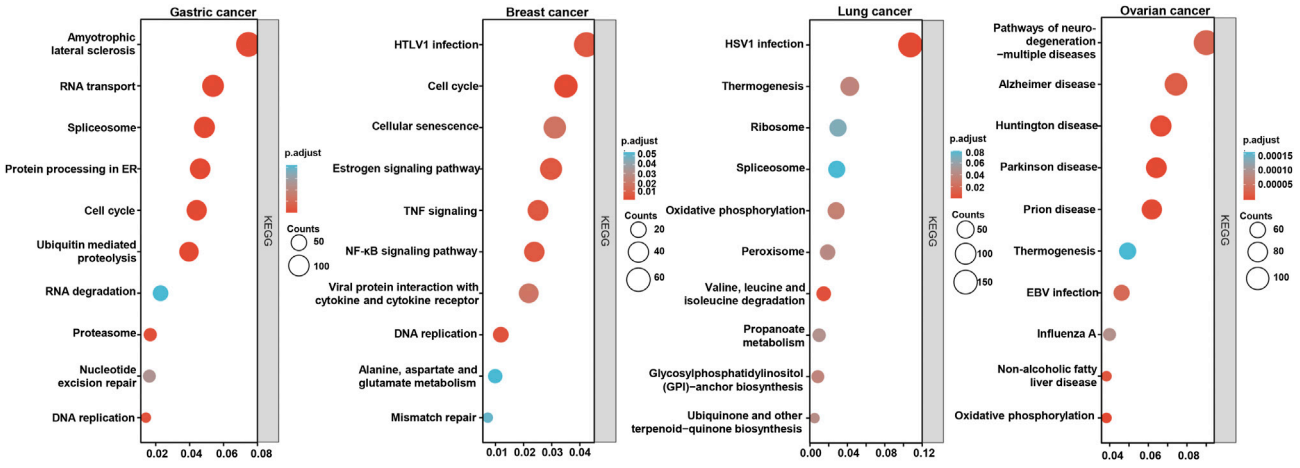


FIGURE 6
The *BTG3*-related signal pathways in cancers The top *BTG3*-related genes were screened and classified into the signal pathways of KEGG using Xiantao database.

mesenchymal transition, and conferred cisplatin resistance in cervical cancer *via* miR-20a- FBXL5/BTG3 signaling (Xiong et al., 2017). Yanagida et al. (2013) demonstrated that the antisense transcript of *BTG3* gene (ASBEL) down-regulated *BTG3* protein and promoted proliferation and tumorigenicity of ovarian clear cell carcinoma. ASBEL knockdown

functioned as a tumor suppressive role in breast cancer cells by up-regulating *BTG3* (Xia et al., 2017). *BTG3* anti-sense transcript mediated the down-regulation of ATF3 expression, which was essential for the proliferation and tumorigenicity of colon cancer cells (Taniue et al., 2016). In combination with these findings, it was suggested that *BTG3* might be employed as a molecular target of cancer gene therapy because of its inhibitory effects on aggressive phenotypes of cancer cells.

A body of evidence indicates that *BTG3* expression is down-regulated in gastric cancer (Gou et al., 2015; Ren et al., 2015), esophageal adenocarcinoma (Du et al., 2015), hepatocellular carcinoma (Lv et al., 2013), lung cancer (Chen et al., 2013), ovarian cancer (Deng et al., 2013), renal cell carcinoma (Majid et al., 2009), and colorectal cancer (Xiong et al., 2017) due to its promoter methylation (Yu et al., 2008; Majid et al., 2009; Lv et al., 2013; Gou et al., 2015), in line with our findings about lung and breast cancers. Additionally, *BTG3* methylation was negatively correlated with its mRNA expression, and was higher in cancer than in normal tissues of the stomach, lung, and breast. In breast and gastric cancer, the correlation between *BTG3* mRNA and aggressive behaviors was the opposite to that between *BTG3* methylation and them. These findings suggested that *BTG3* hypoexpression might be due to its promoter methylation. In contrast, our results showed *BTG3* mRNA overexpression in gastric and ovarian cancers. The controversial findings might be explained by different approaches: previous reports from immunohistochemistry, Western blot or RT-PCR, but the present study from the cDNA chip or transcriptomics.

In addition, BTG3 mRNA expression was positively linked to the differentiation of gastric cancer, in line with our previous report about gastric cancer tissues (Gou et al., 2015). *BTG3* was also reported to induce the differentiation of gastric and colorectal cancer cells, evidenced by a higher level of alkaline phosphatase (Gou et al., 2015; Zheng et al., 2017). These findings suggested that *BTG3* mRNA expression might underlie the molecular mechanisms of gastric cancer differentiation. In addition, *BTG3* expression was found to be negatively associated with the tumor size of lung cancer. Pulmonary adenocarcinoma patients had a higher *BTG3* mRNA expression than squamous cell carcinoma patients. A higher *BTG3* mRNA expression was seen in invasive ductal than lobular carcinomas and negatively correlated with N staging and favorable molecular subtypes (Luminal-type), suggesting that *BTG3* might be involved in the progression of breast cancer, and underlay the mechanisms of molecular subtyping. These results suggested that *BTG3* mRNA was employed to indicate the aggressive behaviors of lung cancer, and histogenesis of lung and breast cancers.

The prognostic significance of *BTG3* expression was analyzed, but controversial. Ren et al. (2015) found that *BTG3* expression was positively correlated with distant metastasis of gastric cancer, and the cancer patients with lower *BTG3* expression had a shorter overall survival time. Deng et al. (2013) demonstrated that *BTG3* expression was negatively associated with a higher incidence of metastasis, a better differentiation, longer disease-free time, and overall survival time of epithelial ovarian cancer as an independent factor of prognosis. However, there was no relationship between *BTG3* protein expression and overall survival rates of the patients with gastric (Gou et al., 2015) or colorectal (Zheng et al., 2017) cancer. In the present study, the positive correlation between *BTG3* mRNA expression and survival rate was seen in gastric and breast cancer patients, but not for lung and ovarian cancers according to Kaplan-Meier plotter. TCGA database showed that *BTG3* mRNA expression was an independent factor for favorable overall survival of lung cancer patients. *BTG3* mRNA was documented to indicate the adverse prognosis for pediatric T-cell acute lymphoblastic leukemia (Gottardo et al., 2007). Therefore, we concluded that the prognostic significance of *BTG3* expression was dependent on cancer type, pathological grouping, distinct methodologies, and different databases. Therefore, it should be careful to employ *BTG3* mRNA as a prognostic marker in clinicopathological practice.

KEGG analysis demonstrated that the *BTG3*-related pathways included cell cycle, pancreatic and insulin secretion, fat digestion and absorption, DNA replication, mismatch repair and homologous recombination in gastric cancer, platelet activation and coagulation, digestion and absorption of protein and fat, metabolism of arachidonic and linoleic acids in lung cancer, cell cycle, salivary and insulin secretion, and DNA replication in breast cancer. The top hub genes mainly contained replication protein and related enzymes in gastric cancer, the key enzymes and proteins for fat metabolism in lung cancer, minichromosome maintenance protein in breast cancer, and ribosomal proteins in ovarian cancer. Therefore, we speculate that *BTG3* might play important role in the cell cycle, fat digestion and metabolism, and protein biosynthesis, which be deeply investigated in the future.

In summary, *BTG3* mRNA might underlie the molecular mechanisms of the histogenesis of gastric, lung, and breast cancers. The paradoxical results about the prognostic significances of *BTG3* mRNA might result from tissue specificity, distinct grouping, and different data sources.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

Conception and design: H-CZ. Collection and assembly of data: HX and C-YZ. Data analysis and interpretation: K-HS and RZ. Manuscript writing: H-CZ. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (21377772D), and National Natural Scientific Foundation of China (81672700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

An, Q., Zhou, Y., Han, C., Zhou, Y., Li, F., and Li, D. (2017). BTG3 overexpression suppresses the proliferation and invasion in epithelial ovarian cancer cell by regulating AKT/GSK3β/β-catenin signaling. *Reprod. Sci.* 24 (10), 1462–1468. doi:10.1177/1933719117691143

Chen, X., Chen, G., Cao, X., Zhou, Y., Yang, T., and Wei, S. (2013). Downregulation of BTG3 in non-small cell lung cancer. *Biochem. Biophys. Res. Commun.* 437 (1), 173–178. doi:10.1016/j.bbrc.2013.06.062

Cheng, Y. C., Chen, P. H., Chiang, H. Y., Suen, C. S., Hwang, M. J., Lin, T. Y., et al. (2015). Candidate tumor suppressor B-cell translocation gene 3 impedes neoplastic progression by suppression of AKT. *Cell Death Dis.* 6 (1), e1584. doi:10.1038/cddis.2014.550

Cheng, Y. C., Lin, T. Y., and Shieh, S. Y. (2013). Candidate tumor suppressor BTG3 maintains genomic stability by promoting Lys63-linked ubiquitination and activation of the checkpoint kinase CHK1. *Proc. Natl. Acad. Sci. U. S. A.* 110 (15), 5993–5998. doi:10.1073/pnas.1220635110

Cui, H., Zhang, S., Zhou, H., and Guo, L. (2017). Direct downregulation of B-cell translocation gene 3 by microRNA-93 is required for desensitizing esophageal cancer to radiotherapy. *Dig. Dis. Sci.* 62 (8), 1995–2003. doi:10.1007/s10620-017-4579-x

Deng, B., Zhao, Y., Gou, W., Chen, S., Mao, X., Takano, Y., et al. (2013). Decreased expression of BTG3 was linked to carcinogenesis, aggressiveness, and prognosis of ovarian carcinoma. *Tumour Biol.* 34 (5), 2617–2624. doi:10.1007/s13277-013-0811-2

Du, Y., Liu, P., Zang, W., Wang, Y., Chen, X., Li, M., et al. (2015). BTG3 upregulation induces cell apoptosis and suppresses invasion in esophageal adenocarcinoma. *Mol. Cell. Biochem.* 404 (1-2), 31–38. doi:10.1007/s11010-015-2363-9

Enkhnaran, B., Zhang, G. C., Zhang, N. P., Liu, H. N., Wu, H., Xuan, S., et al. (2022). microRNA-106b-5p promotes cell growth and sensitizes chemosensitivity to sorafenib by targeting the BTG3/Bcl-xL/p27 signaling pathway in hepatocellular carcinoma. *J. Oncol.* 2022, 1971559. doi:10.1155/2022/1971559

Gottardo, N. G., Hoffmann, K., Beesley, A. H., Freitas, J. R., Firth, M. J., Perera, K. U., et al. (2007). Identification of novel molecular prognostic markers for paediatric T-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* 137 (4), 319–328. doi:10.1111/j.1365-2141.2007.06576.x

Gou, W. F., Yang, X. F., Shen, D. F., Zhao, S., Liu, Y. P., Sun, H. Z., et al. (2015). The roles of BTG3 expression in gastric cancer: A potential marker for carcinogenesis and a target molecule for gene therapy. *Oncotarget* 6 (23), 19841–19867. doi:10.18632/oncotarget.3734

Guéhenneux, F., Duret, L., Callanan, M. B., Bouhas, R., Hayette, S., Berthet, C., et al. (1997). Cloning of the mouse BTG3gene and definition of a new gene family (the BTG family) involved in the negative control of the cell cycle. *Leukemia* 11 (3), 370–375. doi:10.1038/sj.leu.2400599

Lin, T. Y., Cheng, Y. C., Yang, H. C., Lin, W. C., Wang, C. C., Lai, P. L., et al. (2012). Loss of the candidate tumor suppressor BTG3 triggers acute cellular senescence via the ERK- JMJD3-p16 (INK4a) signaling axis. *Oncogene* 31, 3287–3297. doi:10.1038/onc.2011.491

Liu, L., Liu, S., Duan, Q., Chen, L., Wu, T., Qian, H., et al. (2017). MicroRNA-142-5p promotes cell growth and migration in renal cell carcinoma by targeting BTG3. *Am. J. Transl. Res.* 9 (5), 2394–2402.

Lv, C., Wang, H., Tong, Y., Yin, H., Wang, D., Yan, Z., et al. (2018). The function of BTG3 in colorectal cancer cells and its possible signaling pathway. *J. Cancer Res. Clin. Oncol.* 144 (2), 295–308. doi:10.1007/s00432-017-2561-9

Lv, Z., Zou, H., Peng, K., Wang, J., Ding, Y., Li, Y., et al. (2013). The suppressive role and aberrent promoter methylation of BTG3 in the progression of hepatocellular carcinoma. *PLoS One* 8 (10), e77473. doi:10.1371/journal.pone.0077473

Ma, D., Gao, X., Tao, J., Yu, H., and Chai, Z. (2020). Hypoxia-induced downregulation of B-cell translocation gene 3 confers resistance to radiation therapy of colorectal cancer. *J. Cancer Res. Clin. Oncol.* 146 (10), 2509–2517. doi:10.1007/s00432-020-03307-6

Majid, S., Dar, A. A., Ahmad, A. E., Hirata, H., Kawakami, K., Shahryari, V., et al. (2009). BTG3 tumor suppressor gene promoter demethylation, histone modification and cell cycle arrest by genistein in renal cancer. *Carcinogenesis* 30 (4), 662–670. doi:10.1093/carcin/bgp042

Mao, D., Liu, A. H., Wang, Z. P., Zhang, X. W., and Lu, H. (2019). Cucurbitacin B inhibits cell proliferation and induces cell apoptosis in colorectal cancer by modulating methylation status of BTG3. *Neoplasma* 66 (4), 593–602. doi:10.4149/neo_2018_180929N729

Mao, D., Qiao, L., Lu, H., and Feng, Y. (2016). B-cell translocation gene 3 overexpression inhibits proliferation and invasion of colorectal cancer SW480 cells via Wnt/β-catenin signaling pathway. *Neoplasma* 63 (5), 705–716. doi:10.4149/neo_2016_507

Miyai, K., Yoneda, M., Hasegawa, U., Toita, S., Izu, Y., Hemmi, H., et al. (2009). ANA deficiency enhances bone morphogenetic protein-induced ectopic bone formation via transcriptional events. *J. Biol. Chem.* 284 (16), 10593–10600. doi:10.1074/jbc.M807677200

Ou, Y. H., Chung, P. H., Hsu, F. F., Sun, T. P., Chang, W. Y., and Shieh, S. Y. (2007). The candidate tumor suppressor BTG3 is a transcriptional target of p53 that inhibits E2F1. *EMBO J.* 26 (27), 3968–3980. doi:10.1038/sj.emboj.7601825

Peng, L., Li, S., Li, Y., Wan, M., Fang, X., Zhao, Y., et al. (2019). Regulation of BTG3 by microRNA-20b-5p in non-small cell lung cancer. *Oncol. Lett.* 18 (1), 137–144. doi:10.3892/ol.2019.10333

Rahmani, Z. (2006). APRO4 negatively regulates Src tyrosine kinase activity in PC12 cells. *J. Cell Sci.* 119 (4), 646–658. doi:10.1242/jcs.02778

Ren, X. L., Zhu, X. H., Li, X. M., Li, Y. L., Wang, J. M., Wu, P. X., et al. (2015). Down-regulation of BTG3 promotes cell proliferation, migration and invasion and predicts survival in gastric cancer. *J. Cancer Res. Clin. Oncol.* 141 (3), 397–405. doi:10.1007/s00432-014-1826-9

Taniue, K., Kurimoto, A., Takeda, Y., Nagashima, T., Okada-Hatakeyama, M., Katou, Y., et al. (2016). ASBEL-TCF3 complex is required for the tumorigenicity of colorectal cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 113 (45), 12739–12744. doi:10.1073/pnas.1605938113

Wang, L., Mo, H., Jiang, Y., Wang, Y., Sun, L., Yao, B., et al. (2019). MicroRNA-519c-3p promotes tumor growth and metastasis of hepatocellular carcinoma by targeting BTG3. *Biomed. Pharmacother.* 118, 109267. doi:10.1016/j.biopha.2019.109267

Xia, Y., Xiao, X., Deng, X., Zhang, F., Zhang, X., Hu, Q., et al. (2017). Targeting long non-coding RNA ASBEL with oligonucleotide antagonist for breast cancer therapy. *Biochem. Biophys. Res. Commun.* 489 (4), 386–392. doi:10.1016/j.bbrc.2017.05.136

Xiong, Y., Sun, F., Dong, P., Watari, H., Yue, J., Yu, M. F., et al. (2017). iASPP induces EMT and cisplatin resistance in human cervical cancer through miR-20a-FBXL5/BTG3 signaling. *J. Exp. Clin. Cancer Res.* 36 (1), 48. doi:10.1186/s13046-017-0520-6

Yanagida, S., Taniue, K., Sugimasa, H., Nasu, E., Takeda, Y., Kobayashi, M., et al. (2013). ASBEL, an ANA/BTG3 antisense transcript required for tumorigenicity of ovarian carcinoma. *Sci. Rep.* 3, 1305. doi:10.1038/srep01305

Yoneda, M., Suzuki, T., Nakamura, T., Ajima, R., Yoshida, Y., Kakuta, S., et al. (2009). Deficiency of antiproliferative family protein Ana correlates with development of lung adenocarcinoma. *Cancer Sci.* 100 (2), 225–232. doi:10.1111/j.1349-7006.2008.01030.x

Yoshida, Y., Hosoda, E., Nakamura, T., and Yamamoto, T. (2001). Association of ANA, a member of the antiproliferative Tob family proteins, with a Caf1 component of the CCR4 transcriptional regulatory complex. *Jpn. J. Cancer Res.* 92 (6), 592–596. doi:10.1111/j.1349-7006.2001.tb01135.x

Yu, J., Zhang, Y., Qi, Z., Kurtycz, D., Vacano, G., and Patterson, D. (2008). Methylation-mediated downregulation of the B-cell translocation gene 3 (BTG3) in breast cancer cells. *Gene Expr.* 14 (3), 173–182.

Zheng, H. C., He, H. Y., Wu, J. C., Li, J., Zhao, S., Zhao, G. F., et al. (2017). The suppressing effects of BTG3 expression on aggressive behaviors and phenotypes of colorectal cancer: An *in vitro* and vivo study. *Oncotarget* 8 (11), 18322–18336. doi:10.18632/oncotarget.15438

frontiers | Frontiers in Cell and Developmental Biology

TYPE Review
PUBLISHED 21 September 2022
DOI 10.3389/fcell.2022.1006400

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Bo Li,
University of Toronto, Canada
Liping Sun,
The First Affiliated Hospital of China Medical University, China

*CORRESPONDENCE
Hua-chuan Zheng,
zheng_huachuan@hotmail.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 29 July 2022
ACCEPTED 15 August 2022
PUBLISHED 21 September 2022

CITATION
Zheng H-c, Xue H and Zhang C-y (2022), The roles of the tumor suppressor parafibromin in cancer. *Front. Cell Dev. Biol.* 10:1006400. doi: 10.3389/fcell.2022.1006400

The roles of the tumor suppressor parafibromin in cancer

Hua-chuan Zheng[1]*, Hang Xue[1] and Cong-yu Zhang[2]

[1]Department of Oncology and Central Laboratory, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Cancer Center, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

In this review, we discuss parafibromin protein, which is encoded by *CDC73*. A mutation in this gene causes hyperparathyroidism-jaw tumor (HPT-JT) syndrome, an autosomal dominant disease. *CDC73* is transcriptionally downregulated by the Wilms' tumor suppressor gene WT1 and translationally targeted by miR-182-3p and miR-155. In the nucleus, parafibromin binds to RNA polymerase II and PAF1 complex for transcription. Parafibromin transcriptionally increases the expression of c-Myc, decreases CPEB1 expression by interacting with H3M4, and reduces cyclin D1 expression by binding to H3K9. The RNF20/RNF40/parafibromin complex induces monoubiquitination of H2B-K120, and SHP2-mediated dephosphorylation of parafibromin promotes the parafibromin/β-catenin interaction and induces the expression of Wnt target genes, which is blocked by PTK6-medidated phosphorylation. Parafibromin physically associates with the CPSF and CstF complexes that are essential for *INTS6* mRNA maturation. In the cytosol, parafibromin binds to hSki8 and eEF1Bγ for the destabilization of p53 mRNA, to JAK1/2-STAT1 for STAT1 phosphorylation, and to actinin-2/3 to bundle/cross-link actin filaments. Mice with *CDC73* knockout in the parathyroid develop parathyroid and uterine tumors and are used as a model for HPT-JT syndrome. Conditional deletion of *CDC73* in mesenchymal progenitors results in embryos with agenesis of the heart and liver while its abrogation in mature osteoblasts and osteocytes increases cortical and trabecular bone. Heterozygous germline mutations in *CDC73* are associated with parathyroid carcinogenesis. The rates of *CDC73* mutation and parafibromin loss decrease from parathyroid adenoma to atypical adenoma to carcinoma. In addition, down-regulated parafibromin is closely linked to the tumorigenesis, subsequent progression, or poor prognosis of head and neck, gastric, lung, colorectal, and ovarian cancers, and its overexpression might reverse the aggressiveness of these cancer cells. Therefore, parafibromin might be useful as a biological marker of malignancies and a target for their gene therapy.

Abbreviations: HPT-JT, hyperparathyroidism-jaw tumor; PAF1, polymerase-associated factor 1; WT1, Wilms' tumor suppressor 1; CPEB1, cytoplasmic polyadenylation element binding protein 1; H2B-K120, histone H2B at lysine 120; CstF, cleavage stimulation factor; CPSF, cleavage and polyadenylation specificity factor; MEFs, mouse embryonic fibroblasts; PC, parathyroid cancer; ATA, atypical adenoma; AD, adenoma.

KEYWORDS

parafibromin, cancer, tumor suppressor, tumorigenesis, hyperparathyroidism-jaw tumor (HPT-JT) syndrome

Introduction

Hyperparathyroidism-jaw tumor (HPT-JT) syndrome is an autosomal dominant disease that is characterized by parathyroid tumors, fibro-osseous jaw tumors of the mandible or maxilla, and renal disorders (hamartoma, cystic renal disease, or Wilms' tumor) and results from a mutation in *CDC73* (also called *HRPT2*) (Carpten et al., 2002). Parafibromin functions as a tumor suppressor in the tumorigenesis and subsequent progression of parathyroid carcinomas. Parafibromin not only forms a PAF1 (polymerase-associated factor 1) complex for transcriptional events and histone modifications during cell growth and survival, but also contributes to cell mobility. In this review, we summarize the gene structure, biological functions, signaling pathways, and phenotypes of *CDC73* knockout mice and the relationship of genetic and expression alterations in *CDC73* with cancer.

Structure and expression of *CDC73*

CDC73 maps to human chromosome 1q31.2, spans 18.5 kb, comprises 17 exons, and encodes a 2.7-kb mRNA that is translated into a 60-kDa parafibromin comprising 531 amino acids (Carpten et al., 2002). At the transcriptional level, overexpression of Wilms' tumor suppressor 1 (WT1) decreases *CDC73* levels and promotes the proliferation of oral squamous cell carcinoma cells by binding to the *CDC73* promoter (Rather et al., 2014). At the translational level, *CDC73* can be targeted and inhibited by miR-182-3p, and its knockdown can reverse the suppressive effects of miR-182-3p inhibitor on the aggressive phenotypes of oral squamous cell carcinoma cells (Guo et al., 2020). Moreover, the oncogenic miR-155 dramatically reduces parafibromin expression in HEK293 cells (Rather et al., 2013). Using northern blot, *CDC73* mRNA expression is detectable in the heart, placenta, brain, lung, skeletal muscle, liver, pancreas, and kidney. In terms of western blot, the 60-kDa form of parafibromin is observed in human adrenal gland, pancreas, heart, and kidney, while the 40-kDa form is found in the skeletal muscle and heart (Carpten et al., 2002). Immunohistochemically, parafibromin positivity has been seen in gastric glandular cells, hepatocytes, glomerular mesangial cell, renal cortex tubules, and hypophysis in both nuclear and nucleocytoplasmic patterns (Porzionato et al., 2006). The expression profile of parafibormin might be closely linked to its biological functions in the cellular compartments.

Biological functions of parafibromin

In the nucleus (Figure 1), parafibromin binds to the human PAF1-LEO1-CTR9 complex and RNA polymerase II for transcription regulation and 3′ flank modification (Rozenblatt-Rosen et al., 2005). Silencing either parafibromin or PAF1 stimulates cell proliferation and enhances *c-myc* expression due to c-myc protein stabilization and *c-myc* promoter activation, with no alleviation of the *c-myc* transcriptional pause (Lin et al., 2008). Parafibromin can also bind to a histone methyltransferase complex for histone H3 methylation at lysine 4 (Lin et al., 2008) and to the promoter of CPEB1 (cytoplasmic polyadenylation element binding protein 1) for the downregulation of CPEB1 expression (Zhang et al., 2010). Parafibromin interacts with the histone methyltransferase SUV39H1 to induce histone H3K9 methylation and suppress cyclin D1 expression (Yang et al., 2010). The heterodimeric parafibromin/RNF20/40 complex acts as an E3 ubiquitin ligase to ubiquitinate histone H2B at lysine 120 (H2B-K120) (Hahn et al., 2012). Upon SHP2-mediated tyrosine dephosphorylation, parafibromin stably interacts with β-catenin to overcome the parafibromin/SUV39H1-induced transrepression and promote the expression of Wnt/β-catenin target genes, such as *c-Myc* and *cyclin D1* (Takahashi et al., 2011). In addition, parafibromin competitively binds to β-catenin and Gli1 and thereby transactivates Wnt- and Hh-target genes; this activity is strengthened by SHP2 phosphatase, but weakened by PTK6 kinase (Kikuchi et al., 2016). As for RNA modification, Parafibromin cleaves the 3′ end of histone mRNA coupled with polyadenylated tails (Farber et al., 2010), and interacts with cleavage stimulation factor (CstF) and the cleavage and polyadenylation specificity factor (CPSF) complex essential for the maturation of the *INTS6* mRNA 3′ flank (Jo et al., 2014). Taken together, parafibromin is involved in the transcriptional regulation of target genes and mRNA maturation via protein–protein complexes.

In the cytosol (Figure 2), parafibromin physically associates with hSki8 and eEF1Bγ to destabilize *p53* mRNA and inhibit p53-mediated apoptosis (Jo et al., 2014). Parafibromin also interacts with JAK1/2, promotes the formation of the JAK1–JAK2 complex and then the JAK1/2–STAT1 complex, and enhances the JAK-mediated tyrosine phosphorylation of STAT1 upon IFN-γ stimulation (Wei et al., 2015). The N-terminal region of parafibromin binds to actinin-3 and actinin-2 to bundle/cross-link actin filaments for cell mobility in the cytoplasmic compartment (Agarwal et al., 2008), in agreement with the parafibromin expression in the cilia of the bronchial pseudo-stratified columnar epithelium (Xia et al.,

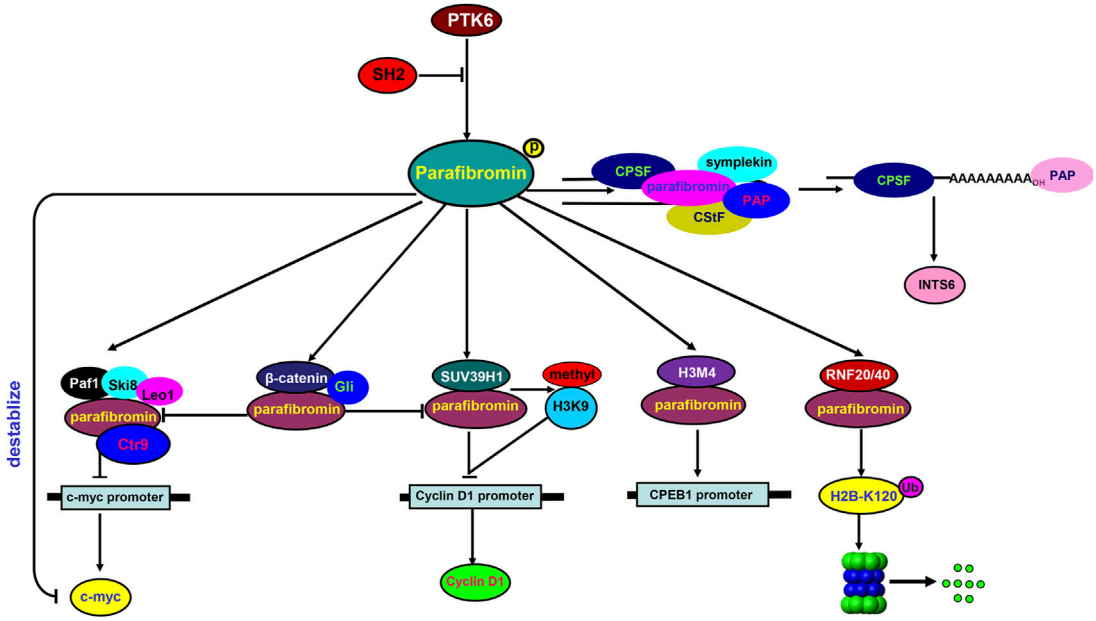


FIGURE 1
Biological functions of parafibromin in the nucleus In the nucleus, parafibromin interacts with PAF1-Ski8-LEO1-CTR9 to suppress *c-Myc* mRNA expression. It can also bind to a H3M4 and H3K9 methyltransferase complex to downregulate the transcription of *CPEB1* and *cyclin D1*, respectively. The ring finger proteins RNF20/RNF40 bind to parafibromin for monoubiquitination of histone H2B at lysine 120 (H2B-K120). For tyrosine dephosphorylation by SHP2, parafibromin competitively interacts with β-catenin and Gli1 to induce the expression of Wnt target genes, which is attenuated by tyrosine phosphorylation via PTK6 kinase. Parafibromin physically associates with CPSF and CstF complexes, which are required for the maturation of the *INTS6* mRNA 3′ flank.

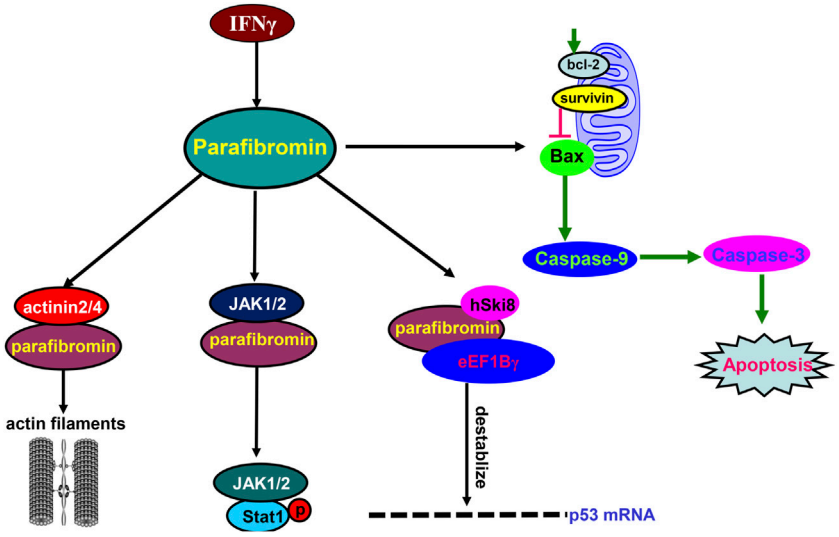


FIGURE 2
Biological functions of parafibromin in the cytosol. In the cytosol, parafibromin physically binds to eEF1Bγ and hSki8 to destabilize *p53* mRNA. Parafibromin interacts with JAK1/2, promotes the interactions of JAK1-JAK2 and JAK1/2-STAT1, and enhances the tyrosine phosphorylation of STAT1 by JAKs after IFN-γ stimulation. Parafibromin interacts with actinin-2 and actinin-3 to bundle/cross-link actin filaments. Finally, parafibromin causes apoptosis by activating caspase-3 and -9 and downregulating the expression of Bcl-2 and survivin.

2011) and fallopian tube (Shen et al., 2016). Parafibromin overexpression induces apoptosis by activating caspase-3 and -9 and by suppressing survivin and Bcl-2 expression (Zhu et al., 2016). Thus, we suggest that parafibromin contributes to mobility, apoptosis, and proliferation via protein–protein interaction.

Phenotypes of *CDC73* knockout mice

Reportedly, *CDC73* double knockout was embryonically lethal in mice from E6.5. Temporal deletion of *CDC73* after E8.5 caused growth retardation and extensive apoptosis (Wang et al., 2008). *CDC73* knockout resulted in severe cachexia and death of the adult mice within 20 days. $CDC73^{-/-}$ mouse embryonic fibroblasts (MEFs) underwent apoptosis, whereas $CDC73^{+/+}$ and $^{+/-}$ MEFs grew normally. The parafibromin/PAF1 complex was found to directly regulate genes related to cell growth and survival, including *H19*, *Hmga1*, *Hmga2*, *Hmgcs2*, *Igf1*, *Igf2*, and *Igfbp4*. The mice with parathyroid-specific deletion of *CDC73* developed parathyroid tumors and could be used as an animal model of HPT-JT syndrome (Walls et al., 2017), and *CDC73* loss in hematopoietic cells was lethal because cell cycle defects in hematopoietic progenitors resulted in bone marrow failure (Saha et al., 2019). Moreover, homozygous knockout of *CDC73* in mesenchymal progenitors (via Dermo1-cre) blocked mesenchymal organ development, such as heart and liver, and displayed extensive apoptosis. The homozygous *CDC73* abrogation in osteoblasts and osteocytes (via Ocn-cre) had no influence on the life span of the mice, but increased cortical and trabecular bone levels and resulted in cytoplasmic RNA accumulation and elevated apoptosis in the osteocytes of the femur (Droscha et al., 2017). These results indicate that *CDC73* knockout mice experience developmental retardation or tumor, possibly via aberrant apoptosis.

Genetic alterations of *CDC73* during parathyroid carcinogenesis

Biallelic mutation in *CDC73* is strongly related to the malignancy of parathyroid tumors, among which parathyroid cancer (PC) is a rare cancer with an unfavorable prognosis (Hahn et al., 2010). *CDC73* germline mutation causing significant conformational alterations in the conserved C-terminal domain or parafibromin loss (identified as "high-impact mutations") were substantially more frequently observed in PC patients than in all other individuals with benign tumor. These high-impact mutations were linked to a 6.6-fold higher risk of PC than low-impact mutations. The mutations were mostly nonsense and frameshift, whereas missense mutations were rare and always disrupted the N terminus of parafibromin (Li et al., 2020). *CDC73* mutation rate gradually decreased from PC to atypical adenoma (ATA) to adenoma (AD). A recurrent 2-bp mutation in exon 7 (c.679_680delAG) was responsible for 50% of all identified mutations (Guarnieri et al., 2017). PCs with *CDC73* mutation displayed a high probability of either recurrence or metastasis (Cetani et al., 2013). Wang et al. (2012) identified six mutations in 6 of 13 PC patients, with three being novel and four being germline. PC patients with *CDC73* mutations were more likely to develop recurrence. Copy number alteration in *CDC73* and chromosomal loss at 1p and 13 were only seen in PC, and *CDC73* hypermethylation was not observed in parathyroid tumors (Sulaiman et al., 2012). Both CpG island hyper-methylation and 5′UTR mutation of *CDC73* rarely silenced parafibromin expression in PC as well (Hahn et al., 2010). Additionally, Masi et al. (2014) found that Ile60Asn mutant parafibromin broke its nucleolar localization and was under-expressed due to proteasomal degradation. Overexpression of the Ile60Asn mutant parafibromin failed to suppress c-myc expression, suggesting that Ile60Asn mutant parafibromin lost the ability to down-regulate c-myc expression. In the combination of these findings, *CDC73* mutations are more common during parathyroid carcinogenesis and are closely associated with the recurrence and metastasis of PC by dysfunction of its encoding protein. However, the promoter methylation or mutation of *CDC73* 5′ flank is rarely responsible for its expression loss.

Parafibromin expression during parathyroid carcinogenesis

Immunohistochemically, parafibromin has been shown to localize to the nucleus and play a tumor-suppressor role. The gradual absence of its nucleolar expression was evident from PC, ATA to AD (Witteveen et al., 2011), consistent with the data of Juhlin et al. (Juhlin et al., 2011). Parafibromin loss in PC was associated it with a 4-fold increased risk of developing local invasion, metastasis and recurrence or metastasis (Witteveen et al., 2011; Kim et al., 2012; Hu et al., 2016). Juhlin et al. (Juhlin et al., 2011) found that the male patients with high-proliferative parathyroid tumors had aberrant parafibromin expression, and PC patients more frequently harbored aberrant parafibromin expression than ATA and AD patients. Parafibromin-negative patients with PC were younger, had larger tumors and more frequent *CDC73* mutation/deletions. PC cells were morphologically characterized by eosinophilic cytoplasm, sheet-like growth, nuclear enlargement, coarse chromatin, perinuclear clear appearance, and branch vasculature (Gill et al., 2019). Although the scoring standard and parafibromin antibodies determined the

immunohistochemical and predictive sensitivity and specificity of parafibromin expression (Hu et al., 2019), a meta-analysis showed parafibromin loss was found to be more common in PC patients than in those with parathyroid ATA, AD, and hyperplasia. Parafibromin immunohistochemistry can be useful for the diagnostic and prognostic evaluation of PC in clinicopathological practice (Pyo and Cho, 2019). Zhu et al. (Zhu et al., 2020), who included 193 PC patients from nine studies, demonstrated that parafibromin immunonegativity might be used to reflect the risk of recurrence, metastasis, and death in PC. These findings suggest that parafibromin expression can be employed to indicate the tumorigenesis, progression and prognosis of PC.

Clinicopathological significance and effects of parafibromin expression on other cancers

Previously, we determined the parafibromin profiles and their clinicopathological significances in head and neck, lung, gastric, colorectal, and ovarian cancers. The downregulated expression of parafibromin in the above-mentioned cancers was negatively correlated with their aggressive variables and adverse prognosis (Zheng et al., 2011; Shen et al., 2016; Zhu et al., 2016; Zheng et al., 2017a; Zheng et al., 2017b). In breast cancer, parafibromin expression was inversely correlated with T stage, clinicopathological stage, local lymphovascular invasion, and C-erbB2 expression (Selvarajan et al., 2008). There was a negative association between the lymph T stage and parafibromin expression in urothelial carcinoma (Karaarslan et al., 2015). Parafibromin expression was positive in 15 cases (50%) of laryngeal squamous cell carcinoma and inversely linked to tumor size and T stage (Cho et al., 2016). Parafibromin expression was immunohistochemically downregulated from normal squamous tissue to dysplasia and then from primary to metastatic cancers of the head and neck and negatively correlated with the N stage, clinicopathological stage, dedifferentiation, and human papillomavirus negativity (Zhang et al., 2015), consistent with the results for colorectal cancer (Zheng et al., 2011). Tongue cancers showed more parafibromin expression than laryngeal cancers (Cho et al., 2016). Parafibromin expression was an independent prognostic factor for the overall or relapse-free survival of head and neck squamous cell carcinomas (Zhang et al., 2015), ovarian cancer (Shen et al., 2016), and colorectal cancer (Zheng et al., 2011). Taken together, downregulated parafibromin expression might be used as a biomarker for tumorigenesis, aggressive behavior, and poor prognosis of non-parathyroid malignancies.

CDC73 is underexpressed in colorectal cancer and its expression is negatively related to colorectal cancer differentiation at both mRNA and protein levels (Zheng et al., 2017a; Zheng et al., 2017b). Xia et al. (2011) reported that *CDC73* mRNA expression was downregulated in lung cancer in comparison with matched normal tissue, in line with the observation in ovarian cancer (Shen et al., 2016). Bioinformatics analysis showed that *CDC73* mRNA levels were higher in gastric, breast, lung, and ovarian cancers than the corresponding normal tissues and were positively correlated with the differentiation and better prognosis of gastric cancer and with the M stage and clinicopathological stage of lung cancer (55). These findings indicate that the clinicopathological significance of *CDC73* expression depends on cancer type and detection method.

The effects of parafibromin on the aggressive phenotypes of cancer cells determine whether it can be used as a target of gene therapy. In head and neck squamous cell carcinoma cells, parafibromin overexpression suppresses cell proliferation, migration, invasion, and epithelial-mesenchymal transition and induces apoptosis and S arrest (Zheng et al., 2017a). Shen et al. (2016) found that parafibromin overexpression inhibited cell proliferation, anti-apoptosis, migration, and invasion, as well as the chemoresistance to cisplatin in ovarian cancer cells by inhibiting PI3K/Akt and suppressing the expression of VEGF and MMP-9. In colorectal cancer cells, nuclear parafibromin inhibited proliferation and tumor growth and induced apoptosis and cell cycle arrest in colorectal cancer cells, but it was the converse for cytosolic parafibromin. Transcriptomically, nuclear parafibromin inhibited PI3K-Akt and FoxO signaling pathways, while the cytosolic form activated the PI3K-Akt pathway and cell mobility. Thus, the overall results indicate that parafibromin could be used as a gene therapy target in the future (Zheng et al., 2017a).

Conclusions and future perspectives

In summary, genetic alteration of *CDC73* and altered parafibromin expression both cause HPT-JT, parathyroid, gastric, colorectal, ovarian, lung, and head and neck carcinogenesis. Its down-regulated expression is negatively correlated with the aggressive behaviors and poor prognosis of these cancers. Thus, its overexpression might suppress the proliferation, migration, and invasion of these cancer cells.

According to recent findings concerning *CDC37*, we believe that the genetic study of *CDC73* should be used to protect against and diagnose genetic PC. Aberrant parafibromin expression should be used to predict the tumorigenesis, aggressive behavior, and poor prognosis of malignancies. Wild-type *CDC37* might be used as a molecular target for gene therapy for cancers on the basis of the parafibromin expression status. In the future, the cytosolic function of parafibromin should be thoroughly investigated in cancer cells and in the cilia of the bronchial pseudo-stratified columnar epithelium and fallopian tube.

Author contributions

H-CZ designed and drafted this manuscript. C-YZ prepared the figure. HX edited and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (21377772D) and National Natural Scientific Foundation of China (81672700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwal, S. K., Simonds, W. F., and Marx, S. J. (2008). The parafibromin tumor suppressor protein interacts with actin-binding proteins actinin-2 and actinin-3. *Mol. Cancer* 7, 65. doi:10.1186/1476-4598-7-65

Carpten, J. D., Robbins, C. M., Villablanca, A., Forsberg, L., Presciuttini, S., Bailey-Wilson, J., et al. (2002). HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat. Genet.* 32 (4), 676–680. doi:10.1038/ng1048

Cetani, F., Banti, C., Pardi, E., Borsari, S., Viacava, P., Miccoli, P., et al. (2013). CDC73 mutational status and loss of parafibromin in the outcome of parathyroid cancer. *Endocr. Connect.* 2 (4), 186–195. doi:10.1530/EC-13-0046

Cho, I., Lee, M., Lim, S., and Hong, R. (2016). Significance of parafibromin expression in laryngeal squamous cell carcinomas. *J. Pathol. Transl. Med.* 50 (4), 264–269. doi:10.4132/jptm.2016.04.24

Droscha, C. J., Diegel, C. R., Ethen, N. J., Burgers, T. A., McDonald, M. J., Maupin, K. A., et al. (2017). Osteoblast-specific deletion of Hrpt2/Cdc73 results in high bone mass and increased bone turnover. *Bone* 98, 68–78. doi:10.1016/j.bone.2016.12.006

Farber, L. J., Kort, E. J., Wang, P., Chen, J., and Teh, B. T. (2010). The tumor suppressor parafibromin is required for posttranscriptional processing of histone mRNA. *Mol. Carcinog.* 49 (3), 215–223. doi:10.1002/mc.20591

Gill, A. J., Lim, G., Cheung, V. K. Y., Andrici, J., Perry-Keene, J. L., Paik, J., et al. (2019). Parafibromin-deficient (HPT-JT type, CDC73 mutated) parathyroid tumors demonstrate distinctive morphologic features. *Am. J. Surg. Pathol.* 43 (1), 35–46. doi:10.1097/PAS.0000000000001017

Guarnieri, V., Seaberg, R. M., Kelly, C., Jean Davidson, M., Raphael, S., Shuen, A. Y., et al. (2017). Erratum to: Large intragenic deletion of CDC73 (exons 4-10) in a three-generation hyperparathyroidism-jaw tumor (HPT-JT) syndrome family. *BMC Med. Genet.* 18, 99. doi:10.1186/s12881-017-0459-7

Guo, J., Su, Y., and Zhang, M. (2020). Circ_0000140 restrains the proliferation, metastasis and glycolysis metabolism of oral squamous cell carcinoma through upregulating CDC73 via sponging miR-182-5p. *Cancer Cell. Int.* 20, 407. doi:10.1186/s12935-020-01501-7

Hahn, M. A., Dickson, K. A., Jackson, S., Clarkson, A., Gill, A. J., and Marsh, D. J. (2012). The tumor suppressor CDC73 interacts with the ring finger proteins RNF20 and RNF40 and is required for the maintenance of histone 2B monoubiquitination. *Hum. Mol. Genet.* 21 (3), 559–568. doi:10.1093/hmg/ddr490

Hahn, M. A., Howell, V. M., Gill, A. J., Clarkson, A., Weaire-Buchanan, G., Robinson, B. G., et al. (2010). CDC73/HRPT2 CpG island hypermethylation and mutation of 5'-untranslated sequence are uncommon mechanisms of silencing parafibromin in parathyroid tumors. *Endocr. Relat. Cancer* 17 (1), 273–282. doi:10.1677/ERC-09-0291

Hu, Y., Bi, Y., Cui, M., Zhang, X., Su, Z., Wang, M., et al. (2019). The influence of surgical extent and paraifbromin staining on the outcome of parathyroid carcinoma: 20-year experience from a single institute. *Endocr. Pract.* 25 (7), 634–641. doi:10.4158/EP-2018-0538

Hu, Y., Liao, Q., Cao, S., Gao, X., and Zhao, Y. (2016). Diagnostic performance of parafibromin immunohistochemical staining for sporadic parathyroid carcinoma: A meta-analysis. *Endocrine* 54 (3), 612–619. doi:10.1007/s12020-016-0997-3

Jo, J. H., Chung, T. M., Youn, H., and Yoo, J. Y. (2014). Cytoplasmic parafibromin/hCdc73 targets and destabilizes p53 mRNA to control p53-mediated apoptosis. *Nat. Commun.* 5, 5433. doi:10.1038/ncomms6433

Juhlin, C. C., Haglund, F., Obara, T., Arnold, A., Larsson, C., and Höög, A. (2011). Absence of nucleolar parafibromin immunoreactivity in subsets of parathyroid malignant tumours. *Virchows Arch.* 459 (1), 47–53. doi:10.1007/s00428-010-1032-3

Karaarslan, S., Yaman, B., Ozturk, H., and Kumbaraci, B. S. (2015). Parafibromin staining characteristics in urothelial carcinomas and relationship with prognostic parameters. *J. Pathol. Transl. Med.* 49 (5), 389–395. doi:10.4132/jptm.2015.08.10

Kikuchi, I., Takahashi-Kanemitsu, A., Sakiyama, N., Tang, C., Tang, P. J., Noda, S., et al. (2016). Dephosphorylated parafibromin is a transcriptional coactivator of the Wnt/Hedgehog/Notch pathways. *Nat. Commun.* 7, 12887. doi:10.1038/ncomms12887

Kim, H. K., Oh, Y. L., Kim, S. H., Lee, D. Y., Kang, H. C., Lee, J. I., et al. (2012). Parafibromin immunohistochemical staining to differentiate parathyroid carcinoma from parathyroid adenoma. *Head. Neck* 34 (2), 201–206. doi:10.1002/hed.21716

Li, Y., Zhang, J., Adikaram, P. R., Welch, J., Guan, B., Weinstein, L. S., et al. (2020). Genotype of CDC73 germline mutation determines risk of parathyroid cancer. *Endocr. Relat. Cancer* 27 (9), 483–494. doi:10.1530/ERC-20-0149

Lin, L., Zhang, J. H., Panicker, L. M., and Simonds, W. F. (2008). The parafibromin tumor suppressor protein inhibits cell proliferation by repression of the c-myc proto-oncogene. *Proc. Natl. Acad. Sci. U. S. A.* 105 (45), 17420–17425. doi:10.1073/pnas.0710725105

Masi, G., Iacobone, M., Sinigaglia, A., Mantelli, B., Pennelli, G., Castagliuolo, I., et al. (2014). Characterization of a new CDC73 missense mutation that impairs Parafibromin expression and nucleolar localization. *PLoS One* 9, e97994. doi:10.1371/journal.pone.0097994

Porzionato, A., Macchi, V., Barzon, L., Masi, G., Iacobone, M., Parenti, A., et al. (2006). Immunohistochemical assessment of parafibromin in mouse and human tissues. *J. Anat.* 209 (6), 817–827. doi:10.1111/j.1469-7580.2006.00657.x

Pyo, J. S., and Cho, W. J. (2019). Diagnostic and prognostic implications of parafibromin immunohistochemistry in parathyroid carcinomaT. *Biosci. Rep.* 39, BSR20181778. doi:10.1042/BSR20181778

Rather, M. I., Nagashri, M. N., Swamy, S. S., Gopinath, K. S., and Kumar, A. (2013). Oncogenic microRNA-155 down-regulates tumor suppressor CDC73 and promotes oral squamous cell carcinoma cell proliferation: Implications for cancer therapeutics. *J. Biol. Chem.* 288 (1), 608–618. doi:10.1074/jbc.M112.425736

Rather, M. I., Swamy, S., Gopinath, K. S., and Kumar, A. (2014). Transcriptional repression of tumor suppressor CDC73, encoding an RNA polymerase II interactor, by Wilms tumor 1 protein (WT1) promotes cell proliferation: Implication for cancer therapeutics. *J. Biol. Chem.* 289 (2), 968–976. doi:10.1074/jbc.M113.483255

Rozenblatt-Rosen, O., Hughes, C. M., Nannepaga, S. J., Shanmugam, K. S., Copeland, T. D., Guszczynski, T., et al. (2005). The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Mol. Cell. Biol.* 25 (2), 612–620. doi:10.1128/MCB.25.2.612-620.2005

Saha, N., Ropa, J., Chen, L., Hu, H., Mysliwski, M., Friedman, A., et al. (2019). The PAF1c subunit CDC73 is required for mouse hematopoietic stem cell maintenance but displays leukemia-specific gene regulation. *Stem Cell. Rep.* 12 (5), 1069–1083. doi:10.1016/j.stemcr.2019.03.010

Selvarajan, S., Sii, L. H., Lee, A., Yip, G., Bay, B. H., Tan, M. H., et al. (2008). Parafibromin expression in breast cancer: A novel marker for prognostication? *J. Clin. Pathol.* 61 (1), 64–67. doi:10.1136/jcp.2007.048694

Shen, D. F., Liu, X., Yang, X. F., Fang, L., Gao, Y., Zhao, S., et al. (2016). The roles of parafibromin expression in ovarian epithelial carcinomas: A marker for differentiation and prognosis and a target for gene therapy. *Tumour Biol.* 37 (3), 2909–2924. doi:10.1007/s13277-015-4103-x

Sulaiman, L., Haglund, F., Hashemi, J., Obara, T., Nordenström, J., Larsson, C., et al. (2012). Genome-wide and locus specific alterations in CDC73/HRPT2-mutated parathyroid tumors. *PLoS One* 7, e46325. doi:10.1371/journal.pone.0046325

Takahashi, A., Tsutsumi, R., Kikuchi, I., Obuse, C., Saito, Y., Seidi, A., et al. (2011). SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver. *Mol. Cell.* 43 (1), 45–56. doi:10.1016/j.molcel.2011.05.014

Walls, G. V., Stevenson, M., Lines, K. E., Newey, P. J., Reed, A. A. C., Bowl, M. R., et al. (2017). Mice deleted for cell division cycle 73 gene develop parathyroid and uterine tumours: Model for the hyperparathyroidism-jaw tumour syndrome. *Oncogene* 36 (28), 4025–4036. doi:10.1038/onc.2017.43

Wang, O., Wang, C., Nie, M., Cui, Q., Guan, H., Jiang, Y., et al. (2012). Novel HRPT2/CDC73 gene mutations and loss of expression of parafibromin in Chinese patients with clinically sporadic parathyroid carcinomas. *PLoS One* 7, e45567. doi:10.1371/journal.pone.0045567

Wang, P., Bowl, M. R., Bender, S., Peng, J., Farber, L., Chen, J., et al. (2008). Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. *Mol. Cell. Biol.* 28 (9), 2930–2940. doi:10.1128/MCB.00654-07

Wei, J., Lian, H., Zhong, B., and Shu, H. B. (2015). Parafibromin is a component of IFN-γ- triggered signaling pathways that facilitates JAK1/2-mediated tyrosine phosphorylation of STAT1. *J. Immunol.* 195 (6), 2870–2878. doi:10.4049/jimmunol.1501111

Witteveen, J. E., Hamdy, N. A., Dekkers, O. M., Kievit, J., van Wezel, T., Teh, B. T., et al. (2011). Downregulation of CASR expression and global loss of parafibromin staining are strong negative determinants of prognosis in parathyroid carcinoma. *Mod. Pathol.* 24 (5), 688–697. doi:10.1038/modpathol.2010.236

Xia, P., Wang, W., Xu, X. Y., Wang, J. P., Takano, Y., and Zheng, H. C. (2011). Parafibromin expression in lung normal tissue and carcinoma: Its comparison with clinicopathological parameters of carcinoma. *Histol. Histopathol.* 26 (8), 1039–1047. doi:10.14670/HH-26.1039

Yang, Y. J., Han, J. W., Youn, H. D., and Cho, E. J. (2010). The tumor suppressor, parafibromin, mediates histone H3 K9 methylation for cyclin D1 repression. *Nucleic Acids Res.* 38 (2), 382–390. doi:10.1093/nar/gkp991

Zhang, J. H., Panicker, L. M., Seigneur, E. M., Lin, L., House, C. D., Morgan, W., et al. (2010). Cytoplasmic polyadenylation element binding protein is a conserved target of tumor suppressor HRPT2/CDC73. *Cell. Death Differ.* 17 (10), 1551–1565. doi:10.1038/cdd.2010.32

Zhang, Z., Yang, X. F., Huang, K. Q., Ren, L., Gou, W. F., Shen, D. F., et al. (2015). The clinicopathological significances and biological functions of parafibromin expression in head and neck squamous cell carcinomas. *Tumour Biol.* 36 (12), 9487–9497. doi:10.1007/s13277-015-3618-5

Zheng, H. C., Gong, B. C., and Zhao, S. (2017). The clinicopathological and prognostic significances of CDC73 expression in cancers: A bioinformatics analysis. *Oncotarget* 8 (56), 95270–95279. doi:10.18632/oncotarget.20446

Zheng, H. C., Liu, J. J., Li, J., Wu, J. C., Yang, L., Zhao, G. F., et al. (2017). The *in vitro* and vivo effects of nuclear and cytosolic parafibromin expression on the aggressive phenotypes of colorectal cancer cells: A search of potential gene therapy target. *Oncotarget* 8 (14), 23603–23612. doi:10.18632/oncotarget.15377

Zheng, H. C., Wei, Z. L., Xu, X. Y., Nie, X. C., Yang, X., Takahashi, H., et al. (2011). Parafibromin expression is an independent prognostic factor for colorectal carcinomas. *Hum. Pathol.* 42 (8), 1089–1102. doi:10.1016/j.humpath.2010.10.024

Zhu, J. J., Cui, Y., Cui, K., Li, X., and Zhang, Z. Y. (2016). Distinct roles of parafibromin in the extracellular environment, cytoplasm and nucleus of osteosarcoma cells. *Am. J. Transl. Res.* 8, 2426–2431.

Zhu, R., Wang, Z., and Hu, Y. (2020). Prognostic role of parafibromin staining and CDC73 mutation in patients with parathyroid carcinoma: A systematic review and meta-analysis based on individual patient data. *Clin. Endocrinol.* 92 (4), 295–302. doi:10.1111/cen.14161

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fgene.2022.1008602

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Jing Ji,
Jiangsu Ocean Universiity, China
An Qin,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Linhua Liu,
liulinhua@hotmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

RECEIVED 01 August 2022
ACCEPTED 08 September 2022
PUBLISHED 23 September 2022

CITATION
Huang S, Chen S, Zhang D, Gao J and Liu L (2022), Enhancer-associated regulatory network and gene signature based on transcriptome and methylation data to predict the survival of patients with lung adenocarcinoma. *Front. Genet.* 13:1008602. doi: 10.3389/fgene.2022.1008602

Enhancer-associated regulatory network and gene signature based on transcriptome and methylation data to predict the survival of patients with lung adenocarcinoma

Shihao Huang[1†], Shiyu Chen[2†], Di Zhang[1], Jiamei Gao[1] and Linhua Liu[1*]

[1]Department of Biochemistry, Institute of Glycobiology, Dalian Medical University, Dalian, Liaoning, China, [2]Department of Laboratory Medicine, Nanxishan Hospital of Guangxi Zhuang Autonomous Region, Guilin, China

Accumulating evidence has proved that aberrant methylation of enhancers plays regulatory roles in gene expression for various cancers including lung adenocarcinoma (LUAD). In this study, the transcriptome and methylation data of The Cancer Genome Atlas (TCGA)-LUAD cohort were comprehensively analyzed with a five-step Enhancer Linking by Methylation/Expression Relationships (ELMER) process. Step 1: 131,371 distal (2 kb upstream from the transcription start site) probes were obtained. Step 2: 10,665 distal hypomethylated probes were identified in an unsupervised mode with the get.diff.meth function. Step 3: 699 probe-gene pairs with negative correlations were screened using the get.pair function in an unsupervised mode. Step 4: After mapping with probes, 768 motifs were obtained and 24 of them were enriched. Step 5: 127 transcription factors (TFs) with differential expressions and negative correlations with methylation levels were screened, which were corresponding to 21 motifs. After the ELMER process, a prognostic "TFs-motifs-genes" regulatory network was constructed. The Least absolute shrinkage and selection operator (LASSO) and Stepwise regression analyses were further applied to identify variables in the TCGA-LUAD cohort and an eight-gene signature was constructed for calculating the risk score. The risk score was verified in two independent validation cohorts. The area under curve values of receiver operating characteristic curves predicting 1-, 3-, and 5-years survival ranged from 0.633 to 0.764. With the increase of the risk scores, both the survival statuses and clinical traits showed a worse tendency. There were significant differences in the degrees of immune cell infiltration, TMB values, and TIDE scores between the high-risk and low-risk groups. Finally, a better-performing prognostic nomogram was integrated with the risk score and other clinical traits. In short, this multi-omics analysis demonstrated the application of ELMER in analyzing enhancer-associated regulatory network in

LUAD, which provided promising strategies for epigenetic therapy and prognostic biomarkers.

KEYWORDS

lung adenocarcinoma, enhancer, methylation, regulatory network, gene signature, prognostic model

Introduction

Enhancer is a DNA sequence in the genome with a length of 50–1,500 bp, which can bind with transcription factors (TFs) to promote the transcription of the target gene. The position of enhancer is not fixed and can be at the near end or the far end of a target gene. The enhancer may be upstream or downstream of its regulatory gene (Field and Adelman, 2020).

Enhancers have been reported to reflect normal and pathogenic cellular conditions (Fishilevich et al., 2017). Some high-throughput identification approaches have been developed to predict the enhancers and corresponding functions (Kleftogiannis et al., 2016). With the progress of functions, enhancers have been found to link with several diseases (Wang et al., 2018). Researchers have tried to build a database for disease-associated enhancers: DiseaseEnhancer (Zhang et al., 2018).

Reviews are describing the roles of enhancers in tumors (Sur and Taipale, 2016). As important regulatory elements of DNA, enhancers participate in several comprehensive regulatory networks of cancer-associated genes. Mutations in tumors often lead to aberrantly regulated enhancers, as well as abnormal expression of growth-related genes (Adhikary et al., 2021). The abnormal regulation can be *trans*-action, such as the activation of a transcription factor or apparent regulatory factors that control enhancer activity. Similarly, abnormal regulation can also be *cis*-action, such as mutation to change enhancer activity or its specificity to the target gene (Chen and Liang, 2020). Investigating the activity regulation and related mechanism of tumor type-specific enhancers at the molecular level may be applied for screening therapeutic targets.

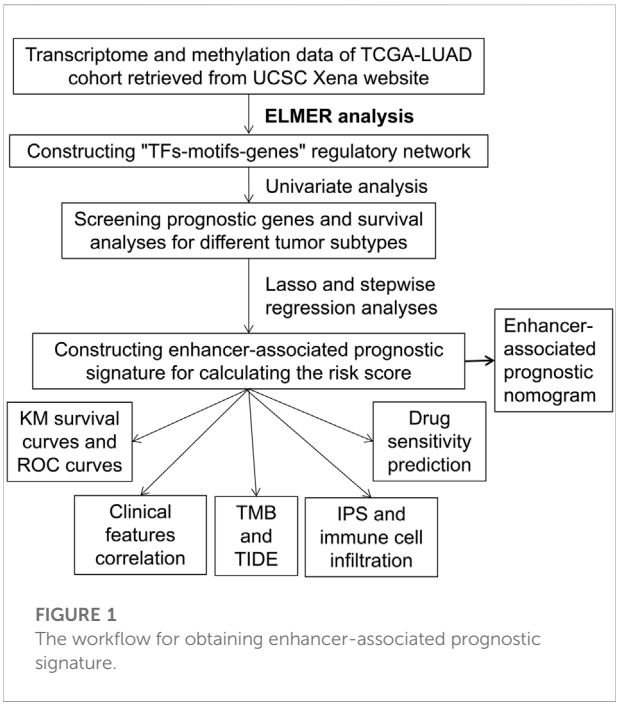


FIGURE 1
The workflow for obtaining enhancer-associated prognostic signature.

Accumulating evidence has suggested that many aberrant methylation sites have been observed on enhancer sequences in cancer cells (Herz, 2016). These abnormal methylations have been proven to link with the expression of a target gene, as well as the disease progression. The methylation state of enhancer regions is the promising next generation of epigenetic biomarkers (Clermont et al., 2016). One study has revealed the abnormal enhancer of hepatocellular carcinoma (HCC) based on multi-omics data (Xiong et al., 2019). By comprehensive analysis of ChIP-seq data, transcriptome data, DNA methylation data, and HiC data, the abnormal enhancer and related transcription disorders in HCC have been described, and the differentially methylated enhancer and its target genes were identified. A prognostic model based on these differentially expressed genes (DEGs) of abnormal enhancers was constructed, which predicted the prognosis of HCC. It was beneficial to the development of epigenetic therapy for HCC (Huang et al., 2022). However, the epigenetic regulation and function of transcription enhancers have been still unclear.

Lung cancer is the leading cause of cancer mortality. It is classified into various histologic subtypes, including adenocarcinoma, squamous carcinoma, non-small cell lung cancer, and small cell lung cancer (Ruiz-Cordero and Devine, 2020). Lung cancer exhibited a good response to novel targeted therapies, such as checkpoint immunotherapy. With the advances in knowledge on molecular characteristics of lung cancer, researchers have found different treatment decisions should be provided to patients with varied gene expression profiles (Chen et al., 2020). For example, some patients may show a better response to immunotherapy, while others should receive targeted therapies and chemotherapy before considering immunotherapy as a single agent (Mazieres et al., 2019). Drug resistance should also be considered (Denisenko et al., 2018). Some studies have tried to find the diagnostic and prognostic prediction markers for lung cancers, such as the immune-related genes signature (Yi et al., 2021), m6A modification (Li et al., 2021), and costimulatory molecule-based signature (Zhang et al.,

2020). All these efforts have contributed to the precision and individual treatment of patients with tumors.

This study aims to find enhancer-associated prognostic biomarkers. The transcriptome and methylation of lung adenocarcinoma (LUAD) cohorts have been integrated. The enhancer-associated regulatory network was constructed after Enhancer Linking by Methylation/Expression Relationships (ELMER) analysis. An enhancer-associated prognostic gene signature has been constructed with the screened target transcription factor and target genes, exhibiting good prognostic prediction performance for patients with lung cancers. The complete analysis route of this study has been provided in Figure 1.

Materials and methods

Datasets

Three cohorts were applied: The Cancer Genome Atlas (TCGA)-LUAD, GSE31210, and GSE8894. Both the transcriptome and methylation data of TCGA-LUAD were retrieved from the UCSC Xena website, and the transcriptome data of GSE31210 and GSE8894 were obtained from the Gene Expression Omnibus (GEO) database. The clinical features were also obtained, and the data from patients with overall survival of <30 days were removed. All transcriptomic data were normalized with $\log_2$ (x+1) method and the combat function of the sva R package was used to exclude batch effects (Leek et al., 2012). The methylation data were normalized with the champ.norm function of the ChAMP R package, and the missing values in methylation data have been filled with function impute.knn (Tian et al., 2017).

Enhancer-associated regulatory network

With the gene methylation and transcription levels of the same batch of samples in the TCGA-LUAD cohort, the sequences of the differential methylation probes were screened with the ELMER R package (Silva et al., 2019), found the enriched motifs, and further predicted the TFs interacting with these motifs. Finally, the "TFs-motifs-genes" regulatory network was constructed. ELMER's main analysis process consisted of the following five parts: 1) Identification of distal probes (probes larger than 2 kb upstream from the transcription start site) from methylation chip data; 2) identification of differences in methylation levels between normal and tumor groups; 3) identification of target genes for differentially methylated probes; 4) identification of motifs enriched with both differentially methylated and target gene-related probes; 5) identification of TFs based on transcriptional differences.

Identification of prognostic regulatory network and tumor subtypes

Based on the transcriptional level of genes included in the regulatory network, we first conducted the univariate analysis to screen prognostic genes, and then performed the unsupervised clustering in the three cohorts with the ConsensusClusterPlus R package by the k-means method (Wilkerson and Hayes, 2010). The clustering process was carried out 1,000 times, involving 80% samples in each iteration. Subsequently, survival analyses were performed for different subtypes.

Construction of enhancer-associated prognostic signature

TCGA-LUAD was applied as the training set, GSE31210 and GSE8894 were applied as the validation sets. Least absolute shrinkage and selection operator (LASSO) and Stepwise regression analyses were applied to further streamline prognostic variables in the TCGA-LUAD cohort and construct a multi-gene COX signature for calculating the risk score of each patient (Tibshirani, 1997). Patients were divided into high-risk and low-risk groups based on the median risk score. Then Kaplan–Meier (KM) survival curves and Receiver Operating Characteristic (ROC) curves were plotted to assess the prediction effect of the model. Independent prognostic analyses were applied for validating the independence of the risk score compared with other clinical features in the three cohorts. The Wilcoxon rank sum test was applied to evaluate correlations between risk score and clinical features in TCGA-LUAD cohort.

Analysis of tumor mutation burden (TMB), tumor immune dysfunction and exclusion (TIDE), and immune micro-environment.

The mutation data of TCGA-LUAD based on VarScan2 (Koboldt et al., 2012) were obtained from TCGA Database. Non-synonymous mutations were calculated and the TMB scores were obtained with the number of variants/the length of exons. Wilcoxon test was used to analyze the difference in TMB values. The Maftools R package was used to calculate and plot the somatic alterations landscapes (Mayakonda et al., 2018). The TIDE scores of the TCGA-LUAD cohort were obtained from TIDE (Jiang et al., 2018). The differences in the immune micro-environment were compared with the proportions of 22 types of immune cells estimated by Cibersort (Chen et al., 2018). The Cibersort R package was applied. The simulation was conducted 1,000 times with the parameter of perm = 1,000, QN = True. The samples with $p > 0.05$ were rejected and removed. The correlation between risk scores and various immune cell infiltration was further analyzed with the Spearman correlation test.

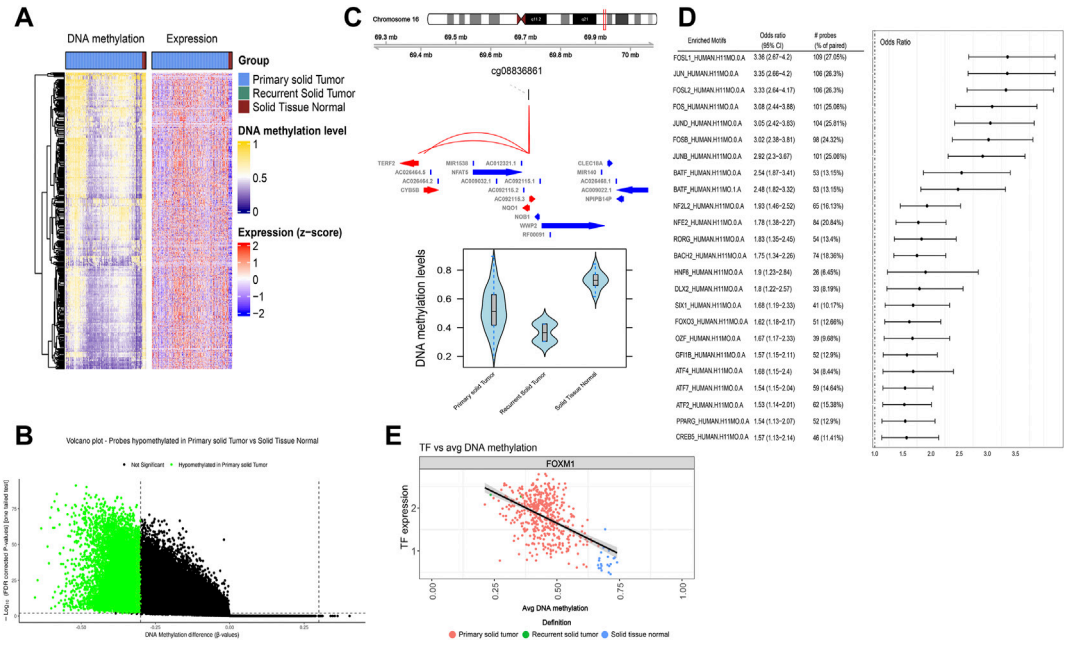


FIGURE 2
Enhancer-associated regulatory network was constructed with ELMER. **(A)** The heatmap of methylation and transcriptomic data in the TCGA-LUAD cohort. **(B)** Volcano plot of probes hypomethylated in primary tumor tissues. **(C)** Example of top ten genes closest to the upstream and downstream of the differentially methylated distal probes. **(D)** Odds ratios of the significantly enriched motifs identified by the get.enriched.motif function. **(E)** Example of correlation plot between the TF expression level and corresponding average DNA methylation level.

Analysis of tumor immunogenicity and drug susceptibility

The tumor immunogenicity was analyzed in patients from high-risk and low-risk groups, which was divided by the median value of the risk score. The immunophenotype scores (IPS) of TCGA-LUAD cohort were downloaded from The Cancer Immunome Atlas database (https://tcia.at/home). Based on the expression status of CTLA4 and PD1 (Charoentong et al., 2017), the high-risk and low-risk groups were further classified into four subgroups: positive CTLA4 and positive PD1; positive CTLA4 and negative PD1; negative CTLA4 and positive PD1; negative CTLA4 and negative PD1. The Wilcoxon nonparametric test was used to compare the differences in IPS between high-risk and low-risk groups in each subgroup. The drug susceptibility has been explored in patients from high-risk and low-risk groups. The pRRophetic R package was applied to analyze IC50 values of six commonly used drugs (Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine, and Paclitaxel).

Integration of the enhancer-associated prognostic nomogram

The predictive efficacy of the risk score for other clinical symptoms was assessed by ROC curves for 1, 3, and 5 years of survival. The risk score and clinical features (Gender, Age, Stage, Prior-malignancy) were integrated into a nomogram using the "rms" R package. The ROC curves of the nomogram and clinical features for the 5-year survival were plotted with the survivalROC R package. The performance of the nomogram was also confirmed by both the KM and calibration curves.

Statistics

Statistical analysis was conducted with the R 4.0.3. The survival analyses were performed with the log-rank test. The comparison between the two groups was executed with the Wilcoxon test or *t*-test. $p < 0.05$ was considered to be significant.

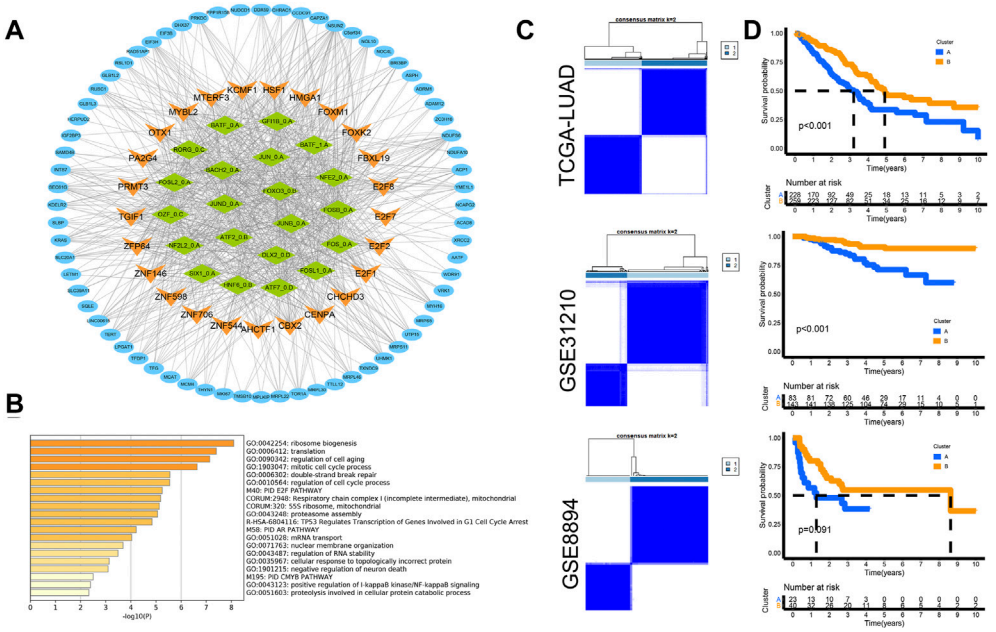


FIGURE 3
Prognostic network was constructed and contributed to tumor subtypes. **(A)** Prognostic regulatory network visualized with the Cytoscape software. **(B)** Function enrichment performed by Metascape. **(C)** Heatmaps of unsupervised consensus matrixes in the three cohorts. **(D)** KM curves revealed significant survival differences in all three cohorts. Log-rank test.

Results

Construction of enhancer-associated regulatory network with ELMER

Both methylation and transcriptomic data of 463 primary tumor tissues and 21 normal controls from the TCGA-LUAD cohort were fed into subsequent ELMER analysis. The heatmap of methylation and transcriptomic data is plotted (Figure 2A). Distal probes are the region where enhancers are enriched. Based on the hg38 reference genome file, we first select 13,1371 distal probes with the get.feature.probe function (Supplementary Material S1). Then, the unsupervised mode is used to identify distal hypomethylated probes with the get.diff.meth function. In detail, for every distal probe, the methylation levels are sorted in all samples within the primary tumor and the normal groups separately, and those samples in the lower quintile (20% samples with the lowest methylation levels) of each group are used to identify whether the probe is hypomethylated in the tumor group, thus obtaining 10,665 distal hypomethylated probes (Supplementary Material S2) with the threshold of false discovery rate (FDR) < 0.01 and $\Delta\beta < -0.3$ (Figure 2B). Next, the GetNearGenes function is applied to identify the top ten genes closest to the upstream and downstream of the distal hypomethylated probes separately, generating probe-gene pairs (Figure 2C). Then, for each probe-gene pair, the inverse correlations between the methylation level of the probe and the expression of the gene were tested. The top 20% and the bottom 20% of all the samples based on the probe's methylation level are extracted as the Methylated (M) group and Unmethylated (U) group. The gene expression levels between M and U groups are compared by the Mann-Whitney *U* test. 669 pairs of statistically significant probe-gene pairs with negative correlations are screened by default parameters using the get.pair function in an unsupervised mode (Supplementary Material S3). Further, the 250 bp base sequence upstream and downstream of the probes screened in the previous step are extracted, mapped to 768 motifs (Supplementary Material S4), and identified 24 significantly enriched motifs by the get.enriched.motif function (Figure 2D). Finally, based on the methylation level, the distal probes corresponding to the same motif were classified as the top 20% M group and the bottom 20% U group. A total of 127 TFs (Lambert et al., 2018) with differential expressions in the two groups and negative correlations with methylation levels are screened by using the get.TFs function in an unsupervised mode (Figure 2E), corresponding to 21 motifs (Supplementary Material S5).

Construction of prognostic network and its contribution to tumor subtypes

With the 21 motifs as links, 127 TFs and 271 target genes are screened in the enhancer-associated regulatory network.

TABLE 1 Clinical characteristics.

Feature	TCGA-LUAD (N = 487)	GSE31210 (N = 226)	GES8894 (N = 63)
Age			
>65	230 (47.23%)	50 (22.12%)	18 (28.57%)
≤65	247 (50.72%)	176 (77.88%)	43 (68.25%)
Unknown	10 (2.05%)	NA	2 (3.17%)
Gender			
Male	226 (46.41%)	105 (46.46%)	34 (53.97%)
Female	261 (53.59%)	121 (53.54%)	29 (46.03%)
Stage			
I	262 (53.80%)	168 (74.34%)	NA
II	114 (23.41%)	58 (25.66%)	NA
III	79 (16.22%)	NA	NA
IV	25 (5.13%)	NA	NA
Unknown	7 (0.01%)	NA	NA

NA: Not Available.

According to expression data in the TCGA-LUAD cohort, 25 TFs and 80 target genes are selected as prognostic genes with the univariate analysis $p < 0.05$ (Supplementary Material S6). Then, the prognostic regulatory network is visualized with the Cytoscape software (Figure 3A). The function enrichment is performed with the Metascape (Zhou et al., 2019) webtool (https://metascape.org/). The results show that the regulatory network mainly affected ribosome biogenesis, translation, cell aging, cell cycle, E2F pathway, and so on (Figure 3B). Then, based on the transcriptional data of the regulatory network, unsupervised clustering analysis is conducted in the three cohorts respectively. Two stable subtypes are obtained (Figure 3C), and the KM curves reveal significant survival differences (Figure 3D).

An eight-gene enhancer-associated prognostic signature

Next, TCGA-LUAD is applied as the training set, with GSE31210 and GSE8894 as the validation sets. Relevant clinical features are presented (Table 1). With the 105 genes in the above regulatory network as the initial variables, we subsequently conduct Lasso regression analysis (Figure 4A) and Stepwise regression analysis, and finally obtain the Cox model consisting of genes. The risk score = (0.3486 × E2F7) + (0.3011 × EIF3H) + (0.2536 × INTS1) + (0.3019 × LPGAT1) + (0.3078 × MCAT) + (0.3006 × SEC61G) + (−0.6044 × SS18L1) + (−0.2837 × TNYN1). The forest plot shows the hazard ratio (HR) of each gene, and the model's concordance index reached 0.7 (Figure 4B). Detailed results of variables within the model are shown in Table 2. The AUC of ROC curves for 1-, 3-, and 5-years survival are 0.764, 0.709, and 0.635 in the TCGA-LUAD cohort, and ranged from 0.633 to 0.71 in the independent validation cohorts GSE31210 and GSE8894 (Figure 4C). KM curves show significant differences between the high-risk and low-risk groups in all three cohorts (Figure 4D). The expression heatmaps of eight genes involved in the model and the risk curves of patients were also visualized. With the increase of the risk scores, the survival statuses become worse, and the expression levels of SS18L1 and THYN1 are gradually decreased, while the expression levels of the other six genes are gradually increased (Figure 4E). The results are consistent with the HR value of each gene in the model. Based on the results of multi-cox prognostic analyses, the risk score is an indicator independent of other factors ($p < 0.05$) for predicting the survival in all three cohorts (Figure 5A). Correlation analysis with clinical traits shows that with the increase of tumor stages and degrees of metastasis, the risk scores exhibit a gradually increasing trend, and the risk scores of patients with PD responses to the primary therapy exhibit significant improvement compared to those with complete response (CR), partial response (PR), and stable disease (SD) responses (Figure 5B).

Differences in TMB, TIDE, and immune micro-environment

TMB score represents the density of non-synonymous mutation distribution in the protein-coding region, which is calculated with the number of non-synonymous mutation sites/the total length of exons. TMB is an effective index to predict the response of immunotherapy, and higher TMB indicates a better immune response. TIDE is the

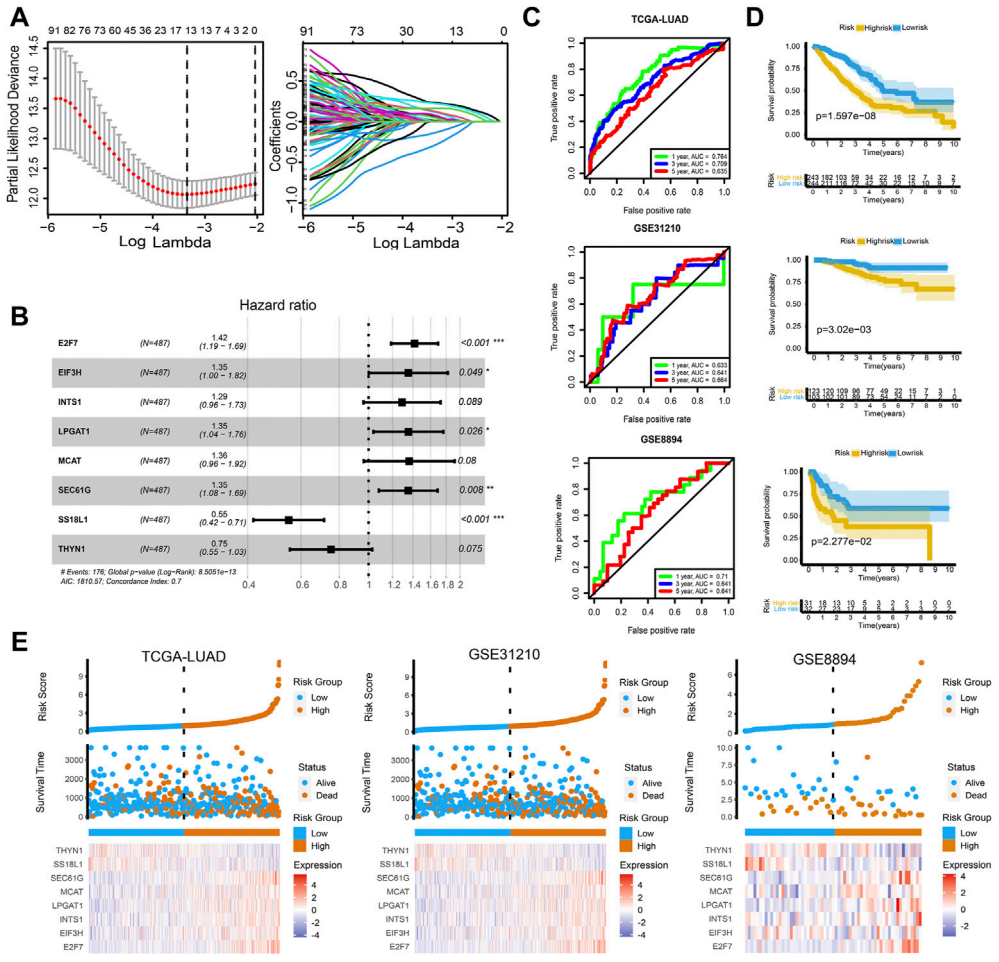


FIGURE 4
An eight-gene enhancer-associated prognostic signature. **(A)** Lasso regression analysis result. Partial Likelihood Deviance profile (left) and coefficients profile (right) changing with the log lambda. **(B)** Forest plot of each gene's hazard ratio (HR) and the model's concordance index. **(C)** ROC curves predicting 1-, 3-, and 5-years survival in the TCGA-LUAD, GSE31210, and GSE8894 cohorts. **(D)** KM curves showed that there were significant differences in survival between the high- and low-risk groups in all three cohorts. **(E)** The expression heatmaps of eight genes in the model and the patients' risk factor correlation curves.

TABLE 2 Detailed results of variables within the enhancer-associated signature.

ID	Coef	HR	HR.95 L	HR.95H	*p*-value
E2F7	0.3486	1.4171	1.1862	1.6930	1.22E-04
EIF3H	0.3011	1.3514	1.0013	1.8239	4.90E-02
INTS1	0.2536	1.2886	0.9622	1.7258	8.89E-02
LPGAT1	0.3019	1.3524	1.0376	1.7627	2.55E-02
MCAT	0.3078	1.3604	0.9639	1.9200	7.99E-02
SEC61G	0.3006	1.3507	1.0811	1.6876	8.13E-03
SS18L1	-0.6044	0.5464	0.4179	0.7144	9.90E-06
THYN1	-0.2837	0.7530	0.5511	1.0287	7.47E-02

computational algorithm for assessing tumor microenvironment from gene expression profiles. TIDE involves a set of gene expression markers to evaluate tumor immune evasion, including dysfunction of tumor-infiltrating cytotoxic T lymphocytes (CTL) and exclusion of CTL by immunosuppressive factors. TIDE is a quantitative index of immune escape, and higher TIDE indicates more serious the immune escape of the tumor.

To explore the characteristics of the immune microenvironment in patients with high and low risks, we compare the TMB values, TIDE scores, and 22 types of immune cell infiltration in these two groups. The high-risk group shows significantly higher TMB values (Figure 6A), and

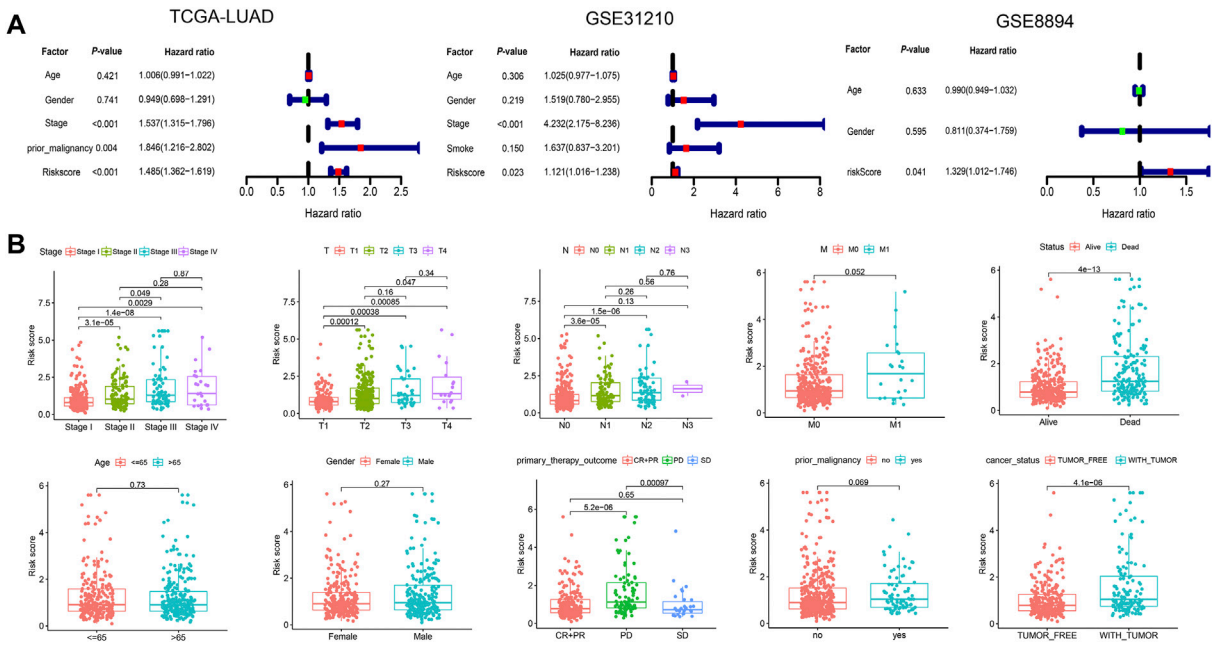


FIGURE 5
The risk score was an independent factor for survival. **(A)** Multi-cox analyses of the risk score and other clinical features in the three cohorts. **(B)** Differential analyses of the risk scores in patients with different clinical traits. With the increase of tumor stages and degrees of metastasis, the risk scores showed a gradually increasing trend, and the risk scores of patients with PD responses to the primary therapy were significantly improved compared to those with complete response (CR), partial response (PR), and stable disease (SD) responses.

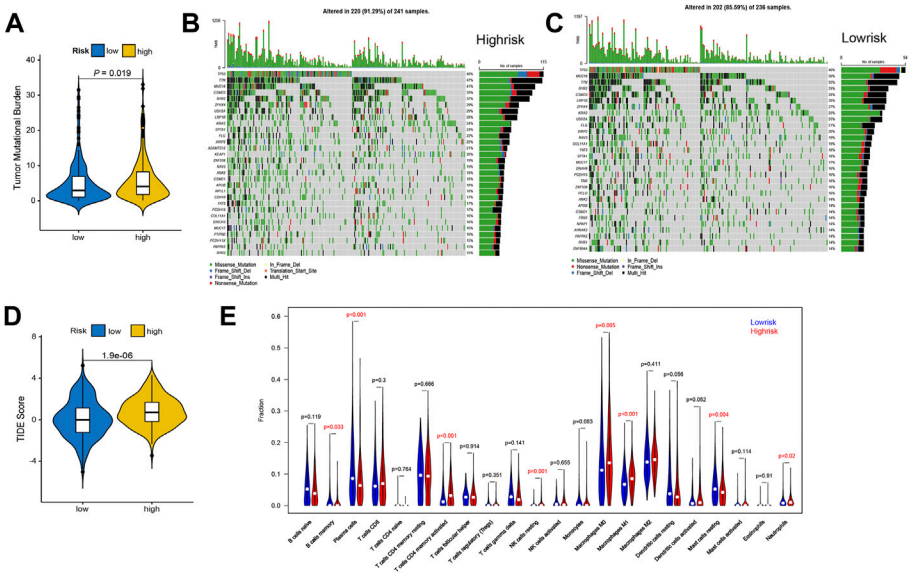


FIGURE 6
Differences of tumor mutation burden and immune micro-environment in high- and low-risk patients. **(A)** The violin diagram shows the TMB difference between the two groups. **(B)** Mutation profile of the top 30 genes with the biggest mutation frequency in high-risk samples. **(C)** Mutation landscape of the top 30 genes with the biggest mutation frequency in low-risk samples. **(D)** Violin diagram of the TIDE scores in the two groups. **(E)** The violin diagram of the differences between two groups in the infiltration of 22 immune cells.

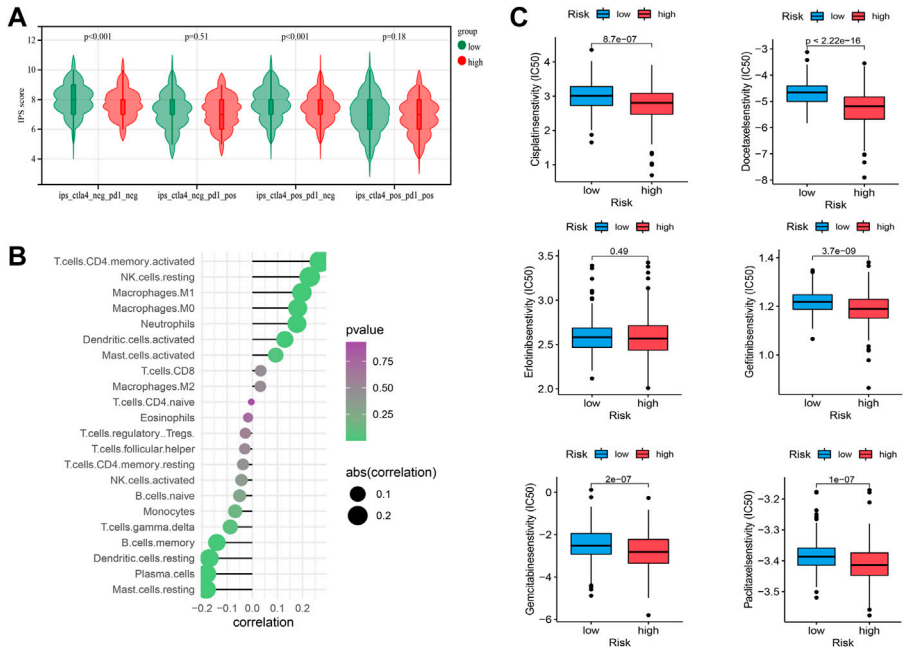


FIGURE 7
Tumor immunogenicity and drug susceptibility. **(A)** IPS score in low-risk and high-risk groups, which were classified into four subgroups: positive CTLA4 and positive PD1; positive CTLA4 and negative PD1; negative CTLA4 and positive PD1; negative CTLA4 and negative PD1, respectively. **(B)** The correlation between risk scores and various immune cell infiltration. **(C)** The IC50 values of Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine, and Paclitaxel in low-risk and high-risk groups.

higher percentages of gene mutations (Figure 6B), compared to that of the low-risk group (Figure 6C). The high-risk group also presents higher TIDE scores than the low-risk group (Figure 6D). Eight of the 22 immune cell types show statistical differences in the infiltration degree. Among them, macrophage M0, macrophage M1, and T cell CD4 memory activated significantly increase in the high-risk group, while plasma cells and mast cells resting decrease significantly (Figure 6E). T cell CD4 memory activated exhibits a significant positive correlation with the risk score, while mast cells resting presents a significant negative correlation with the risk score (Figure 7B).

Tumor immunogenicity and drug susceptibility

The IPS is included for evaluating tumor immunogenicity. Higher IPS suggested higher immunogenicity, exhibiting a potential higher response rate to immunotherapy. The IPS score of the low-risk group is significantly higher than that of the high-risk group in subgroups of positive CTLA4 and positive PD1, negative CTLA4 and negative PD1 (Figure 7A). The results suggest that the patients in the low-risk group show a better response to immunotherapy. The IC50 values of six commonly used drugs (Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine and Paclitaxel) are calculated in low-risk and high-risk groups. For all six drugs except for Erlotinib, LC50 is significantly higher in low-risk groups than that of high-risk groups (Figure 7C). It suggests that the patients in the high-risk group may be more sensitive to these drugs, which is accompanied by the high potential of immunosuppression.

An integrated enhancer-associated prognostic nomogram

We first compare the predictive accuracy of the risk score with various clinical traits (Age, Gender, Stage, and Prior-malignancy status). ROC analyses of multiple indicators show that the risk score is more accurate compared to other clinical traits in predicting 1- and 3-years survival (Figure 6B–Figure 8A), and slightly lower than Stage in predicting 5-years survival (Figure 8C). Then, a nomogram including all the variables in the TCGA-LUAD cohort is integrated with multi-cox regression analysis (Figure 8D). The nomogram exhibited the highest accuracy in the ROC curve for predicting 5-years survival, higher than the risk score alone (Figure 8E). And the performance is also validated with the KM curve (Figure 8F) and the calibration curve (Figure 8G).

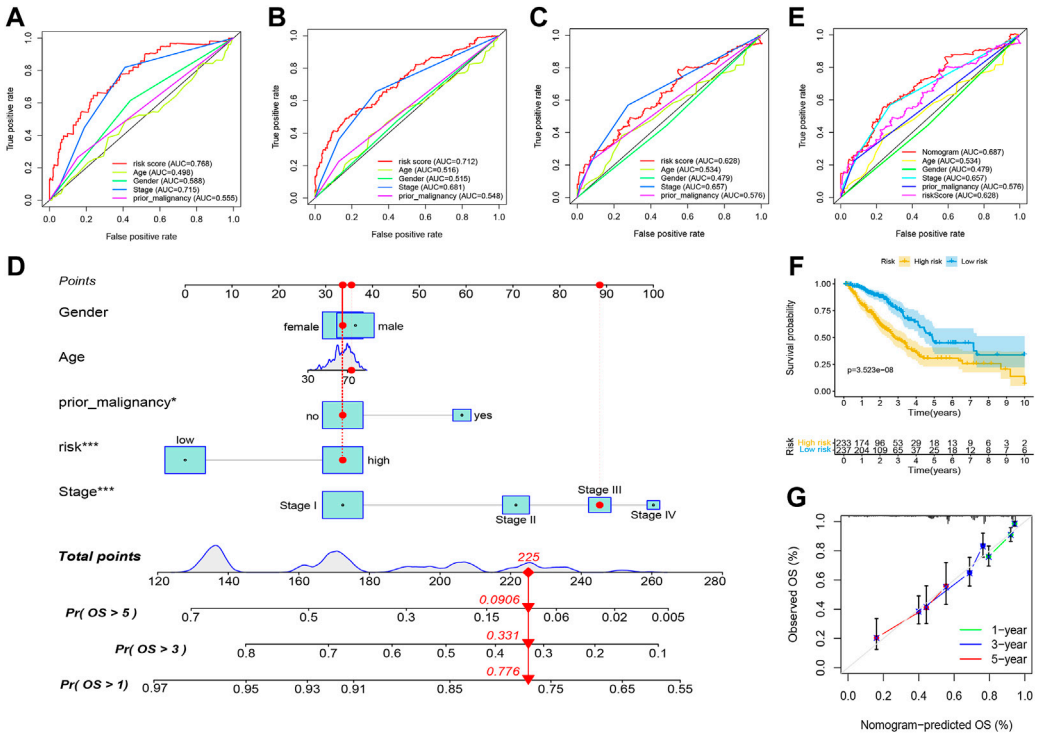


FIGURE 8
Integration of the enhancer-associated prognostic nomogram. **(A)** ROC analyses of the risk score with other clinical traits (Age, Gender, Stage, and Prior-malignancy status) in predicting 1-year survival. **(B)** ROC analyses of multiple indicators in predicting 3-years survival. **(C)** ROC analyses of multiple indicators in predicting 5-years survival. **(D)** Nomogram including all the variables in the TCGA-LUAD cohort integrated with multi-cox regression analysis. **(E)** ROC analyses of the nomogram with other clinical traits in predicting 5-years survival. **(F)** KM curve of the high- and low-risk patients distinguished by the nomogram. **(G)** Calibration curve of the nomogram for the survival prediction.

Discussion

Lung cancer has been a common type of cancer (Mao et al., 2016). Its prognosis has been generally poor, especially for the advanced stages (Jones and Baldwin, 2018). As a huge burden on society (Bade and Dela Cruz, 2020; Wu et al., 2021), the high incidence and mortality of lung cancer put forward higher and more urgent demands for its early diagnosis and treatment (Rodriguez-Canales et al., 2016). With the development of molecular biology tools, the genomic information of lung cancer has been better profiled at the molecular level, which provided new treatment options and improved outcomes for patients (Parikh, 2019). Several studies have tried to find diagnostic and prognostic prediction markers for lung cancers. One prospective cohort study included 426 patients with complete surgical resection of stages I to III LUAD reported that a computational machine-learning prediction model integrating genomic and clinicopathologic features could better predict the risk of recurrence, compared with the current TNM system. It would provide recommendations for adjuvant therapy after surgical resection of LUAD (Jones et al., 2021). Another study performed on LUAD patients in eastern China detected molecular alterations with a customized DNA panel. Some of the specific mutations may make effects on the efficacy of targeted therapies, CCAAT enhancer binding protein alpha (*CEBPA*) mutations affected the efficacy of EGFR-tyrosine kinase inhibitors. The erb-b2 receptor tyrosine kinase 2 (*ERBB2*), CEBPA and transcription factor 7 like 2 (*TCF7L2*) mutated tumors tend to have higher TMB. The targeted DNA panel may be helpful for personalized treatment decisions of LUAD patients (Liu et al., 2021).

For better understanding the molecular characteristics of lung cancers, several studies have involved various bioinformatic tools to comprehensively analyze the omics data based on microarray or sequencing analysis of patients. Based on the microarray datasets of three cohorts of lung cancer, a meta-analysis has been performed. There were 50 upregulated and 87 downregulated genes overlapped in three datasets, which were included in following analysis. With the protein-protein interaction (PPI) networks, 22 core genes were identified, which were all significantly associated with poor survival. Finally, KEGG pathway enrichment reanalysis screened five key genes, which exhibit a relationship with certain drugs. The identified key genes

can be candidate targets for both the treatment and prognosis of lung cancer (Wang et al., 2021). With the synthetic analysis of the transcriptome sequencing dataset and a non-coding RNA sequence dataset of small-cell lung cancer, the differentially expressed genes and miRNAs can be screened. After function enrichment, the molecular mechanisms were identified with the PPI network. Finally, 19 overlapping target genes and 32 corresponding regulatory miRNAs were screened. The bioinformatics analysis involving multi-omics data can assist in exploring the roles of target genes, miRNA, and TFs, which may better understand the potential molecular pathways (Mao et al., 2019).

After screening the potential candidate genes based on bioinformatics analysis, they can be further validated experimentally. A comprehensive bioinformatics analysis revealed that, the decreased expression of immunoglobulin superfamily member 10 (IGSF10) was associated with the shortened overall survival duration of patients with lung cancer. In subsequent experimental validation, IGSF10-knockout cells presented significantly increased proliferation and adhesion capability, revealed by MTT, colony formation assay, and Transwell assay, respectively. Further, Western blotting suggested that, the IGSF10-knockout can activate the integrin-β1/FAK pathway, presented as the upregulated protein expression levels of integrin-β1, phosphorylated (p)-FAK and p-AKT (Ling et al., 2020). Another study has analyzed the RNA sequencing data and revealed circXPO1, a novel circular RNA (circRNA) in LUAD. The circXPO1 was derived from a well-established cancer therapeutic target, XPO1, which was highly expressed in LUAD tissues compared with paired controls. High circXPO1 expression was correlated with worse overall survival. Mechanically, circXPO1 could bind with IGF2BP1 and enhance CTNNB1 mRNA stability, and subsequently promote LUAD progression (Huang et al., 2020). A similar study verified that circ-CAMK2A enhanced LUAD metastasis by regulating the miR-615-5p/fibronectin one pathway. Circ-CAMK2A upregulated the expression level of fibronectin one by sponging miR-615-5p, thus promoting MMP2 and MMP9 expression to stimulate the metastasis of LUAD (Du et al., 2019). In short, the combination of bioinformatics analysis and experimental verification can better clarify the significance of certain target biomarkers or promising pathways.

In our study, the comprehensive multi-omics analysis has also been applied to screen the target genes, which may be significant to prognosis prediction of lung cancer. The transcriptome and methylation data of the TCGA-LUAD cohort involving 463 primary tumor tissues and 21 normal controls were obtained for subsequent ELMER analysis. A total of 127 TFs corresponding to 21 motifs and 271 target genes were screened for constructing the subsequent enhancer-associated regulatory network. 25 TFs and 80 target genes were selected as prognostic genes with the univariate analysis $p < 0.05$. With TCGA-LUAD as the training set, the Cox model involving eight genes was selected with LASSO regression analysis and Stepwise regression analysis. The risk score = (0.3486 × E2F7) + (0.3011 × EIF3H) + (0.2536 × INTS1) + (0.3019 × LPGAT1) + (0.3078 × MCAT) + (0.3006 × SEC61G) + (−0.6044 × SS18L1) + (−0.2837 × TNYN1). With the increase of the risk scores, the survival statuses became worse, as well as the clinical traits including tumor stages, metastasis degree, and treatment responses. The risk score exhibited prognostic prediction accuracy with GSE31210 and GSE8894 as the validation sets.

The multi-omics study has been proved as a good tool for the epigenetic regulation of functional enhancers. In a study on HCC, methyl-binding DNA capture sequencing was firstly performed on both tumor and control tissues. The data revealed abnormal enhancer hypermethylation patterns. Then, the single-base resolution whole-genome bisulfite sequencing (WGBS) was performed to screen enhancers with differential methylation. Then, CCAAT/enhancer-binding protein-beta (C/EBPβ) enhancer was selected for further function mechanism. The survival analysis indicated that hypomethylation of C/EBPβ enhancer was related to the poor prognosis of patients with HCC. The mechanism has been also investigated experimentally (Xiong et al., 2019). This study has inspired several studies to perform multi-omics analysis (Cui et al., 2021). By involving methylome, transcriptome, and 3D genomic data, the researchers comprehensively analyzed enhancer methylation regulome and identified enhancer methylation-enhancer TF-target gene expression. They found that the enhancer-regulated core TFs could further shape their enhancer methylation, thus forming the enhancer methylation-driven core transcriptional regulatory circuitries, which can be served as innovative therapy targets and prognostic risk biomarkers (Pan et al., 2022). In another study integrating ChIP-seq, RNA-seq, and WGBS data, the enhancers with differential expression and differential methylation were identified, as well as the associated differentially expressed genes. A model based on six enhancer-associated genes was constructed with regression analysis, exhibiting excellent predictive accuracy (Huang et al., 2022). In addition to screening biomarkers, the combination of epigenetic and transcriptional data can also demonstrate the mechanism. For example, the aberrant methylation of promoters and enhancers could activate critical cell cycle-related pathways and inhibit several metabolic pathways, thus affecting the progression of HCC (Huang et al., 2021). In short, the integrative analysis of multi-omics data can help us find new and more effective function targets in various diseases.

One comprehensive study has summarized more than 30 bioinformatics approaches for enhancer identification. With the advances in biological technologies, several data resources have been involved for screening enhancers, such as evolutionary conservation data, histone marks, Open chromatin, Transcription factor-binding sites, Sequencing features, Screening data, and eRNA expression. These data types can be combined in different ways to generate feature vectors that describe DNA regions. After feature selection, the feature vectors feed computational models that make

decisions using unsupervised and/or supervised algorithms, such as Clustering, Classification, Graphical models, and Regression. The outcome is a list of identified enhancer regions. However, one of the major challenges is how to assess the correctness of predicted enhancers, because there is no large, sufficiently comprehensive, and experimentally validated enhancer set for humans. One possible way of validation is to link the predicted enhancers to their target genes (Kleftogiannis et al., 2016). The ELMER analysis applied in this study has involved five steps, which can construct a "TFs-motifs-genes" regulatory network. The formation of a regulatory network integrating TF, motif, and target gene further guaranteed the correctness of predicted candidates. However, one limitation of this study may be the lack of experimental evidence for validation. The prognostic prediction signature for LUAD has been only validated in patients from another two LUAD cohorts.

Conclusion

This study has comprehensively analyzed the transcriptome and methylation data of a LUAD cohort. ELMER analysis has been performed to screen motifs, motif-associated TFs, and target genes. The "TFs-motifs-genes" regulatory network was constructed. After regression analysis, the Cox model involving eight genes was constructed. The enhancer-associated prognostic gene signature can be applied as a risk score for predicting the survival status of patients. With the increase of the risk scores, both the survival statuses and clinical traits showed a worse tendency in patients with lung adenocarcinoma. The multi-omics bioinformatics analysis can be a good tool for obtaining more information at the epigenetic level.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: www.ncbi.nlm.nih.gov/geo/, GSE31210, GSE8894.

Author contributions

SH, SC, and LL performed the conception and design of this study. DZ and JG collected the datasets from databases and contributed to the organization and typesetting of the figures. SH and SC conducted the bioinformatic analyses. SH and LL wrote the initial edition of the manuscript. SC and LL checked the statistical results. All the authors read and approved the submitted article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1008602/full#supplementary-material

SUPPLEMENTARY MATERIAL S1
131,371 distal probes selected with the get.feature.probe function.

SUPPLEMENTARY MATERIAL S2
10,665 distal hypomethylated probes identified with the get.diff.meth function in the tumor group.

SUPPLEMENTARY MATERIAL S3
669 pairs of statistically significant probe-gene pairs with negative correlations were screened by default parameters using the get.pair function in an unsupervised mode.

SUPPLEMENTARY MATERIAL S4
768 motifs enriched within the 250bp base sequence upstream and downstream of the probes in the 669 probe-gene pairs.

SUPPLEMENTARY MATERIAL S5
127 TFs with differential expressions in the two groups and negative correlations with methylation levels were screened by using the get.TFs function.

SUPPLEMENTARY MATERIAL S6
25 TFs and 80 target genes were selected as prognostic genes with the univariate analysis p value <0.05.

References

Adhikary, S., Roy, S., Chacon, J., Gadad, S. S., and Das, C. (2021). Implications of enhancer transcription and eRNAs in cancer. *Cancer Res.* 81 (16), 4174–4182. doi:10.1158/0008-5472.Can-20-4010

Bade, B. C., and Dela Cruz, C. S. (2020). Lung cancer 2020. *Clin. Chest Med.* 41 (1), 1–24. doi:10.1016/j.ccm.2019.10.001

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, H., and Liang, H. (2020). A high-resolution map of human enhancer RNA loci characterizes super-enhancer activities in cancer. *Cancer Cell* 38 (5), 701–715. e705. doi:10.1016/j.ccell.2020.08.020

Chen, J., Yang, H., Teo, A. S. M., Amer, L. B., Sherbaf, F. G., Tan, C. Q., et al. (2020). Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 52 (2), 177–186. doi:10.1038/s41588-019-0569-6

Cui, D., Li, W., Jiang, D., Wu, J., Xie, J., and Wu, Y. (2021). Advances in multi-omics applications in HBV-associated hepatocellular carcinoma. *Front. Med.* 8, 754709. doi:10.3389/fmed.2021.754709

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis.* 9 (2), 117. doi:10.1038/s41419-017-0063-y

Du, J., Zhang, G., Qiu, H., Yu, H., and Yuan, W. (2019). The novel circular RNA circ-CAMK2A enhances lung adenocarcinoma metastasis by regulating the miR-615-5p/fibronectin 1 pathway. *Cell Mol. Biol. Lett.* 24, 72. doi:10.1186/s11658-019-0198-1

Field, A., and Adelman, K. (2020). Evaluating enhancer function and transcription. *Annu. Rev. Biochem.* 89, 213–234. doi:10.1146/annurev-biochem-011420-095916

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., et al. (2017). GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017. doi:10.1093/database/bax028

Helgason, P. L., Parolia, A., Liu, H. H., and Helgason, C. D. (2016). DNA methylation at enhancer regions Novel avenues for epigenetic biomarker development. *Front. Biosci.* 21 (2), 430–446. doi:10.2741/4399

Herz, H. M. (2016). Enhancer deregulation in cancer and other diseases. *Bioessays* 38 (10), 1003–1015. doi:10.1002/bies.201600106

Huang, P., Xu, M., Han, H., Zhao, X., Li, M. D., and Yang, Z. (2021). Integrative analysis of epigenome and transcriptome data reveals aberrantly methylated promoters and enhancers in hepatocellular carcinoma. *Front. Oncol.* 11, 769390. doi:10.3389/fonc.2021.769390

Huang, P., Zhang, B., Zhao, J., and Li, M. D. (2022). Integrating the epigenome and transcriptome of hepatocellular carcinoma to identify systematic enhancer aberrations and establish an aberrant enhancer-related prognostic signature. *Front. Cell Dev. Biol.* 10, 827657. doi:10.3389/fcell.2022.827657

Huang, Q., Guo, H., Wang, S., Ma, Y., Chen, H., Li, H., et al. (2020). A novel circular RNA, circXPO1, promotes lung adenocarcinoma progression by interacting with IGF2BP1. *Cell Death Dis.* 11 (12), 1031. doi:10.1038/s41419-020-03237-8

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Jones, G. D., Brandt, W. S., Shen, R., Sanchez-Vega, F., Tan, K. S., Martin, A., et al. (2021). A genomic-pathologic annotated risk model to predict recurrence in early-stage lung adenocarcinoma. *JAMA Surg.* 156 (2), e205601. doi:10.1001/jamasurg.2020.5601

Jones, G. S., and Baldwin, D. R. (2018). Recent advances in the management of lung cancer. *Clin. Med.* 18 (2), s41–s46. doi:10.7861/clinmedicine.18-2-s41

Kleftogiannis, D., Kalnis, P., and Bajic, V. B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinform* 17 (6), 967–979. doi:10.1093/bib/bbv101

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22 (3), 568–576. doi:10.1101/gr.129684.111

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172 (4), 650–665. doi:10.1016/j.cell.2018.01.029

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Li, Y., Gu, J., Xu, F., Zhu, Q., Chen, Y., Ge, D., et al. (2021). Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. *Brief. Bioinform* 22 (4). doi:10.1093/bib/bbaa225

Ling, B., Liao, X., Huang, Y., Liang, L., Jiang, Y., Pang, Y., et al. (2020). Identification of prognostic markers of lung cancer through bioinformatics analysis and *in vitro* experiments. *Int. J. Oncol.* 56 (1), 193–205. doi:10.3892/ijo.2019.4926

Liu, J., Xu, W. Y., Ye, M., Liu, Z., and Li, C. (2021). Genetic alteration profiling of Chinese lung adenocarcinoma and its effect on targeted therapy efficacy. *Front. Oncol.* 11, 726547. doi:10.3389/fonc.2021.726547

Mao, Y., Xue, P., Li, L., Xu, P., Cai, Y., Chu, X., et al. (2019). Bioinformatics analysis of mRNA and miRNA microarray to identify the key miRNA-gene pairs in small-cell lung cancer. *Mol. Med. Rep.* 20 (3), 2199–2208. doi:10.3892/mmr.2019.10441

Mao, Y., Yang, D., He, J., and Krasna, M. J. (2016). Epidemiology of lung cancer. *Surg. Oncol. Clin. N. Am.* 25 (3), 439–445. doi:10.1016/j.soc.2016.02.001

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Mazieres, J., Drilon, A., Lusque, A., Mhanna, L., Cortot, A. B., Mezquita, L., et al. (2019). Immune checkpoint inhibitors for patients with advanced lung cancer and oncogenic driver alterations: Results from the IMMUNOTARGET registry. *Ann. Oncol.* 30 (8), 1321–1328. doi:10.1093/annonc/mdz167

Pan, X., Li, X., Sun, J., Xiong, Z., Hu, H., Ning, S., et al. (2022). Enhancer methylation dynamics drive core transcriptional regulatory circuitry in pan-cancer. *Oncogene* 41 (26), 3474–3484. doi:10.1038/s41388-022-02359-x

Parikh, A. R. (2019). Lung cancer genomics. *ama* 48 (1), 78–83. doi:10.5644/ama2006-124.244

Rodriguez-Canales, J., Parra-Cuentas, E., and Wistuba, II (2016). Diagnosis and molecular classification of lung cancer. *Cancer Treat. Res.* 170, 25–46. doi:10.1007/978-3-319-40389-2_2

Ruiz-Cordero, R., and Devine, W. P. (2020). Targeted therapy and checkpoint immunotherapy in lung cancer. *Surg. Pathol. Clin.* 13 (1), 17–33. doi:10.1016/j.path.2019.11.002

Silva, T. C., Coetzee, S. G., Gull, N., Yao, L., Hazelett, D. J., Noushmehr, H., et al. (2019). ELMER v.2: An R/bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* 35 (11), 1974–1977. doi:10.1093/bioinformatics/bty902

Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. *Nat. Rev. Cancer* 16 (8), 483–493. doi:10.1038/nrc.2016.62

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., et al. (2017). ChAMP: Updated methylation analysis pipeline for illumina BeadChips. *Bioinformatics* 33 (24), 3982–3984. doi:10.1093/bioinformatics/btx513

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Wang, Y., Zhou, Z., Chen, L., Li, Y., Zhou, Z., and Chu, X. (2021). Identification of key genes and biological pathways in lung adenocarcinoma via bioinformatics analysis. *Mol. Cell Biochem.* 476 (2), 931–939. doi:10.1007/s11010-020-03959-5

Wang, Z., Zhang, Q., Zhang, W., Lin, J. R., Cai, Y., Mitra, J., et al. (2018). Hedd: Human enhancer disease database. *Nucleic Acids Res.* 46 (D1), D113–d120. doi:10.1093/nar/gkx988

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, F., Wang, L., and Zhou, C. (2021). Lung cancer in China: Current and prospect. *Curr. Opin. Oncol.* 33 (1), 40–46. doi:10.1097/cco.0000000000000703

Xiong, L., Wu, F., Wu, Q., Xu, L., Cheung, O. K., Kang, W., et al. (2019). Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. *Nat. Commun.* 10 (1), 335. doi:10.1038/s41467-018-08245-z

Yi, M., Li, A., Zhou, L., Chu, Q., Luo, S., and Wu, K. (2021). Immune signature-based risk stratification and prediction of immune checkpoint inhibitor's efficacy for lung adenocarcinoma. *Cancer Immunol. Immunother.* 70 (6), 1705–1719. doi:10.1007/s00262-020-02817-z

Zhang, C., Zhang, Z., Sun, N., Zhang, Z., Zhang, G., Wang, F., et al. (2020). Identification of a costimulatory molecule-based signature for predicting prognosis risk and immunotherapy response in patients with lung adenocarcinoma. *Oncoimmunology* 9 (1), 1824641. doi:10.1080/2162402x.2020.1824641

Zhang, G., Shi, J., Zhu, S., Lan, Y., Xu, L., Yuan, H., et al. (2018). DiseaseEnhancer: A resource of human disease-associated enhancer catalog. *Nucleic Acids Res.* 46 (D1), D78–d84. doi:10.1093/nar/gkx920

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fgene.2022.1006636

OPEN ACCESS

EDITED BY
Wei Liu,
Arizona State University, United States

REVIEWED BY
Zhao Xiaoyun,
Shenyang Pharmaceutical University, China
Guoliang Zheng,
China Medical University, China
Yutian Zou,
Sun Yat-sen University Cancer Center (SYSUCC), China

*CORRESPONDENCE
Hua-chuan Zheng,
zheng_huachuan@hotmail.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

RECEIVED 29 July 2022
ACCEPTED 10 October 2022
PUBLISHED 21 October 2022

CITATION
Zheng H-c, Xue H, Zhang C-y, Shi K-h and Zhang R (2022), The roles of BTG1 mRNA expression in cancers: A bioinformatics analysis.
Front. Genet. 13:1006636.
doi: 10.3389/fgene.2022.1006636

The roles of *BTG1* mRNA expression in cancers: A bioinformatics analysis

Hua-chuan Zheng[1]*, Hang Xue[1], Cong-yu Zhang[2], Kai-hang Shi[3] and Rui Zhang[4]

[1]Department of Oncology, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Cancer Center, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China, [3]Department of Dermatology, The Affiliated Hospital of Chengde Medical University, Chengde, China, [4]Department of Colorectal Surgery, Liaoning Cancer Hospital, Shenyang, China

BTG1 (B-cell translocation gene 1) may inhibit proliferation and cell cycle progression, induce differentiation, apoptosis, and anti-inflammatory activity. The goal of this study was to clarify the clinicopathological and prognostic significances of *BTG1* mRNA expression and related signal pathways in cancers. Using the Oncomine, TCGA (the cancer genome atlas), xiantao, UALCAN (The University of ALabama at Birmingham Cancer data analysis Portal), and Kaplan-Meier plotter databases, we undertook a bioinformatics study of *BTG1* mRNA expression in cancers. *BTG1* expression was lower in gastric, lung, breast and ovarian cancer than normal tissue due to its promoter methylation, which was the opposite to *BTG1* expression. *BTG1* expression was positively correlated with dedifferentiation and histological grading of gastric cancer ($p < 0.05$), with squamous subtype and young age of lung cancer ($p < 0.05$), with infrequent lymph node metastasis, low TNM staging, young age, white race, infiltrative lobular subtype, Her2 negativity, favorable molecular subtyping, and no postmenopause status of breast cancer ($p < 0.05$), and with elder age, venous invasion, lymphatic invasion, and clinicopathological staging of ovarian cancer ($p < 0.05$). *BTG1* expression was negatively correlated with favorable prognosis of gastric, lung or ovarian cancer patients, but the converse was true for breast cancer ($p < 0.05$). KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis showed that the top signal pathways included cytokine-cytokine receptor interaction, cell adhesion molecules, chemokine, immune cell receptor and NF (nuclear factor)-κB signal pathways in gastric and breast cancer. The top hub genes mainly contained CD (cluster of differentiation) antigens in gastric cancer, FGF (fibroblast growth factor)-FGFR (FGF receptor) in lung cancer, NADH (nicotinamide adenine dinucleotide): ubiquinone oxidoreductase in breast cancer, and ribosomal proteins in ovarian cancer. *BTG1* expression might be employed as a potential marker to indicate carcinogenesis and subsequent progression, even prognosis.

KEYWORDS

Btg1, bioinformatics analysis, carcinogenesis, aggressiveness, prognosis

Introduction

BTG1 (B-cell translocation gene 1) is reported to suppress cell proliferation and cell cycle progression, and induce cell differentiation due to its interaction with the myogenic factor MyoD (Busson et al., 2005), protein arginine methyltransferase 1 (Lin et al., 1996), and human carbon catabolite repressor protein-associative factor 1 (Bogdan et al., 1998). BTG1 has also been demonstrated to promote Hoxb9-induced transcription to suppress proliferation in HeLa cells (Prévôt et al., 2000). Additionally, BTG1 mediates the apoptotic induction, as evidenced by BTG1 overexpression in apoptotic cells (Corjay et al., 1998) and the contribution of BTG1 to anti-sense Bcl-2- induced cytotoxicity (Nahta et al., 2006). Liu et al. (2015) found that BTG1 potentiated apoptosis and suppressed proliferation in renal clear cell carcinoma by interacting with PRMT1. BTG1 could reverse the miR-22-induced inhibition of autophagy (Zhang et al., 2015), while miR-4295 significantly promoted proliferation, colony formation, and migration of bladder cancer cell *via* directly targeting BTG1 (Nan et al., 2016). By inhibiting BTG1, miR-511 strengthened the proliferation of human hepatoma cells, while miR-301A promoted the development of colitis-associated cancer (He et al., 2017; Zhang et al., 2017). BTG1 functioned as a direct target of miR-330-3p, and miR-27a-3p in hepatocellular carcinoma and ovarian cancer cells, thereby weakened cell viability, migration and invasion, and promoted cell apoptosis (Li et al., 2019; Zhao et al., 2019). BTG1 was shown to prevent antigen from inducing molecular features of *in vitro* allergic reactions as a direct target of miR-183-5p (Kim et al., 2020). Ni et al. (2021) found that miR-141-5p enhanced the proliferation and inhibited apoptosis by targeting *BTG1* in cervical Cancer cells. Cheng et al. (2021) revealed that the exosomal miR-301a-3p promotes the proliferation and invasion of nasopharyngeal carcinoma squamous cells by targeting *BTG1*.

Su et al. (2019) reported that BTG1 overexpression triggered G_1/S phase cell cycle arrest and increased apoptosis in HCT-116 cells *via* the ERK (extracellular regulated protein kinases)/MEK (map kinase kinases) signaling pathway. Zhu et al. (2015) showed that BTG1 enhanced the radiation sensitivity of human breast cancer by inducing cell cycle arrest, the formation of reactive oxygen species, chromosomal aberrations and apoptosis *via* inhibition of the PI3K/Akt signaling pathway. BTG1 overexpression was also found to suppress proliferation, tumor growth and lung metastasis, induce differentiation, autophagy, and apoptosis, and mediate chemosensitivity in colorectal or gastric cancer cells (Zheng et al., 2015; Zhao et al., 2017). Xue et al. (2021) found that chidamide triggered BTG1-mediated autophagy and reverses the chemotherapy resistance in relapsed/refractory B-cell lymphoma. However, BTG1 overexpression was found to promote invasion and metastasis of colorectal cancer in our previous study (Zhu et al., 2014; Szász et al., 2016; Zhao et al., 2020).

In the present study, we aimed to clarify the clinicopathological and prognostic significances of *BTG1* mRNA expression in cancers by bioinformatics analysis of high-throughput cDNA array and RNA sequencing using online Oncomine, TCGA(the cancer genome atlas), xiantao, UALCAN (The University of ALabama at Birmingham Cancer data analysis Portal) and Kaplan-Meier plotter. According to the cancer types of Kaplan-Meier plotter (before 2018), we chose gastric, lung, breast and ovarian cancers for *BTG1* analysis.

Methods

Oncomine database analysis

The individual gene expression level of *BTG1* mRNA was analyzed using Oncomine (www.oncomine.org), a cancer microarray database and a web-based data mining platform for a new discovery from genome-wide expression analyses. We compared the differences in *BTG1* mRNA levels between normal tissue and cancer. All data were log-transformed, with median centered per array centered and standard deviation normalized to each array.

TCGA database analysis

The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) database was used to download expression data (RNA-seqV2) and clinicopathological data of gastric ($n = 392$), lung ($n = 865$), breast ($n = 1,093$), and ovarian ($n = 304$) cancer patients. We integrated the raw data, analyzed *BTG1* expression in the cancers, and compared it with clinicopathological and prognostic data from the cancer patients. A student t test was used to compare the means. Kaplan-Meier survival plots were generated with survival curves compared by log-rank statistic. Cox's proportional hazards model was employed for multivariate analysis. Two-sided $p < 0.05$ was considered statistically significant. SPSS 17.0 software was employed to analyze all the data.

GEO analysis

The mRNA expression profile of GSE38666 (platform: Affymetrix-GPL570) and GSE26712 (platform: Affymetrix-GPL96) was obtained from NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/), and R was used for the analysis of the

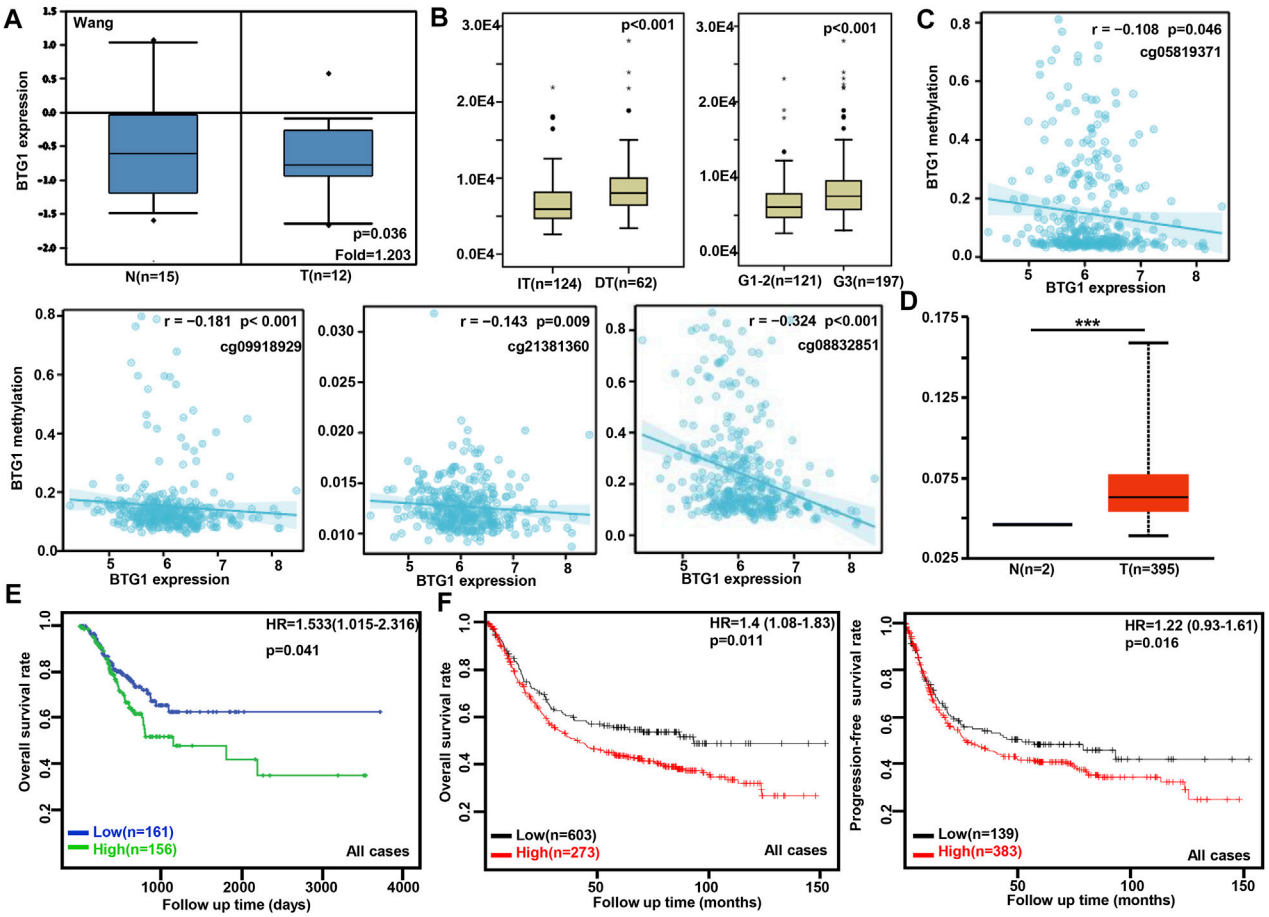


FIGURE 1
The clinicopathological and prognostic significances of *BTG1* mRNA expression in gastric cancer. Wang's **(A)** dataset was used for bioinformatics analysis to explore *BTG1* expression in gastric cancer. A lower *BTG1* expression was detectable in gastric cancer than that in normal mucosa ($p < 0.05$). TCGA database showed that *BTG1* was negatively correlated with Lauren's classification and histological staging of gastric cancer **(B)** $p < 0.05$. The negative relationship between *BTG1* expression and methylation was analyzed in gastric cancer using xiantao database **(C)**. Its methylation was higher in gastric cancer than normal tissues in UALCAN **(D)**. *BTG1* expression was negatively correlated with overall survival rate of the cancer patients by TCGA **(E)** $p < 0.05$. It was the same for the overall and progression-free survival rates of the patients with gastric cancer according to the data from Kaplan-Meier plotter **(F)** $p < 0.05$. Note: N, normal tissue; T, tumor; HR, hazard ratio; IT, intestinal-type; DT, diffuse-type.

BTG1 mRNA expression between ovarian cancer and normal tissues.

Kaplan-Meier plotter analysis

The prognostic significance of *BTG1* mRNA in gastric, lung, breast and ovarian cancers was also analyzed by the Kaplan-Meier plotter. (http://kmplot.com).

UALCAN analysis

The expression and methylation of *BTG1* gene and the relationship between BTG1 expression and immune cells infiltration were analyzed using UALCAN database (http://ualcan.path.uab.edu). They were also compared to clinicopathologic and prognostic parameters associated with gastric, lung, breast, and ovarian cancers.

Xiantao analysis

The expression and methylation of *BTG1* gene and the relationship between BTG1 expression and immune cells infiltration were analyzed using xiantao platform (https://www.xiantao.love/).

Additionally, we found the differential and related genes using xiantao. The PPI (protein-protein interaction) network was built using the differential genes, and the important hub genes were identified. These genes were submitted to KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis in order to build signal pathways.

TABLE 1 The prognostic significance of *BTG1* mRNA in gastric cancer.

Clinicopathological features	Overall survival		Progression-free survival	
	Hazard ratio	*p*-value	Hazard ratio	*p*-value
Sex				
Female	0.69 (0.45–1.07)	0.098	0.65 (0.43–0.99)	0.045
Male	1.65 (1.22–2.24)	0.001	1.45 (1.08–1.94)	0.012
T				
2	1.65 (1.08–2.53)	0.019	1.44 (0.9–2.32)	0.13
3	0.75 (0.51–1.08)	0.12	—	—
4	0.54 (0.2–1.48)	0.23	—	—
N				
0	1.98 (0.84–4.67)	0.11	1.68 (0.72–3.91)	0.22
1–3	1.29 (0.99–1.69)	0.062	1.24 (0.92–1.67)	0.16
1	1.49 (0.98–2.27)	0.063	1.33 (0.89–1.98)	0.16
2	0.72 (0.46–1.13)	0.15	0.79 (0.51–1.21)	0.28
3	2.07 (1.16–3.71)	0.013	1.71 (0.97–3.01)	0.061
M				
0	—	—	1.28 (0.94–1.75)	0.12
1	1.37 (0.99–1.91)	0.058	1.63 (0.85–3.13)	0.14
TNM staging				
I	1.67 (0.54–5.18)	0.37	0.5 (0.16–1.57)	0.23
II	2.5 (1.05–5.97)	0.032	1.89 (0.84–4.26)	0.12
III	0.81 (0.56–1.19)	0.28	0.74 (0.48–1.13)	0.16
IV	1.73 (1.14–2.65)	0.0099	1.38 (0.92–2.08)	0.11
Differentiation				
Well-differentiated	—	—	—	—
Moderately-differentiated	0.62 (0.32–1.21)	0.16	0.66 (0.35–1.24)	0.2
Poorly-differentiated	0.61 (0.37–1.01)	0.05	0.59 (0.37–0.94)	0.026
Lauren's classification				
Intestinal-type	1.49 (1.03–2.13)	0.031	1.38 (0.95–2.02)	0.092
Diffuse-type	1.18 (0.82–1.68)	0.38	0.81 (0.56–1.18)	0.27
Mixed-type	3.52 (1.11–11.15)	0.023	0.5 (0.18–1.35)	0.16
Her2 positivity				
–	1.27 (0.97–1.66)	0.078	1.23 (0.91–1.65)	0.18
+	1.73 (1.11–2.71)	0.015	1.42 (0.89–2.27)	0.14
Perforation				
–	0.57 (0.38–0.86)	0.0059	0.61 (0.42–0.9)	0.011
Treatment				
Surgery alone	1.56 (1.13–2.14)	0.0063	1.31 (0.96–1.77)	0.083
5-FU-based adjuvant	0.21 (0.06–0.71)	0.0058	0.16 (0.05–0.55)	0.00095
Other adjuvant	0.6 (0.25–1.47)	0.26	1.54 (0.69–3.44)	0.29

Results

The clinicopathological and prognostic significances of *BTG1* mRNA expression in gastric cancer

According to Wang's database, we found that *BTG1* mRNA expression was lower in gastric cancer than in normal tissues (Figure 1A, $p < 0.05$). In TCGA data, *BTG1* expression was positively correlated with dedifferentiation and histological grading of gastric cancer (Figure 1B, $p < 0.05$). Using the xiantao tool, we discovered a negative correlation between *BTG1* mRNA and its methylations (cg05819371, cg09918929, cg21381360, and cg08832851) (Figure 1C, $p < 0.05$). The UALCAN showed that *BTG1* methylation was higher in gastric cancer than in normal tissues (Figure 1D, $p < 0.05$). According to the TCGA data, *BTG1* mRNA

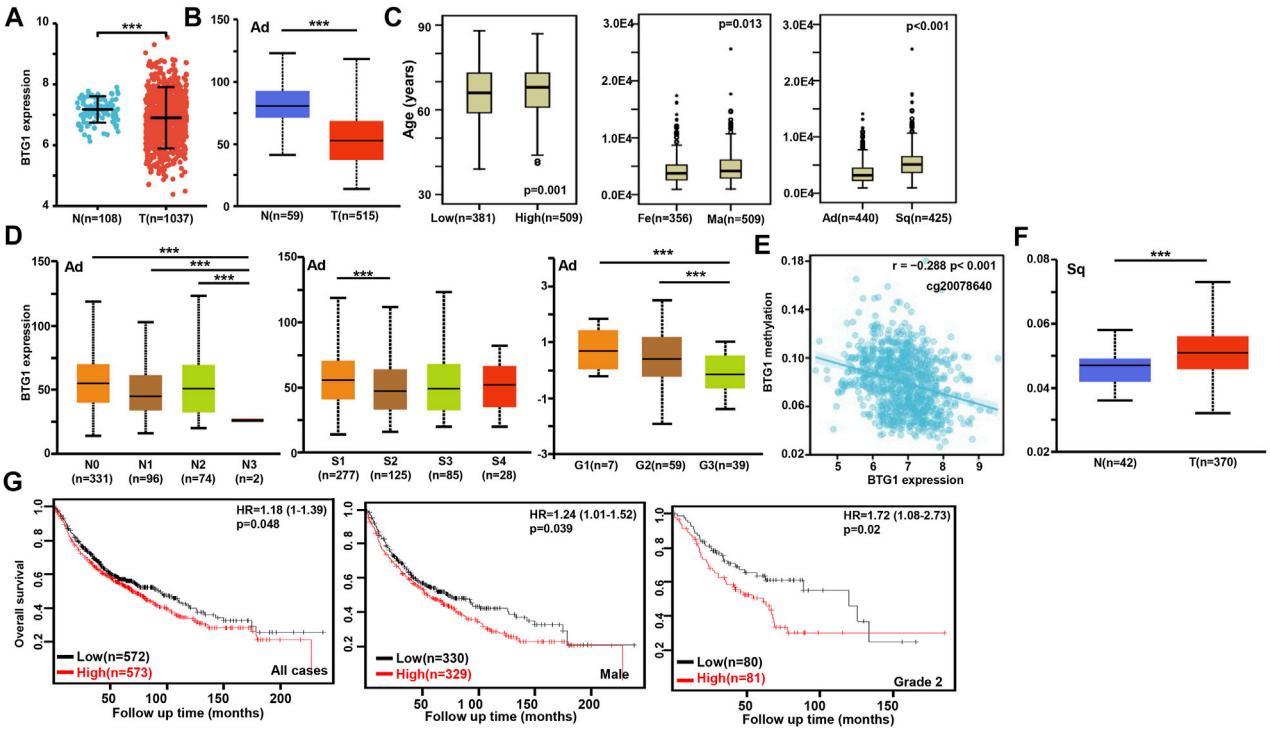


FIGURE 2
The clinicopathological and prognostic significances of *BTG1* mRNA expression in lung cancer. Xiantao **(A)** and UALCAN **(B)** datasets were employed for bioinformatics analysis to analyze BTG1 expression during lung carcinogenesis. *BTG1* expression was downregulated in lung cancer, compared with normal tissue ($p < 0.05$). *BTG1* expression was compared with age, gender and histological subtyping of the cancer patients by TCGA **(C)**, and with N staging, clinicopathological staging and histological grading by UALCAN **(D)**. The negative relationship between *BTG1* expression and methylation was analyzed in lung cancer using xiantao database **(E)**. Its methylation was higher in lung cancer than normal tissues in UALCAN **(F)**. The correlation between *BTG1* expression and overall or post-progression survival rate of the patients with lung cancer was analysis using Kaplan-Meier plotter **(G)**. Note: Ad, adenocarcinoma; Sq, squamous cell carcinoma; Fe, female; Ma, male; S, stage; HR, hazard ratio.

expression was negatively related to the overall survival of the gastric cancer patients (Figure 1E, $p < 0.05$). By the Kaplan-Meier plotter, a higher *BTG1* expression was negatively correlated with overall and progression-free survival rates of all cancer patients, male or perforating cancer patients and patients receiving 5-FU-based adjuvant (Figure 1F; Table 1, $p < 0.05$). As shown in Table 1, negative correlation between overall survival and *BTG1* expression was found in the patients with stage II and IV, T2, N3, intestinal and mixed or Her2-positive cancers ($p < 0.05$). It was similar for progression-free survival in female, male, or poorly-differentiated cancer patients ($p < 0.05$). Negative association between *BTG1* expression and overall prognosis was observed in the cancer patients only receiving surgical operation (Table 1, $p < 0.05$).

The clinicopathological and prognostic significances of *BTG1* mRNA expression in lung cancer

According to xiantao (Figure 2A) and UALCAN databases (Figure 2B), *BTG1* expression was lower in lung cancer than in normal tissue $p < 0.05$). In TCGA data, *BTG1* expression was higher in squamous cell carcinoma than adenocarcinoma, in male than female cancer patients, and in elder than younger cancer patients (Figure 2C, $p < 0.05$). In UALCAN data, *BTG1* expression was lower in N_3 than N_0, N_1 and N_2, S_2 than S_1, and G_3 than G_1 and G_2 cancer (Figure 2D, $p < 0.05$). There was a negative correlation between *BTG1* mRNA and methylation (cg20078640, cg05819371, cg09918929, cg06551025, cg13132650, cg04211745, cg04100724, cg21381360, cg25218905 and cg08832851) by xiantao (Figure 2E, $p < 0.05$). *BTG1* methylation was higher in lung cancer than in normal tissues by UALCAN (Figure 2F, $p < 0.05$). According to Kaplan-Meier plotter, we found that the higher *BTG1* expression was negatively correlated with the overall rate of all, male or G_2 cancer patients (Figure 2G, $p < 0.05$). The Cox's risk proportional analysis showed that younger age, lymph node metastasis, TNM staging and *BTG1* hypoexpression were independent factors for worse prognosis of the patients with lung cancer (Table 2, $p < 0.05$).

TABLE 2 Multivariate analysis of hazard factors of the prognosis of the patients with lung cancer.

Clinicopathological features	Hazard ratio (95% CI)	*p*-value
Age (<60/≧60 years)	2.175 (1.467–3.226)	<0.001
Gender (Female/male)	0.589 (0.081–4.266)	0.600
T stage (T_{1-2}/T_{3-4})	0.713 (0.401–1.269)	0.250
Lymph node metastasis (−/+)	1.761 (1.066–2.911)	0.027
TNM staging (I-II/III-IV)	2.341 (1.343–4.083)	0.003
Histological classification (Ad/Sq)	1.090 (0.650–1.828)	0.744
BTG1 mRNA expression (low/high)	0.503 (0.329–0.770)	0.002

Ad, adenocarcinoma; Sq, squamous cell carcinoma; CI, confidence interval.

The clinicopathological and prognostic significances of *BTG1* mRNA expression in breast cancer

BTG1 was more expressed in breast normal tissue than in cancer according to xiantao (Figure 3A) and UALCAN databases (Figure 3B, $p < 0.05$). TCGA database showed that *BTG1* expression was negatively associated with lymph node metastasis and TNM staging of breast cancer (Figure 3C, $p < 0.05$). There was a negative correlation between *BTG1* mRNA and methylation (cg05819371, cg09918929, cg21381360, and cg08832851) by xiantao (Figure 3D, $p < 0.05$). *BTG1* methylation was higher in breast cancer than normal tissues (Figure 3E, $p < 0.05$), N1 than N0, and Luminal than triple-negative breast cancer (TNBC) patients (Figure 3F, $p < 0.05$) by UALCAN. As summarized in Table 3, *BTG1* mRNA expression was positively correlated with young patients, white race, infiltrative lobular carcinoma, Her2 negativity, better molecular subtyping, and no postmenopause status ($p < 0.05$). TCGA data showed that *BTG1* expression was positively associated with a high overall survival rate of breast cancer patients (Figure 3G $p < 0.05$). According to Kaplan-Meier plotter, a higher *BTG1* expression was positively correlated with overall survival rates of all or luminal-B cancer patients (Figure 3H, $p < 0.05$). Patients with Her2-negative and luminal-B cancer who had high *BTG1* expression had a longer time without distant metastasis than those who had low *BTG1* expression ($p < 0.05$, data not shown). There appeared to be a positive relationship between *BTG1* expression and the progression-free survival rate of cancer patients without chemotherapy or margin invasion ($p < 0.05$, data not shown). The overall survival rate of the patient with ER (estrogen receptor)-negative, grade-3, or luminal-B cancer was higher in the group of high *BTG1* expression than that in its low expression ($p < 0.05$, data not shown).

The clinicopathological and prognostic significances of *BTG1* mRNA expression in ovarian cancer

Then, we used xiantao database to perform bioinformatics analysis and found that *BTG1* expression was lower in ovarian cancers than normal tissues (Figure 4A, $p < 0.05$), in line with GEO data (GSE38666 and GSE26712, Figure 4B). As shown in Table 4, *BTG1* expression was positively associated with elder age, venous invasion, lymphatic invasion, and FIGO (International Federation of Gynecology and Obstetrics) staging of ovarian cancer ($p < 0.05$). According to Kaplan-Meier plotter, a higher *BTG1* expression was negatively correlated with the overall survival rates of Dubulk suboptimal and p53-mutant cancer patients (Figure 4C, $p < 0.05$). Stage I + II, II, II + III, III, and III + IV cancer patients with high *BTG1* expression showed a short progression-free survival time than those with its low expression (Figure 4B, $p < 0.05$). There appeared a negative relationship between *BTG1* expression and the overall survival rate of the cancer patients with paclitaxel treatment (Figure 4B, $p < 0.05$).

The relationship between *BTG1* mRNA expression and the infiltrating immune cells in cancers

According to xiantao, *BTG1* mRNA expression was positively related to the infiltration of $CD8^+$ T cells, TFH, T cells, cytotoxic cells, DC, B cells, aDC, Th1 cells, iDC, TReg, macrophages, NK cells, eosinophils, mast cells and T helper cells, pDC, NK CD56dim cells and Tem, but negatively to Th17 cells in gastric cancer (Figure 5A, $p < 0.05$). It was positively related to the infiltration of Tcm and Tgd, but negatively to T helper cells, neutrophils, aDC, NK CD56bright cells, DC, iDC, NK cells, Th1 cells, pDC, $CD8^+$ T cells, TFH, Tem, eosinophils, macrophages and Th17 cells in lung cancer (Figure 5B, $p < 0.05$). It was positively related to the infiltration of T helper cells, $CD8^+$ T cells, T cells, mast cells, B cells, iDC, TFH, cytotoxic cells, neutrophils, Tcm, Th1 cells, DC, Tem, macrophages, eosinophils, NK CD56bright cells, pDC, aDC, NK cells, TReg, Tgd and Th17 cells in breast cancer (Figure 5C, $p < 0.05$). It was positively related to the infiltration of T helper cells, DC, eosinophils, iDC, macrophages, TFH, neutrophils, T cells, Th2 cells, Tem, Tcm, $CD8^+$ T cells, NK CD56dim cells, Th1 cells, cytotoxic cells and B cells in ovarian cancer (Figure 5D, $p < 0.05$).

The *BTG1*-related genes and pathways in cancers

On the xiantao platform, we discovered the genes that differed between low and high expression groups of *BTG1* mRNA in cancers. KEGG analysis showed that the top signal

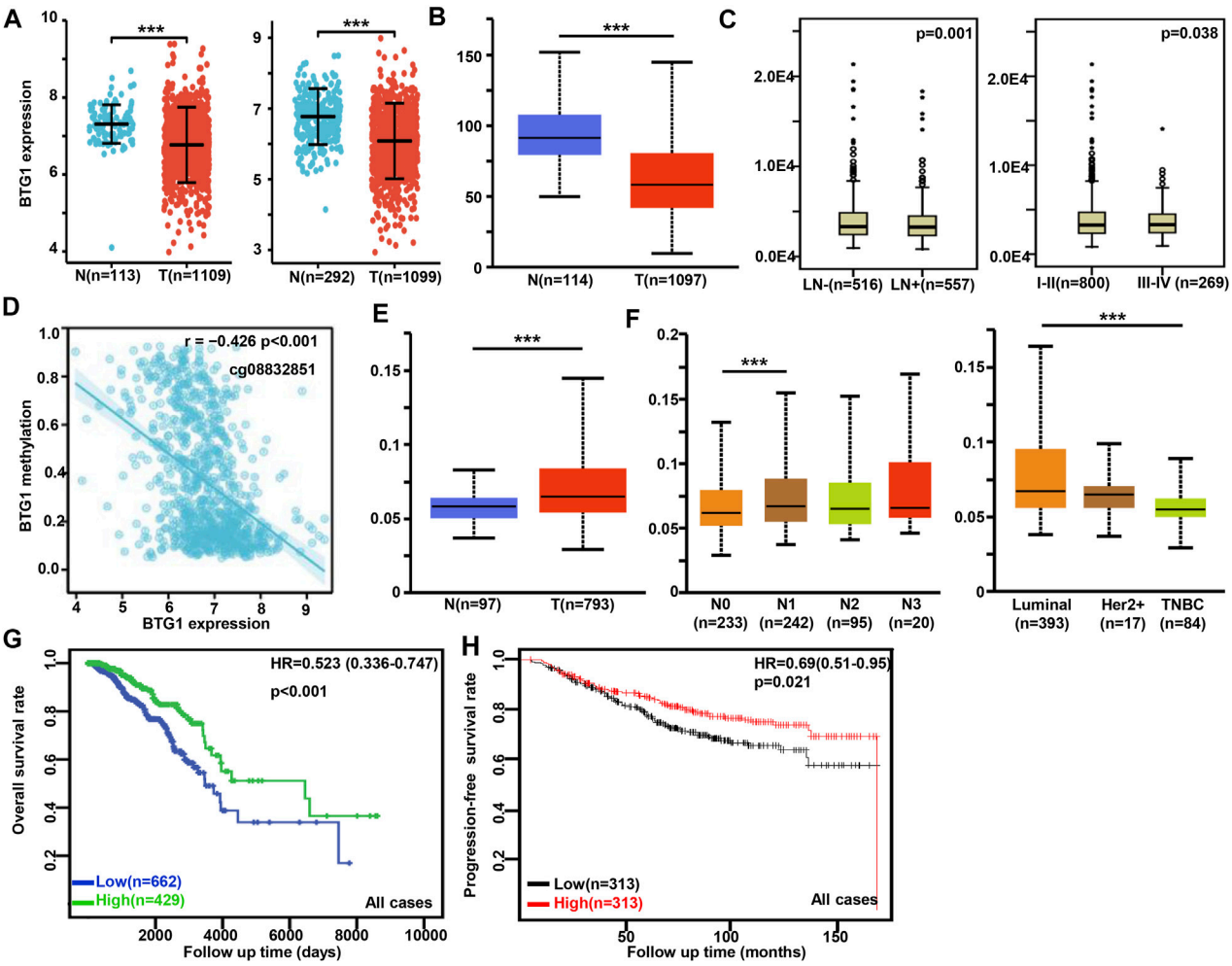


FIGURE 3
The clinicopathological and prognostic significances of *BTG1* mRNA expression in breast cancer. Xiantao **(A)** and UALCAN **(B)** dataset was used for bioinformatics analysis to explore *BTG1* expression in breast cancer. A lower *BTG1* expression was detectable in breast cancer than that in normal mucosa ($p < 0.05$). TCGA database showed that *BTG1* was negatively correlated with lymph node metastasis (LN) and clinicopathological staging of breast cancer **(C)** $p < 0.05$. The negative relationship between *BTG1* expression and methylation was analyzed in breast cancer using xiantao database **(D)**. Its methylation was higher in breast cancer than normal tissues in UALCAN **(E)**, and compared with N staging and molecular subtyping of breast cancer **(F)**. *BTG1* expression was positively correlated with overall survival rate of the cancer patients by TCGA **(G)** $p < 0.05$ and progression-free survival by Kaplan-Meier plotter **(H)** $p < 0.05$. Note: N, normal tissue; T, tumor; TNBC, triple-negative breast cancer; HR, hazard ratio.

pathways included cytokine-cytokine receptor interaction, cell adhesion molecules, chemokine, B cell receptor and NF(nuclear factor)-κB signal pathways in gastric cancer, focal adhesion, lysosome, cell cycle, MAPK (mitogen-activated protein kinase), AMPK (Adenosine 5'-monophosphate-activated protein kinase) and TNF (tumor necrosis factor) signal pathways in lung cancer, cytokine-cytokine receptor interaction, chemokine, T cell and NF-κB signal pathways in breast cancer, and neural diseases, oxidative phosphorylation, non-alcoholic fatty liver disease and ribosome in ovarian cancer (Figure 6A). In addition, the STRING was used to identify the PPI pairs and the cytoscape to find out the top 10 nodes ranked by degree (Figure 6B). The top hub genes mainly contained CD (cluster of differentiation) antigens in gastric cancer, FGF (fibroblast growth factor)-FGFR (FGF receptor) and downstream genes in lung cancer, mitochondrial NADH (nicotinamide adenine dinucleotide): ubiquinone oxidoreductase in breast cancer and ribosomal proteins in ovarian cancer.

According to the xiantao database, the *BTG1*-correlated genes in cancers were analyzed and subjected to the KEGG analysis (Figure 7). The *BTG1*-correlated genes were involved in cell adhesion, chemokine, T cell receptor and PD-L1 signal pathways for gastric cancer, endocytosis, infection-related diseases, and spliceosome, ubiquitin-mediated proteolysis and PD-L1 signal pathway for lung cancer, cytokine-cytokine

TABLE 3 The correlation of *BTG1* mRNA expression with clinicopathological features of breast cancer.

Characteristic	Variables	Low expression	High expression	*p* -value
Age (years), n (%)	≦60	281 (25.9%)	320 (29.5%)	0.022
	>60	260 (24%)	222 (20.5%)	
Race, n (%)	Asian	34 (3.4%)	26 (2.6%)	0.021
	Black or African American	104 (10.5%)	77 (7.7%)	
	White	354 (35.6%)	399 (40.1%)	
Histological type, n (%)	IDC	414 (42.4%)	358 (36.6%)	<0.001
	I LC	66 (6.8%)	139 (14.2%)	
Her2 status, n (%)	Negative	260 (35.8%)	298 (41%)	0.003
	Indeterminate	8 (1.1%)	4 (0.6%)	
	Positive	96 (13.2%)	61 (8.4%)	
PAM50, n (%)	Normal	9 (0.8%)	31 (2.9%)	<0.001
	Luminal A	263 (24.3%)	299 (27.6%)	
	Luminal B	123 (11.4%)	81 (7.5%)	
	Her2+	47 (4.3%)	35 (3.2%)	
	Basal-like	99 (9.1%)	96 (8.9%)	
Menopause status, n (%)	Pre	102 (10.5%)	127 (13.1%)	0.043
	Peri	17 (1.7%)	23 (2.4%)	
	Post	374 (38.5%)	329 (33.8%)	

IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma.

receptor interaction, chemokine, JAK-STAT, FoxO and TNF signal pathway for breast cancer, endocytosis, cAMP, oxytoxin, cGMP and GnRH signal pathways for ovarian cancer.

Discussion

Overexpression of BTG1 inhibited the proliferation, migration and invasion of hepatocytes, thyroid, nasopharynx, esophagus, breast and non-small cell lung cancer cells and induced apoptosis and cell cycle arrest by downregulating the expression of Cyclin D1, Bcl-2 and MMP-9 (Sun et al., 2014a; Sun et al., 2014b; Sun et al., 2014c; Sun et al., 2014d; Lu et al., 2014; Sheng et al., 2014). In ovarian cancer, BTG1 expression caused lower growth rate, high cisplatin sensitivity, G_1 arrest, apoptosis, and decreased migration and invasion by down-regulating the expression of PI3K (phosphatidylinositol 3-kinase), PKB (protein kinase B), Bcl-xL, survivin, VEGF (vascular epithelial growth factor), and MMP (matrix metalloproteinase)-2 (Zhao et al., 2013). The chemosensitivity of BTG1 transfectants to paclitaxel, cisplatin, MG132 (proteasome inhibitor) or SAHA (histone deacetylase inhibitor) was positively correlated with its apoptotic induction of colorectal cancer cells (Zhao et al., 2017). BTG1 overexpression suppressed tumor growth and lung metastasis of gastric and colorectal cancer cells by inhibiting proliferation, and enhancing autophagy and apoptosis in xenograft models (Zheng et al., 2015; Zhao et al., 2017). Taken together, BTG1 should be used as a molecular target for cancer gene therapy.

Jung et al. (2018) found that *BTG1* expression was lower in colorectal cancer than in control, and in metastatic cancer than in primary cancer, due to the hypermethylation of the *BTG1* promoter. BTG1 expression was decreased in hepatocellular, thyroid, nasopharyngeal, esophageal, breast, colorectal, and non-small cell lung cancers, and negatively associated with aggressive behaviors (Sun et al., 2014a; Sun et al., 2014b; Sun et al., 2014c; Sun et al., 2014d; Lu et al., 2014; Sheng et al., 2014; Zhu et al., 2014). Decreased BTG1 expression in gastric cancer was positively correlated with depth of invasion, lymphatic and venous invasion, lymph node metastasis, TNM staging and worse prognosis (Zhu et al., 2015), but the lower BTG1 expression in ovarian cancer was positively correlated with FIGO staging (Sheng et al., 2014). Reportedly, the downregulated *BTG1* expression is positively correlated with its promoter methylation in colorectal, gastric and ovarian cancers (Kanda et al., 2015; Zheng et al., 2015; Kim et al., 2017). Here, we found that *BTG1* expression was lower in gastric, lung, breast and ovarian cancer than normal tissue due to its promoter methylation, which was opposite of *BTG1* expression. Moreover, *BTG1* expression was positively correlated with dedifferentiation and histological grading of gastric cancer, and venous invasion, lymphatic invasion, and TNM staging of ovarian cancer, but negatively associated with lymph node metastasis and TNM staging of breast cancer. In breast cancer, the

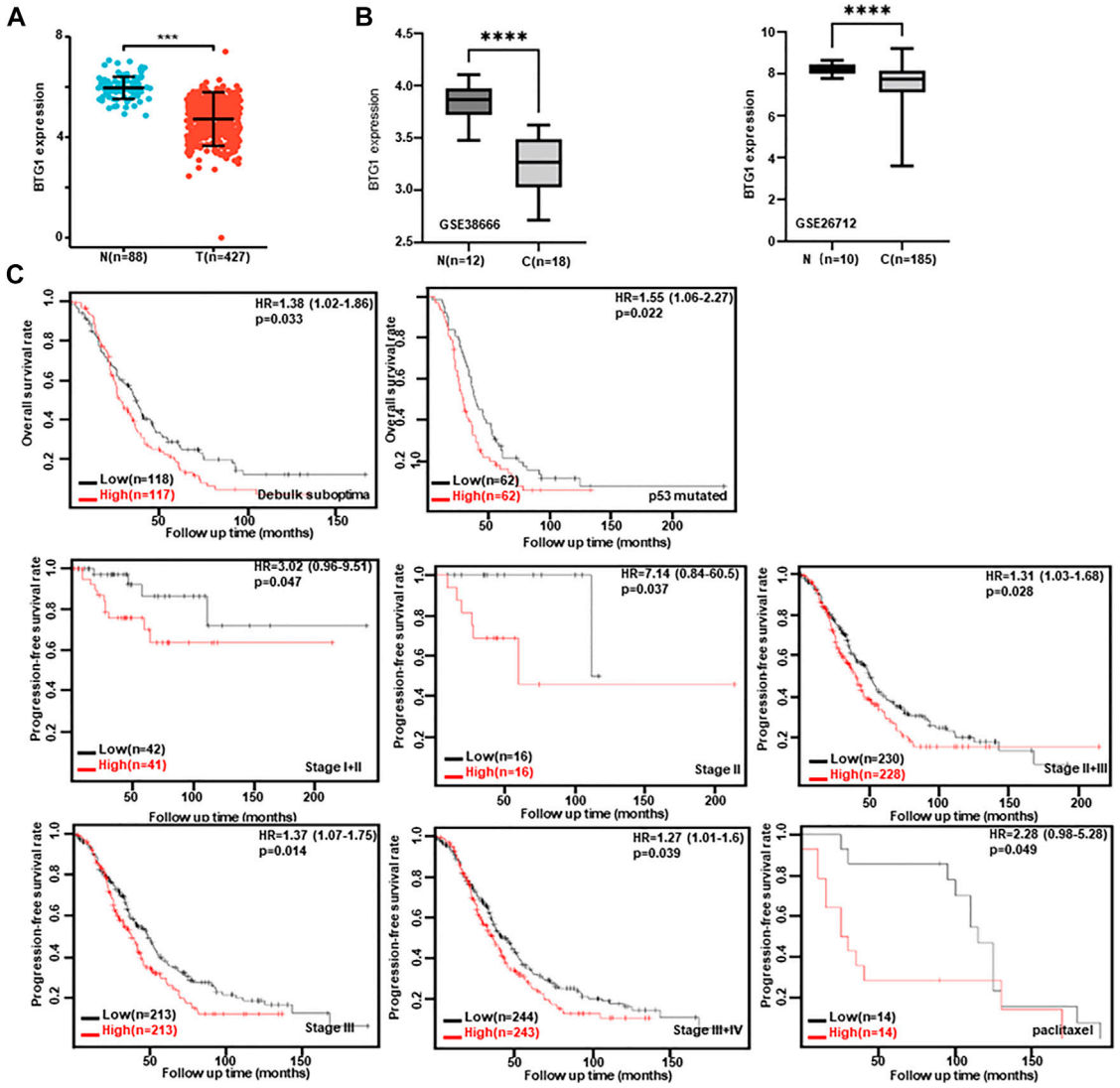


FIGURE 4
The clinicopathological and prognostic significances of *BTG1* mRNA expression in ovarian cancer. Both xiantao **(A)** and GEO **(B)** databases were employed for bioinformatics analysis to observe *BTG1* expression in ovarian cancer. A lower *BTG1* expression was detectable in ovarian carcinoma than normal tissue ($p < 0.05$). The correlation between *BTG1* expression and overall, or post-progression survival rate of the patients with ovarian cancer, even stratified by different clinicopathological parameters **(C)** $p < 0.05$. HR, hazard ratio.

relation of *BTG1* expression to TNM staging was the opposite to that between *BTG1* methylation and TNM staging. *BTG*1 expression was higher in pulmonary squamous cell carcinoma than in adenocarcinoma, which provided another evidence that squamous cell carcinoma but not adenocarcinoma showed higher *BTG1* expression than normal tissue. These findings suggested that aberrant *BTG1* expression was positively correlated with carcinogenesis, histogenesis and subsequent progression because of its aberrant promoter methylation. The contradictory data in our and other studies could be attributed to sample selection, different methodologies, and tissue specificity.

Kanda et al. (2015) reported that downregulation of *BTG1* mRNA in gastric cancer was positively associated with shorter disease-specific and recurrence-free survival of the patients with gastric cancer as an independent prognostic factor. BTG1 expression was adversely linked to poor prognosis of patients with hepatocellular, thyroid, nasopharyngeal, esophageal, breast, or non-small cell lung cancer (Sun et al., 2014b; Sun et al., 2014c; Sun et al., 2014d; Lu et al., 2014; Sheng et al., 2014). According to the Kaplan-Meier plotter, we found that *BTG1* expression was negatively correlated with favorable prognosis of gastric, lung, or ovarian cancer patients, but not for breast cancer patients. The correlation between *BTG1* expression

TABLE 4 The correlation of *BTG1* mRNA expression with clinicopathological features of ovarian cancer.

Characteristic	Variables	Low expression	High expression	*p*-value
Age (years), n (%)	≦60	112 (29.6%)	96 (25.3%)	0.108
	>60	77 (20.3%)	94 (24.8%)	
Venous invasion, n (%)	No	21 (20%)	20 (19%)	0.480
	Yes	27 (25.7%)	37 (35.2%)	
Lymphatic invasion, n (%)	No	30 (20.1%)	18 (12.1%)	0.036
	Yes	43 (28.9%)	58 (38.9%)	
FIGO stage, n (%)	Stage 1	0 (0%)	1 (0.3%)	0.022
	Stage 2	12 (3.2%)	11 (2.9%)	
	Stage 3	138 (36.7%)	157 (41.8%)	
	Stage 4	38 (10.1%)	19 (5.1%)	

FIGO, federation of gynecology and obstetrics.

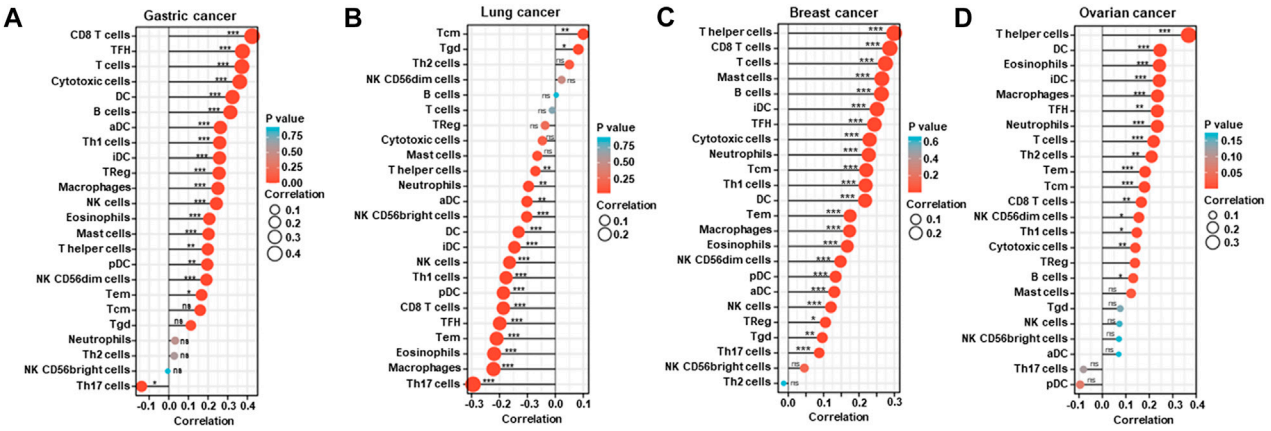


FIGURE 5
The relationship between *BTG1* mRNA expression and immune infiltration in cancers. The enrichment of immune cells was explored between low and high expression of *BTG1* in gastric **(A)**, lung **(B)**, breast **(C)**, and ovarian **(D)** cancers using xiantao. ns, not significant; *p < 0.05; **p < 0.01; ***p < 0.001.

and prognosis of the cancer patients did not parallel with the alteration in *BTG1* expression in cancer tissues. Additionally, TCGA data showed the same results about the prognostic significance of *BTG1* expression to Kaplan-Meier plotter's in gastric and breast cancers although Kaplan-Meier plotter is based on cDNA array and the TCGA experiment on RNA sequencing. The tissue specificity might account for the phenomenon of the relationship between *BTG1* expression and prognosis.

Reportedly, human papillomavirus E7 peptide (p11-20) induced the pro-apoptotic *BTG1* expression in mature dendritic cells from human peripheral blood monocytes (Bard et al., 2004). BTG1 has been discovered to regulate quiescence in naïve and memory T cells (Kim et al., 2022). Loss of BTG1 expression causes glucocorticoid resistance both by decimating glucocorticoid receptor (GR) expression and the recruitment of BTG1/PRMT1 complex to GR promoter (van Galen et al., 2010). In the present study, we found that *BTG1* mRNA expression was closely linked to the infiltration of immune cells, indicating the possible role of *BTG1* in immunotherapy in future. The relationship between *BTG1* expression and infiltrating immune cells in gastric, breast and ovarian cancer was opposite to that in lung cancer, which might be attributable to its inclusion of lung adenocarcinoma and squamous cell carcinoma.

Because m6A-hypomethylation regulated FGFR4 phosphorylates GSK-3β and activates β-catenin/TCF4 signaling to drive anti-HER2 resistance, FGFR4 inhibition enhances susceptibility to anti-HER2 therapy in resistant breast cancer by reducing glutathione synthesis and Fe^{2+} efflux efficiency *via* the β-catenin/TCF4-SLC7A11/FPN1 axis and labile iron pool accumulation (Zou et al., 2022). Reportedly, TNBCs were classified into immune Phenotypes A and B

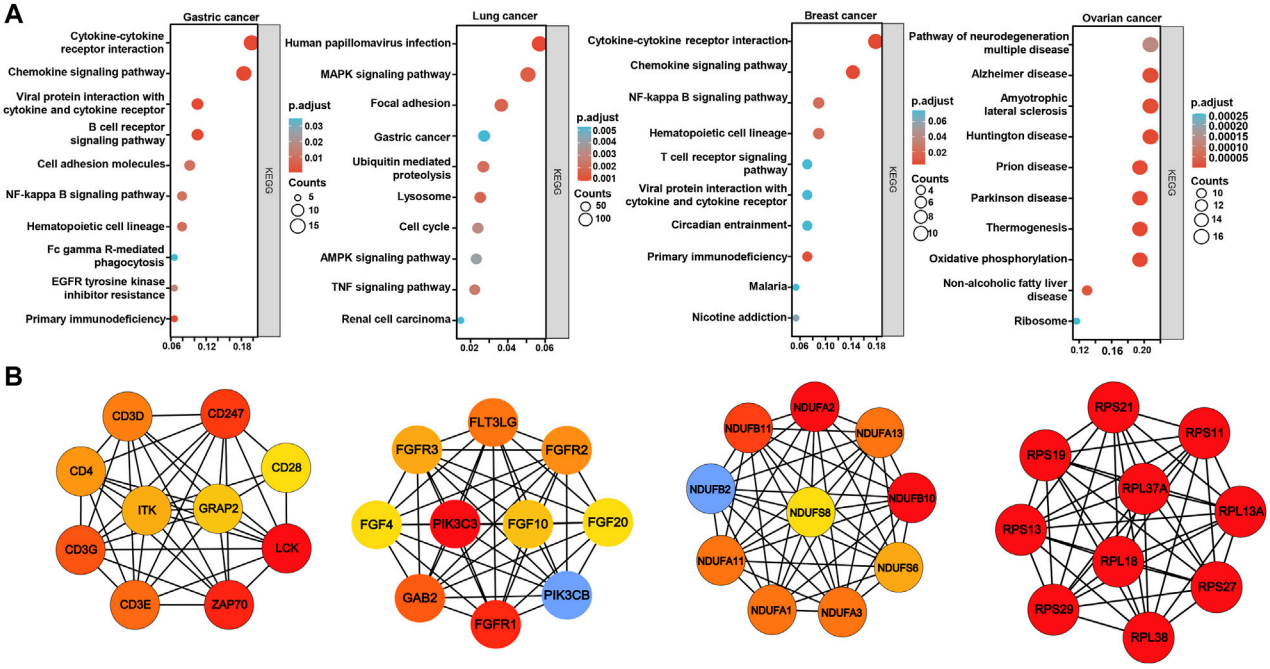


FIGURE 6
The differential genes and related signal pathways about *BTG1* in cancers The differential genes of *BTG1* were subjected to the signal pathway analysis using KEGG **(A)**. STRING was used to identify the protein-protein interaction network of differential genes about *BTG1* in cancers, and cytoscape was employed to find out the top 10 hub nodes ranked by degree **(B)**.

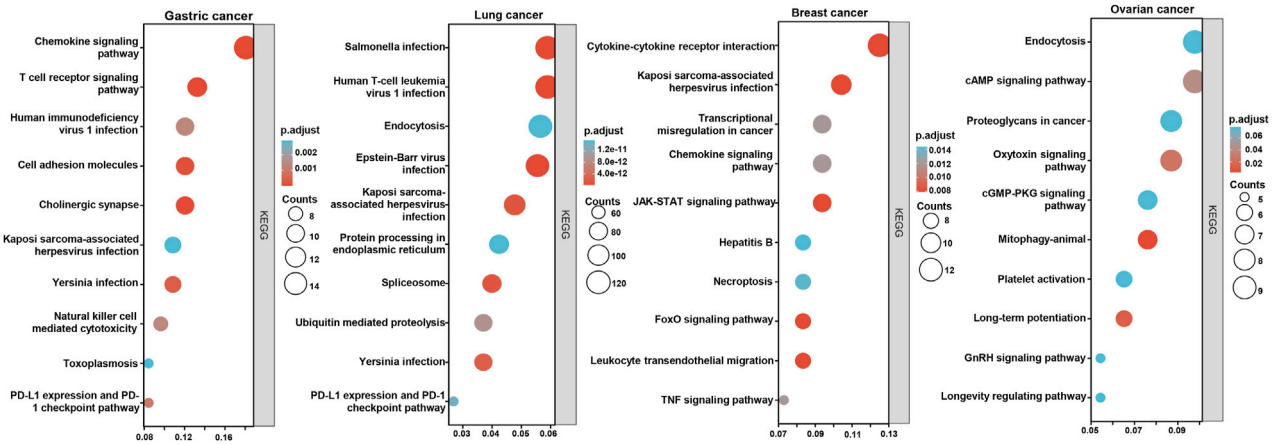


FIGURE 7
The *BTG1*-related signal pathways in cancers. The top *BTG1*-related genes were screened and classified into the signal pathways of KEGG using xiantao database.

according to the infiltration of stromal γδ T cells, $CD4^+$ T cells, monocytes, M1 and M2 macrophages. In Phenotype A, there were rich immune-related pathways, and a stronger level of PD-L1, PD-1 and CTLA-4 (Zheng et al., 2020). Here, we found that *BTG1* mRNA expression was remarkably linked to the infiltration of immune cells, Her2 expression and molecular subtyping of breast cancer, demonstrating that *BTG1* would be employed to guide the target therapy for breast cancer patients.

KEGG analysis demonstrated that the *BTG1*-related pathways included cytokine-cytokine receptor interaction, cell adhesion

molecules, chemokine, immune cell receptor and NF-κB signal pathways in gastric and breast cancers. The top hub genes mainly contained CD antigens in gastric cancer, FGF-FGFR and downstream genes in lung cancer, mitochondrial NADH: ubiquinone oxidoreductase in breast cancer, and ribosomal proteins in ovarian cancer respectively. These bioinformatics data provided novel clues to the molecular mechanisms of BTG1 in malignancies.

In conclusion, aberrant *BTG1* mRNA expression was closely linked to carcinogenesis, cancer aggressiveness, and worse prognosis of cancer patients in a tissue-specific manner. The study's limitation is that the results from the Oncomine, TCGA, xiantao, UALCAN, and Kaplan-Meier plotter datasets were not validated using real-time RT-PCR, even after laser capture dissection because the specificity of cDNA chip or RNA sequencing is not enough high.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found in the article/Supplementary Material.

Author contributions

Conception and design: H-CZ Collection and assembly of data: HX and C-YZ Data analysis and interpretation: K-HS and RZ. Manuscript writing: H-CZ. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (21377772D) and National Natural Scientific Foundation of China (81672700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GZ declared a shared affiliation with the author RZ to the handling editor at time of review.

Publisher's note

References

Bard, M. P., Hegmans, J. P., Hemmes, A., Luider, T. M., Willemsen, R., Severijnen, L. A., et al. (2004). Proteomic analysis of exosomes isolated from human malignant pleural effusions. *Am. J. Respir. Cell Mol. Biol.* 31 (1), 114–121. doi:10.1165/rcmb.2003-0238OC

Bogdan, J. A., Adams-Burton, C., Pedicord, D. L., Sukovich, D. A., Benfield, P. A., Corjay, M. H., et al. (1998). Human carbon catabolite repressor protein (CCR4)-associative factor 1: Cloning, expression and characterization of its interaction with the B-cell translocation protein BTG1. *Biochem. J.* 33, 471–481. doi:10.1042/bj3360471

Busson, M., Carazo, A., Seyer, P., Grandemange, S., Casas, F., Pessemesse, L., et al. (2005). Coactivation of nuclear receptors and myogenic factors induces the major BTG1 influence on muscle differentiation. *Oncogene* 24 (10), 1698–1710. doi:10.1038/sj.onc.1208373

Cheng, Q., Li, Q., Xu, L., and Jiang, H. (2021). Exosomal microRNA-301a-3p promotes the proliferation and invasion of nasopharyngeal carcinoma cells by targeting BTG1 mRNA. *Mol. Med. Rep.* 23 (5), 328. doi:10.3892/mmr.2021.11967

Corjay, M. H., Kearney, M. A., Munzer, D. A., Diamond, S. M., and Stoltenborg, J. K. (1998). Antiproliferative gene BTG1 is highly expressed in apoptotic cells in macrophage-rich areas of advanced lesions in Watanabe heritable hyperlipidemic rabbit and human. *Lab. Invest.* 78 (7), 847–858.

He, C., Yu, T., Shi, Y., Ma, C., Yang, W., Fang, L., et al. (2017). MicroRNA 301A promotes intestinal inflammation and colitis-associated cancer development by inhibiting BTG1. *Gastroenterology* 152 (6), 1434–1448. doi:10.1053/j.gastro.2017.01.049

Jung, Y. Y., Sung, J. Y., Kim, J. Y., and Kim, H. S. (2018). Down-regulation of B-Cell translocation gene 1 by promoter methylation in colorectal carcinoma. *Anticancer Res.* 38 (2), 691–697. doi:10.21873/anticanres.12274

Kanda, M., Oya, H., Nomoto, S., Takami, H., Shimizu, D., Hashimoto, R., et al. (2015). Diversity of clinical implication of B-cell translocation gene 1 expression by histopathologic and anatomic subtypes of gastric cancer. *Dig. Dis. Sci.* 60 (5), 1256–1264. doi:10.1007/s10620-014-3477-8

Kim, J. Y., Do, S. I., Bae, G. E., and Kim, H. S. (2017). B-cell translocation gene 1 is downregulated by promoter methylation in ovarian carcinoma. *J. Cancer* 8 (14), 2669–2675. doi:10.7150/jca.21037

Kim, M., Jo, H., Kwon, Y., Kim, Y., Jung, H. S., and Jeoung, D. (2020). Homoharringtonine inhibits allergic inflammations by regulating NF-κB-miR-183-5p-BTG1 axis. *Front. Pharmacol.* 11, 1032. doi:10.3389/fphar.2020.01032

Kim, S. H., Jung, I. R., and Hwang, S. S. (2022). Emerging role of anti-proliferative protein BTG1 and BTG2. *BMB Rep.* 55 (8), 380–388. doi:10.5483/bmbrep.2022.55.8.092

Li, E., Han, K., and Zhou, X. (2019). microRNA-27a-3p Down-regulation inhibits malignant biological behaviors of ovarian cancer by targeting BTG1. *Open Med.* 14, 577–585. doi:10.1515/med-2019-0065

Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S., and Herschman, H. R. (1996). The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J. Biol. Chem.* 271 (25), 15034–15044. doi:10.1074/jbc.271.25.15034

Liu, C., Tao, T., Xu, B., Lu, K., Zhang, L., Jiang, L., et al. (2015). BTG1 potentiates apoptosis and suppresses proliferation in renal cell carcinoma by interacting with PRMT1. *Oncol. Lett.* 10 (2), 619–624. doi:10.3892/ol.2015.3293

Lu, Y. F., Sun, G. G., Liu, Q., Yang, C. R., and Cheng, Y. J. (2014). BTG1 expression in thyroid carcinoma: Diagnostic indicator and prognostic marker. *Int. J. Oncol.* 45 (4), 1574–1582. doi:10.3892/ijo.2014.2543

Nahta, R., Yuan, L. X., Fiterman, D. J., Zhang, L., Symmans, W. F., Ueno, N. T., et al. (2006). B cell translocation gene 1 contributes to antisense Bcl-2-mediated apoptosis in breast cancer cells. *Mol. Cancer Ther.* 5 (6), 1593–1601. doi:10.1158/1535-7163.MCT-06-0133

Nan, Y. H., Wang, J., Wang, Y., Sun, P. H., Han, Y. P., Fan, L., et al. (2016). MiR-4295 promotes cell growth in bladder cancer by targeting BTG1. *Am. J. Transl. Res.* 8 (11), 4892–4901.

Ni, Z., Shen, Y., Wang, W., Cheng, X., and Fu, Y. (2021). miR-141-5p Affects the cell proliferation and apoptosis by targeting BTG1 in cervical Cancer. New York: Cancer Biother Radiopharm. in press.

Prévôt, D., Voeltzel, T., Birot, A. M., Morel, A. P., Rostan, M. C., Magaud, J. P., et al. (2000). The leukemia-associated protein Btg1 and the p53-regulated protein Btg2 interact with the homeoprotein Hoxb9 and enhance its transcriptional activation. *J. Biol. Chem.* 275 (1), 147–153. doi:10.1074/jbc.275.1.147

Sheng, S. H., Zhao, C. M., and Sun, G. G. (2014). BTG1 expression correlates with the pathogenesis and progression of breast carcinomas. *Tumour Biol.* 35 (4), 3317–3326. doi:10.1007/s13277-013-1437-0

Su, C., Huang, D. P., Liu, J. W., Liu, W. Y., and Cao, Y. O. (2019). miR-27a-3p regulates proliferation and apoptosis of colon cancer cells by potentially targeting BTG1. *Oncol. Lett.* 18 (3), 2825–2834. doi:10.3892/ol.2019.10629

Sun, G. G., Lu, Y. F., Cheng, Y. J., and Hu, W. N. (2014). The expression of BTG1 is downregulated in NSCLC and possibly associated with tumor metastasis. *Tumour Biol.* 35 (4), 2949–2957. doi:10.1007/s13277-013-1379-6

Sun, G. G., Lu, Y. F., Cheng, Y. J., Yang, C. R., Liu, Q., Jing, S. W., et al. (2014). Expression of BTG1 in hepatocellular carcinoma and its correlation with cell cycles, cell apoptosis, and cell metastasis. *Tumour Biol.* 35 (12), 11771–11779. doi:10.1007/s13277-014-2298-x

Sun, G. G., Wang, Y. D., Cheng, Y. J., and Hu, W. N. (2014). BTG1 underexpression is an independent prognostic marker in esophageal squamous cell carcinoma. *Tumour Biol.* 35 (10), 9707–9716. doi:10.1007/s13277-014-2245-x

Sun, G. G., Wang, Y. D., Cheng, Y. J., and Hu, W. N. (2014). The expression of BTG1 is downregulated in nasopharyngeal carcinoma and possibly associated with tumour metastasis. *Mol. Biol. Rep.* 41 (9), 5979–5988. doi:10.1007/s11033-014-3475-0

Szász, A. M., Lánczky, A., Nagy, Á., Förster, S., Hark, K., Green, J. E., et al. (2016). Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1, 065 patients. *Oncotarget* 7 (31), 49322–49333. doi:10.18632/oncotarget.10337

van Galen, J. C., Kuiper, R. P., van Emst, L., Levers, M., Tijchon, E., Scheijen, B., et al. (2010). BTG1 regulates glucocorticoid receptor autoinduction in acute lymphoblastic leukemia. *Blood* 115 (23), 4810–4819. doi:10.1182/blood-2009-05-223081

Xue, K., Wu, J. C., Li, X. Y., Li, R., Zhang, Q. L., Chang, J. J., et al. (2021). Chidamide triggers BTG1- mediated autophagy and reverses the chemotherapy resistance in the relapsed/refractory B-cell lymphoma. *Cell Death Dis.* 12 (10), 900. doi:10.1038/s41419-021-04187-5

Zhang, H., Tang, J., Li, C., Kong, J., Wang, J., Wu, Y., et al. (2015). MiR-22 regulates 5-FU sensitivity by inhibiting autophagy and promoting apoptosis in colorectal cancer cells. *Cancer Lett.* 356, 781–790. doi:10.1016/j.canlet.2014.10.029

Zhang, S. Q., Yang, Z., Cai, X. L., Zhao, M., Sun, M. M., Li, J., et al. (2017). miR-511 promotes the proliferation of human hepatoma cells by targeting the 3'UTR of B cell translocation gene 1 (BTG1) mRNA. *Acta Pharmacol. Sin.* 38 (8), 1161–1170. doi:10.1038/aps.2017.62

Zhao, S., Chen, S. R., Yang, X. F., Shen, D. F., Takano, Y., Su, R. J., et al. (2017). BTG1 might be employed as a biomarker for carcinogenesis and a target for gene therapy in colorectal cancers. *Oncotarget* 8 (5), 7502–7520. doi:10.18632/oncotarget.10649

Zhao, S., Xue, H., Hao, C. L., Jiang, H. M., and Zheng, H. C. (2020). BTG1 overexpression might promote invasion and metastasis of colorectal cancer *via* decreasing adhesion and inducing epithelial-mesenchymal transition. *Front. Oncol.* 10, 598192. doi:10.3389/fonc.2020.598192

Zhao, X., Chen, G. Q., and Cao, G. M. (2019). Abnormal expression and mechanism of miR-330-3p/BTG1 axis in hepatocellular carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 23 (16), 6888–6898. doi:10.26355/eurrev_201908_18728

Zhao, Y., Gou, W. F., Chen, S., Takano, Y., Xiu, Y. L., and Zheng, H. C. (2013). BTG1 expression correlates with the pathogenesis and progression of ovarian carcinomas. *Int. J. Mol. Sci.* 14 (10), 19670–19680. doi:10.3390/ijms141019670

Zheng, H. C., Li, J., Shen, D. F., Yang, X. F., Zhao, S., Wu, Y. Z., et al. (2015). BTG1 expression correlates with pathogenesis, aggressive behaviors and prognosis of gastric cancer: A potential target for gene therapy. *Oncotarget* 6 (23), 19685–19705. doi:10.18632/oncotarget.4081

Zheng, S., Zou, Y., Xie, X., Liang, J. Y., Yang, A., Yu, K., et al. (2020). Development and validation of a stromal immune phenotype classifier for predicting immune activity and prognosis in triple-negative breast cancer. *Int. J. Cancer* 147 (2), 542–553. doi:10.1002/ijc.33009

Zhu, R., Li, W., Xu, Y., Wan, J., and Zhang, Z. (2015). Upregulation of BTG1 enhances the radiation sensitivity of human breast cancer *in vitro* and *in vivo*. *Oncol. Rep.* 34 (6), 3017–3024. doi:10.3892/or.2015.4311

Zhu, Y., Qiu, P., and Ji, Y. (2014). TCGA-Assembler: Open-source software for retrieving and processing TCGA data. *Nat. Methods* 11 (6), 599–600. doi:10.1038/nmeth.2956

Zou, Y., Zheng, S., Xie, X., Ye, F., Hu, X., Tian, Z., et al. (2022). N6-methyladenosine regulated FGFR4 attenuates ferroptotic cell death in recalcitrant HER2-positive breast cancer. *Nat. Commun.* 13 (1), 2672. doi:10.1038/s41467-022-30217-7

frontiers | Frontiers in Cell and Developmental Biology

TYPE Review
PUBLISHED 08 November 2022
DOI 10.3389/fcell.2022.1012179

OPEN ACCESS

EDITED BY
Wei Liu,
Arizona State University, United States

REVIEWED BY
Qing Feng,
Nanjing Medical University, China
Tao Wang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Yue Wang,
Nankai University, China

*CORRESPONDENCE
Hua-chuan Zheng,
zheng_huachuan@hotmail.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 05 August 2022
ACCEPTED 05 October 2022
PUBLISHED 08 November 2022

CITATION
Zheng H-c, Xue H and Jiang H-m (2022), The roles of ING5 in cancer: A tumor suppressor.
Front. Cell Dev. Biol. 10:1012179.
doi: 10.3389/fcell.2022.1012179

The roles of ING5 in cancer: A tumor suppressor

Hua-chuan Zheng[1]*, Hang Xue[1] and Hua-mao Jiang[2]

[1]Department of Oncology and Central Laboratory, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Department of Urology, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

As a Class II tumor suppressor, ING5 contains nuclear localization signal, plant homeodomain, novel conserved region, and leucine zipper-like domains. ING5 proteins form homodimer into a coil-coil structure, and heterodimers with ING4, histone H3K4me3, histone acetyltransferase (HAT) complex, Tip60, Cyclin A1/CDK2, INCA1 and EBNA3C for the transcription of target genes. The acetylated proteins up-regulated by ING5 are preferentially located in nucleus and act as transcription cofactors, chromatin and DNA binding functions, while those down-regulated by ING5 mostly in cytoplasm and contribute to metabolism. ING5 promotes the autoacetylation of HAT p300, p53, histone H3 and H4 for the transcription of downstream genes (Bax, GADD45, p21, p27 and so forth). Transcriptionally, YY1 and SRF up-regulate ING5 mRNA expression by the interaction of YY1-SRF-p53-ING5 complex with ING5 promoter. Translationally, ING5 is targeted by miR-196, miR-196a, miR-196b-5p, miR-193a-3p, miR-27-3p, miR-200b/200a/429, miR-1307, miR-193, miR-222, miR-331-3p, miR-181b, miR-543 and miR-196-b. ING5 suppresses proliferation, migration, invasion and tumor growth of various cancer cells *via* the suppression of EGFR/PI3K/Akt, IL-6/STAT3, Akt/NF-κB/NF-κB/MMP-9 or IL-6/CXCL12 pathway. ING5-mediated chemoresistance is closely linked to anti-apoptosis, overexpression of chemoresistant genes, the activation of PI3K/Akt/NF-κB and Wnt/β-catenin signal pathways. Histologically, ING5 abrogation in gastric stem-like and pdx1-positive cells causes gastric dysplasia and cancer, and conditional ING5 knockout in pdx1-positive and gastric chief cells increases MNU-induced gastric carcinogenesis. Intestinal ING5 deletion increases AOM/DSS-induced colorectal carcinogenesis and decreases high-fat-diet weight. The overexpression and nucleocytoplasmic translocation of ING5 are seen during carcinogenesis, and ING5 expression was inversely associated with aggressive behaviors and poor prognosis in a variety of cancers. These findings indicated that ING5 might be used for a molecular marker for carcinogenesis and following progression, and as a target for gene therapy if its chemoresistant function might be ameliorated.

Abbreviations: CAF, cancer-associated fibroblast; DSB, double-strand break; EBV, Epstain-Barr virus; EMT, epithelial- mesenchymal transition; HAT, histone acetyltransferase; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HNSCC, head and neck squamous cell carcinoma; ING, inhibitor of growth; KO, knockout; LOH, loss of heterozygosity; LZL, leucine zipper-like; NER, nucleotide excision repair; NCR, novel conserved region; NLS, nuclear localization signal; PASMC, pulmonary arterial smooth muscle cell; PHD, plant homeodomain; WT, wild-type.

KEYWORDS
signal pathway, biological function, ING5, cancer, tumor suppressor

Introduction

Carcinogenesis is a complex biological process, and characterized by both frequent genetic and epigenetic changes, including the inactivation of tumor suppressor genes (TSG) and the activation of oncogenes. The chromosomal deletion, mutation and hypermethylation of TSGs result in their loss (Class I) and inactivation (Class II), finally to malignantly transform the normal cells. Inhibitor of growth (ING) family members belong to Class II TSG and play a critical role in initiation, promotion and development of cancers (Zheng et al., 2011; Dantas et al., 2019). In the review, we demonstrate that

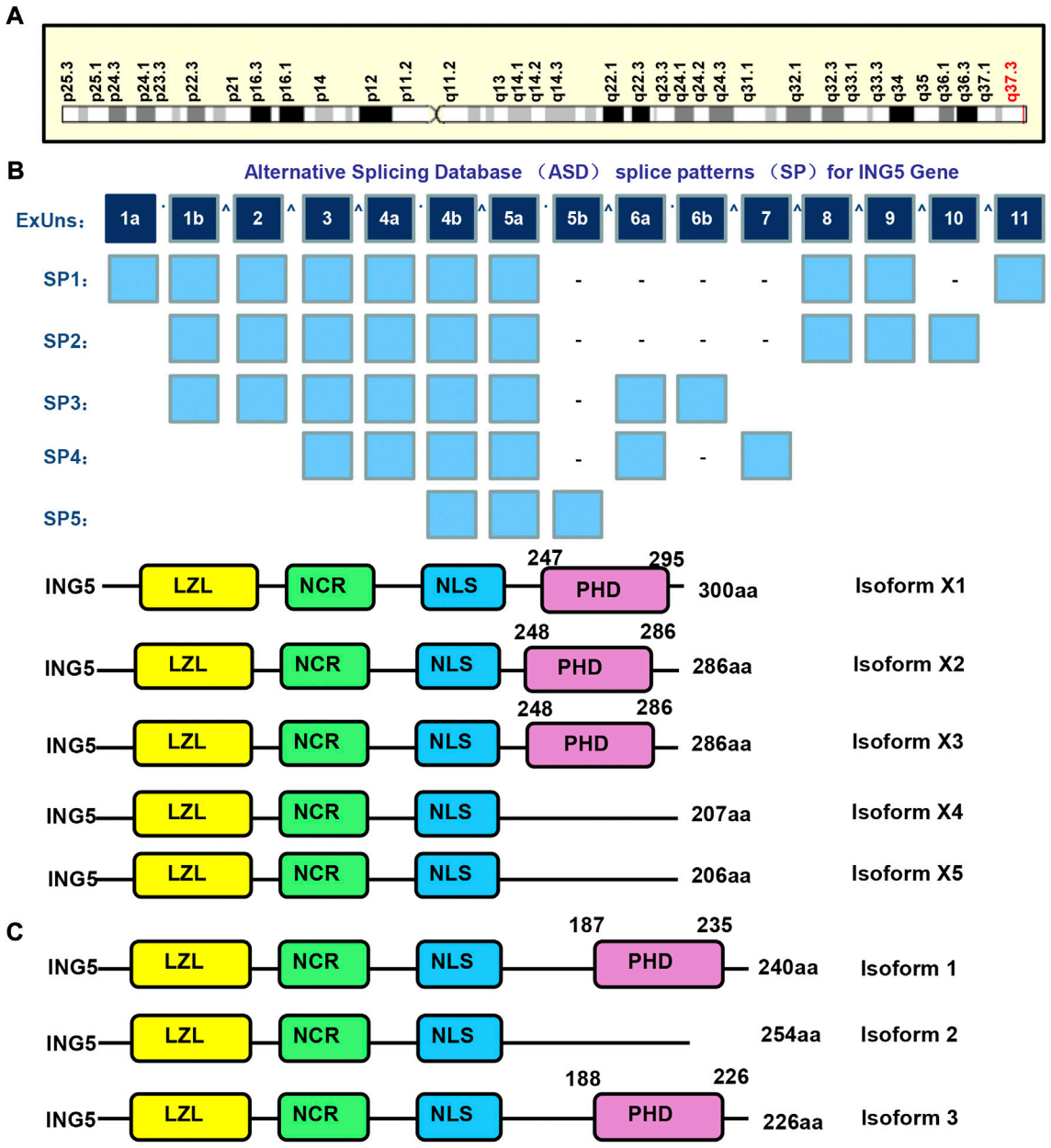


FIGURE 1
The structures of ING5 gene and protein. ING5 is localized in human chromosome 2q37.3 **(A)**, and theoretically spliced into 5 forms and encodes 5 proteins **(B)**. The 3 isoforms of ING5 are produced according to Genbank **(C)**. LZL, leucine zipper-Like motif; NCR: novel conserved region; NLS, nuclear localization signal; PHD, plant homeodomain.

ING5 interacts with ING4, histone H3K4me3, HAT (histone acetyltransferase) complex, Tip60, Cyclin A1/CDK2, INCA1 and EBNA3C, and promotes the autoacetylation of HAT p300, p53, histone H3 and H4 for the transcription of downstream genes. Transcriptionally, YY1 and SRF up-regulate ING5 mRNA expression by the interaction of YY1-SRF-p53-ING5 complex with ING5 promoter. Translationally, ING5 is targeted by miR-196, miR-193a-3p, miR-27-3p and so forth. ING5 suppresses aggressive phenotypes of various cancer cells *via* EGFR/PI3K/Akt, IL-6/STAT3, Akt/NF-κB/NF-κB/MMP-9 or IL-6/CXCL12 pathway. Histologically, conditional ING5 abrogation causes gastric dysplasia and cancer, and increases chemically-induced gastrointestinal carcinogenesis. In human cancer samples, the up-regulated expression of ING5 was in gastric, breast, and colorectal cancers, but down-regulated in head and neck squamous cell carcinoma (HNSCC), lung cancer, osteosarcoma, prostate cancer, ovarian cancer, hepatocellular carcinoma (HCC), esophageal cancer, and thyroid cancer. These results suggest that ING5 might be used for a molecular marker for carcinogenesis and following progression, and as a target for gene therapy.

The structure and functions of ING5

The ING (inhibitor of growth) family is composed of ING1–5, and involved in apoptosis, senescence and DNA damage. They function as epigenetic readers of H3K4Me3 histone, and contribute to HAT or deacetylase complex formation. Their gatekeeper functions are evidenced by their arrest of the cell cycle, stabilization of p53, or maintenance of HAT activity. In contrast, their caretaker functions are evidenced by the nucleotide excision repair (NER) and double-strand break (DSB) repair of ING1-, and ING2- and ING3-mediated accumulation of ATM, BRCA1, and 53BP1 (Archambeau et al., 2019; Dantas et al., 2019). Among them, ING5 maps to human chromosome 2q37.3 (Figure 1A), contains 11 exons, theoretically spliced into 5 forms and encodes 5 proteins (Figure 1B). The 3 isoforms of ING5 are produced according to Genbank (Figure 1C). ING5 protein contains plant homeodomain (PHD), nuclear localization signal (NLS), novel conserved region (NCR), and leucine zipper-like (LZL) domains. Their NLS favors nucleic translocation, LZL mediates interaction with leucine zipper proteins, PHD interacts with such chromatin-interacting proteins as methylated histone H3, and NCR binds to HAT complex to remodel chromatin and modulate transcription (Zhang et al., 2017a; Archambeau et al., 2019; Dantas et al., 2019).

ING5 proteins form homodimer *via* their amino-terminal domain, and fold independently into a coil-coil structure. The PHD domains of the homodimer are chemically equivalent and independent of the rest of ING5. ING5 protein forms heterodimer with ING4 protein as well (Ormaza et al., 2019). ING5 binds to histone H3K4me3 and forms 2 HAT complexes, H3-MOZ-BRPF-MORF-ING5 and H4-JADE-HBO1-ING5. Both MCM and HAT complexes are involved in DNA replication because ING5 silencing abolishes DNA synthesis, and HBO1 silencing promotes the progression of S phase (Zhang et al., 2017a; Wu et al., 2018; Archambeau et al., 2019; Dantas et al., 2019; Ormaza et al., 2019). Zhang et al. (2011) identified ING5 as a partner of inhibitor of Cyclin A1 (INCA) that interacted with Cyclin A1/CDK2. Linzen et al. (2015) demonstrated that ING5 was phosphorylated at T152 by Cyclin E/CDK2 and Cyclin A/CDK2, which was repressed by CDK2 inhibitor p27KIP1. Wang et al. (2012) demonstrated ING5 suppressed the proliferation of mesenchymal stem cells by down-regulating CDK2 expression. ING5 and p53 bound to amino-terminal residues 129 to 200 of Epstein-Barr virus (EBV) EBNA3C. The conserved domain of ING5 competitively interacted with both p53 and EBNA3C. EBNA3C significantly inhibited ING5-mediated regulation of p53 transcription (Saha et al., 2011) (Figure 2).

ING5 profoundly influences protein lysine acetylation with 122 of 163 acetylation peptides significantly up-regulated, and 72 of 100 acetylation peptides down-regulated by ING5. The acetylated proteins up-regulated by ING5 are preferentially located in nucleus, and act as transcription cofactors, chromatin and DNA binding factors, while those down-regulated by ING5 are principally in cytosol, and involved in cellular metabolism. ING5 promotes HAT p300 autoacetylation at K1555, K1558, K1560, K1647 and K1794, resulting in the activation of HAT and subsequent acetylation of p53 at K382 and histone H3 at K18. HAT inhibitor impairs ING5-mediated acetylation of H3K18 and p53K382, and subsequently down-regulates both p21 and Bax expression (Zhang et al., 2017b). Ormaza et al. (2019) found that ING5 acetylated histone H3 by HBO1 complex, and H4 by MOZ/MORF complex. Liu et al. (2013) found that ING5 facilitated Tip60 to acetylate p53 at K120 in response to DNA damage by the formation of p53-Tip60 complex. Acetylated p53 at K120 interacted with and activated the promoters of the apoptotic genes, Bax and GADD45. ING5 was reported to activate the cyclin-dependent kinase inhibitor p21 promoter and up-regulate p21 expression, and acetylate p53 protein at K382 (Shiseki et al., 2003) (Figure 2). Wang et al. (2018) found that ING5 enhanced the expression of stem cell markers (Oct4, Olig2 and Nestin) to promote self-renewal, prevent lineage differentiation and increase stem-like population of brain tumor initiating cells.

The biological effects of ING5 expression on cancer cells

ING5 inhibits proliferation, glucose catabolism, anti-apoptosis, migration, invasion, epithelial- mesenchymal transition (EMT) or lung metastasis, and induces cell

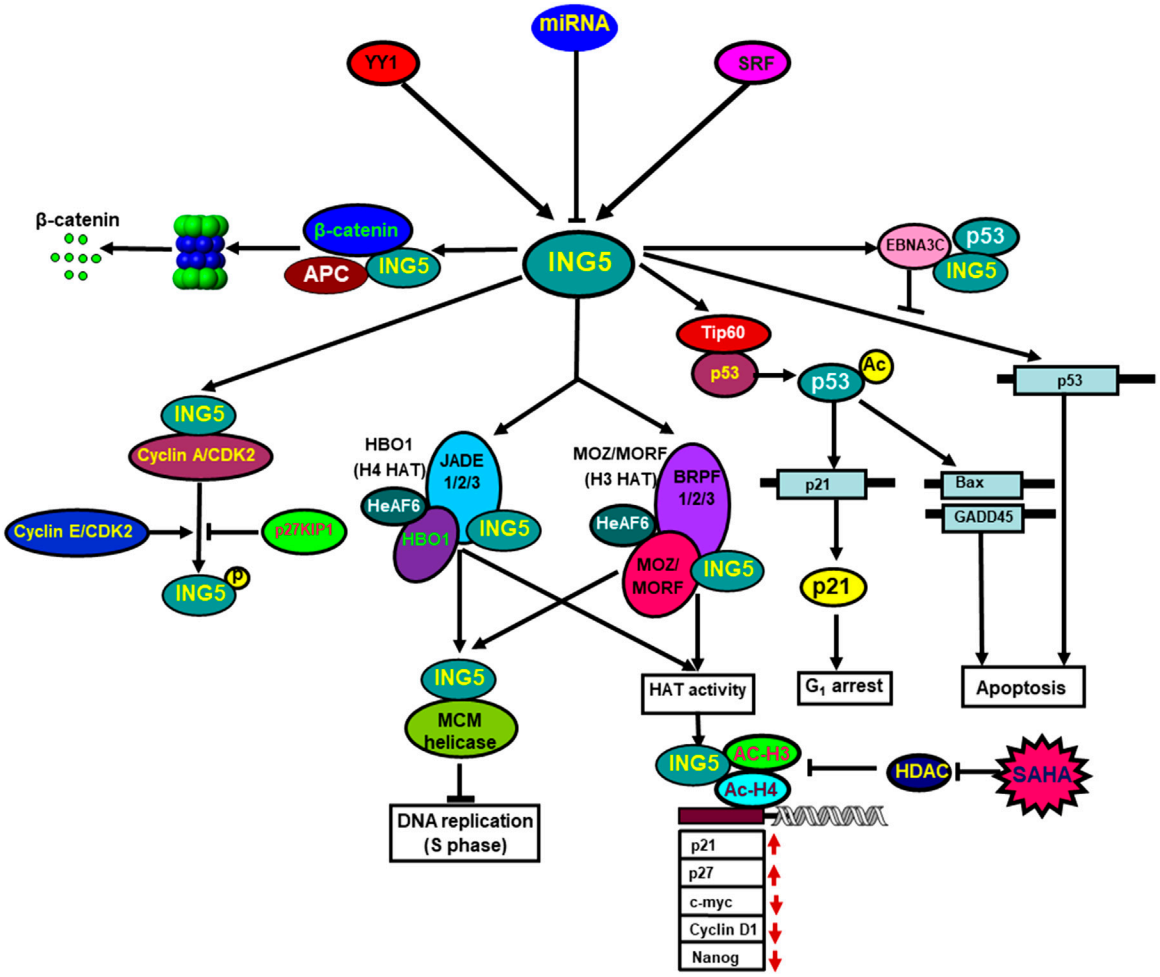


FIGURE 2
The biological functions of ING5. Transcriptionally, YY1 and SRF can up-regulate ING5 mRNA expression by the interaction of YY1-SRF-p53-ING5 complex with ING5 promoter. Translationally, ING5 expression is suppressed by as such miRNA targets as miR-196a, miR-193a-3p, miR-27-3p, miR-1307, miR-193, miR-222, miR-331-3p, miR-543 and so forth. ING5 interacts with histone H3K4me3 and is involved in the formation of two different HAT (histone acetyltransferase) complexes, (H4-HBO1-JADE-ING5 and H3-MOZ-MORF-BRPF-ING5), which contribute to DNA replication *via* ING5 and MCM complexes. ING5 binds to Cyclin A1/CDK2, and is phosphorylated by Cyclin E/CDK2 and Cyclin A/CDK2, which is repressed by CDK2 inhibitor p27KIP1. ING5 interacts with Epstain-Barr virus EBNA3C, which suppresses ING5-mediated transcription of p53. Additionally, ING5 assisted Tip60 to acetylate p53, and acetylated p53 at K120 binds to and activates the promoters of Bax and GADD45. ING5 also activates the p21 promoter to induce p21 expression *via* p53 acetylation at Lys-382. HAT and SAHA exposure recruited ING5 and acetylated histones H3 and H4 to the promoters of c-myc, nanog, Cyclin D1, p21, and p27, thereby regulating their transcription.

arrest, senescence, autophagy, fat accumulation, and chemoresistance in gastric, colorectal, breast, lung, ovarian cancer cells, neuroblastoma and glioma cells (Gou et al., 2015; Zhao et al., 2016; Ding et al., 2017; Zhao et al., 2017; Zheng et al., 2017; Wu et al., 2018; Yu et al., 2022; Zheng et al., 2022). ING5 inhibited tumor growth, blood supply or lung metastasis of gastric, colorectal, ovarian, breast or lung cancer cells by suppressing proliferation, and inducing autophagy and apoptosis in nude xenograft models. ING5-mediated chemoresistance was closely linked to anti-apoptosis, overexpression of chemoresistant genes, the activation of PI3K/Akt/NF-κB and Wnt/β-catenin signal pathways (Gou et al., 2015; Zhao et al., 2016; Ding et al., 2017; Zhao et al., 2017; Zheng et al., 2017; Wu et al., 2018; Yu et al., 2022). Qi and Zhang (2014) found that intact ING5 inhibited the proliferation and anti-apoptosis in tongue squamous cell carcinoma cells. In addition, 2 mutants of ING5 (aa 107-226 and 1-184) can facilitate cellular senescence with Cyclin E and CDK2 hypoexpression. In our study, we found that all wild-type and mutant ING5 had either the LZL or PHD domain, and weakened proliferation, migration and invasion of gastric cancer cells, which was closely linked to cdc-2, VEGF, and MMP-9 hypoexpression (Zheng et al., 2022).

ING5 was also demonstrated to inhibit proliferation, glucose catabolism, migration, invasion, EMT, and tumor growth, and promote apoptosis, cell cycle arrest, senescence, or autophagy in prostate cancer cells by activating p53 and inactivating Akt (Barlak et al., 2020), in lung cancer cells by phosphorylating β-catenin at S33/37, and suppressing IL-6/STAT3 and EGFR/PI3K/Akt pathways (Zhang et al., 2015; Liu et al., 2017; Liu et al., 2019), in neuroblastoma cells by acetylating histones (Wu et al., 2018), in colorectal cancer cells through PI3K/Akt pathway (Xin et al., 2020), in osteosarcoma cells *via* Smad pathway (Zhang et al., 2018a), in esophageal squamous carcinoma cells through IL-6/CXCL12 (Wang et al., 2021) and Akt/NF-κB/MMP-9 (Zhang et al., 2018b) pathways, and in breast cancer cells *via* PI3K/Akt/NF-κB pathway (Zhao et al., 2015; Cui et al., 2017), in glioma cells (Zhao et al., 2017) and gastric cancer cells (Gou et al., 2015) *via* either β-catenin/TCF-4 or PI3K/Akt pathway, and thyroid cancer *via* c-Met/PI3K/Akt signaling pathway (Gao and Han, 2018). These findings suggest that ING5 reverses the aggressiveness mostly through PI3K/Akt signal pathway.

Wang et al. (2018) reported that ING5 promoted stemness and self-renewal, and prevented lineage differentiation in glioblastoma cells *via* Ca^{2+} and follicle-stimulating hormone. Chen et al. (2017) demonstrated that ING5 inhibited cell proliferation and chemoresistance, and promoted cell apoptosis in ovarian cancer cells. Li et al. (2015) found that ING5 silencing increased the chemoresistance and inhibited the DNA damage response pathway in bladder cancer, while it was the converse for ING5 expression in chemoresistant bladder cancer cells. Our previous work showed that ING5 overexpression caused chemoresistance and lipogenesis of cancer cells, and chemoresistant cells to 5-FU, cisplatin and sorafenib showed ING5 overexpression and lipogenesis, suggesting that ING5-mediated lipogenesis results in drug resistance (Xuan et al., 2022).

ING5 knockdown elevated cell viability and impaired cell cycle G_0/G_1 phase arrest and apoptosis in PC-12 and SH-SY5Y cells after oxygen-glucose deprivation (Zhang et al., 2019). Wu et al. (2018) demonstrated that HAT recruited ING5 and acetylated histones H3 and H4 to the promoters of c-myc, nanog, Cyclin D1, p21, and p27, thereby up-regulating p21 and p27 expression, and down-regulating c-myc, Cyclin D1 and nanog expression. SAHA treatment attenuated the histone deacetylase inhibitor and finally strengthened the formation of ING5-acetyl- H3-acetyl-H4 (Figure 2).

The regulation of ING5 expression

At the transcriptional level, our group found out a suppressor between −2000 and −1000 bp, and an enhancer between −800 and −100 bp although two promoter sequences were predicted from -1995 to −1690 bp and from -1973 to −1400 bp. The trans-acting factors of ING5 promoters are predicted to be SRF (−717 to −678 bp), CTCF (−48 to 0 bp), YY1 (−48 to −25 bp), Sp1 (−44 to −30 bp; 32 to 20), PPAR-γ1 (−24 to −25 bp), WT1 (−10 to 1 bp), and Pax-5 (−1 to 25 bp). EMSA and ChiP indicated that only SRF and YY1 could interact with the *ING5* promoter (−50 to 0 bp). In gastric cancer cells, either SRF or YY1 silencing decreased *ING5* mRNA and protein expression. SRF was found to interact with YY1, p53 and ING5, and either SFR or YY1 was co-localized with p53 or ING5 in the nuclei of cancer cells (Zheng et al., 2022). These findings indicate that SRF, YY1, ING5 and p53 form a complex to interact with the ING5 promoter and up-regulate its expression in gastric cancer cells.

At the translational level, both miR-196 and miR-196b-5p target ING5 in colorectal cancer cells (Pourdavoud et al., 2020; Xin et al., 2020). By targeting ING5, miR-27-3p promotes the G_1-S phase transition in osteosarcoma cells (Ye et al., 2018) and miR-196b-5p inhibits the apoptosis of pancreatic cancer cells (Ma et al., 2021). In ovarian cancer cells, miR-200b/200a/429 targets ING5 which blocks miR-200b/200a/429-induced malignant transformation (Guan et al., 2020), and miR-1307 targets and down-regulated ING5 expression (Chen et al., 2017). miR-1307- mediated apoptosis and chemosensitivity were *in vivo* and vitro reversed by ING5 knockdown. CAF (cancer-associated fibroblast)-derived exosomal miR-196a confers cisplatin resistance in HNSCC cells by targeting ING5 (Qin et al., 2019), and miR-193a-3p induces the chemoresistance of breast cancer cells by targeting ING5, in line with the effects of ING5 silencing (Li et al., 2015). In triple-negative breast cancer cells, miR-193 bound to the 3′-UTR of ING5 mRNA to inhibit its expression and miR-193 inhibitor suppressed cell proliferation through ING5/PI3K/Akt pathway (Xu et al., 2020). CASC2 acted as a competitor of miR-222 to up-regulate ING5 expression in pulmonary arterial smooth muscle cells (PASMCs). CASC2-mediated inhibitory effects of hypoxia on the migration and proliferation of PASMCs could be weakened by either miR-222 inhibitor or ING5 overexpression (Han et al., 2020). miR-331-3p could promote proliferation of HCC cells by targeting ING5, which was strengthened by hepatitis B virus (HBV) (Cao et al., 2015). HBx protein of HBV accelerated the proliferation of HCC cells by up-regulating miR-181b to target ING5 (Xie et al., 2018). In neuroblastoma cells, ING5 is a negatively-regulated target gene of miR-376c-3p (Zhang et al., 2019), and SAHA down-regulates miR-196-b and miR-543 expression to facilitate the translation of ING5 protein (Wu et al., 2018). miR-193 induces proliferation and CDK2 activity in bone mesenchymal stem cells by targeting ING5 (Wang et al., 2012).

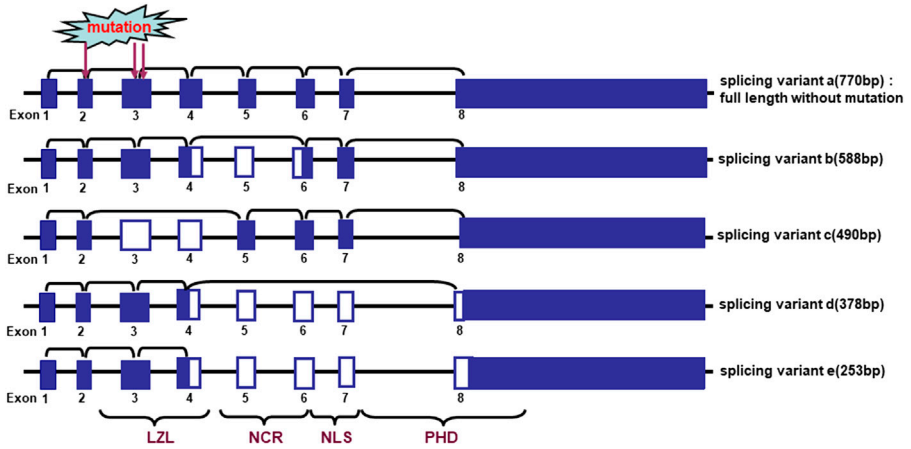


FIGURE 3
Alternative splicing of ING5 mRNA in oral squamous cell carcinoma. The schematic representation indicates the genomic structure and alternatively splicing variants of ING5. LZL, leucine zipper-Like motif; NCR: novel conserved region; NLS, nuclear localization signal; PHD, plant homeodomain.

The effects of ING5 knockout on carcinogenesis and lipogenesis

ING5 was conditionally deleted in gastric pit cells using Capn8-cre mice, in gastric parietal cells using Atp4b-cre mice, in gastric pdx-1-positive cells using pdx1-cre mice, in gastric stem-like cells using K19-cre mice, in gastric chief cells using PGC-cre mice, and in intestinal epithelial cells using villin-cre mice. The normal gastric epithelium was observed in Capn8-cre; ING5$^{f/f}$ mice and regenerative dysplasia was detected in PGC-cre; ING5$^{f/f}$ mice. However, well-, moderately-, or poorly-differentiated adenocarcinoma was observed in Atp4b-cre; ING5$^{f/f}$,pdx1-cre; ING5$^{f/f}$ and K19-cre; ING5$^{f/f}$ mice, even in pdx1-cre; ING5$^{f/f}$ and K19-cre; ING5$^{f/f}$ and wild-type (WT) mice exposed to MNU. After the treatment with MNU, chemically-induced gastric cancer and dysplasia were more common in K19-cre; ING5$^{f/f}$ and Pdx1-cre; ING5$^{f/f}$ mice than those in WT mice. These data demonstrated that ING5 knockout (KO) might cause spontaneous and chemically-induced gastric carcinogenesis (Zheng et al., 2022). After exposure to AOM/DSS, the cancer formation rate of intestinal KO mice was high in comparison to that of WT mice (Yu et al., 2022).

After the oral administration of high-fat diet, villin-cre; ING5$^{f/f}$ showed a lower weight, a slighter fatty liver and a lighter fat weight than WT mice (Yu et al., 2022). He et al. (2021) found that liver-specific reconstitution of JFK expression resulted in hepatic lipid accumulation in JFK KO mice by destabilizing ING5 and inhibiting fatty acid β-oxidation. Therefore, we hypothesized that ING5 might play an important role in fat absorption and fatty acid β-oxidation.

The clinicopathological and prognostic significances of ING5 in cancers

Cengiz et al. (2007) found that 3 frequently-deleted regions at 2q37.3, 2q33-35 and 2q21-24 harboring ING5 were detected in oral cancers, and loss of heterozygosity (LOH) of chromosome 2q in 85% (33/39) of oral squamous carcinomas, indicating an important role of ING5 deletion in oral carcinogenesis (Cengiz et al., 2010). There were 5 different alternative splicing variants of ING5, which resulted in ING5 mRNA hypoexpression in 61% of the oral tumors as compared to the matched normal samples (Figure 3). These missense mutations of ING5 gene were located to NCR and LZL domains of ING5 protein (Cengiz et al., 2007; Cengiz et al., 2010).

In our previous study, ING5 protein is primarily observed the cytoplasm and nucleus in the breast, intestine, stomach, and tongue, lung. Totally, ING5 expression was detected in 40.6% (400/986) of cancer tissues, among which ING5 was more frequently expressed in colorectal (56.3%), breast (79.9%) and endometrial (50.0%) cancers, and less in pancreatic cancer (22.6%) and HCC (14.5%) (Yang et al., 2019). Although a high ING5 expression was observed in gastric, breast, and colorectal cancers (Xing et al., 2011; Zhao et al., 2015; Wang et al., 2021; Zheng et al., 2022), ING5 was found to be expressed in osteosarcoma, prostate cancer, ovarian cancer, HCC, lung cancer, esophageal cancer, and thyroid cancer at a low level (Qi and Zhang, 2014; Cao et al., 2015; Zhang et al., 2015; Cui et al., 2017; Zheng et al., 2017; Zhang et al., 2018a). The low nuclear expression of ING5 and its nucleocytoplasmic translocation were involved in the tumorigenesis of gastric cancer, breast cancer, HNSCC, and colorectal cancer (Li et al., 2010; Xing et al., 2011;

Zheng et al., 2011; Ding et al., 2017). Bilgiç et al. (2018) found that ING5 was under-expressed in the atrophic gastritis and gastric cancer compared with the normal mucosa.

As for the clinicopathological and prognostic significances, ING5 mRNA expression was negatively correlated with vascular and lymphatic invasion, and clinicopathological staging of ovarian cancers. There was an adverse relationship between ING5 mRNA expression and the overall or progression-free survival of the ovarian cancer patients with stage I, Grade 3, and Grade 2, indicating the prognostic significance of ING5 depends on the cancer type, subtyping and molecular level (Zheng et al., 2017). In gastric, colorectal and breast cancers, nuclear ING5 expression was negatively associated with depth of invasion, tumor size, dedifferentiation, and TNM staging, opposite to the finding of cytoplasmic ING5 expression (Xing et al., 2011; Zheng et al., 2011; Ding et al., 2017). ING5 immunoreactivity was less expressed in ovarian mucinous and serous adenocarcinoma than miscellaneous subtypes, and negatively correlated with differentiation and low ki-67 expression of ovarian cancer (Zhang et al., 2018b). In lung cancer, ING5 protein was distinctly expressed in small cell carcinoma < large cell carcinoma < adenocarcinoma < squamous cell carcinoma (Zhao et al., 2016). The nuclear ING5 expression was positively linked to a better prognosis of gastric cancer patients, albeit not independent (Xing et al., 2011), in agreement with the report of Zhang et al. about lung cancer (Zhang et al., 2015).

Conclusion and perspectives

In summary, ING5 protein can acetylate HAT p300, p53, histone H3 and H4 for the transcription of downstream genes *via* the formation of homodimer or heterodimers with ING4, histone H3K4me3, HAT complex, Tip60, Cyclin A1/CDK2, INCA1 and EBNA3C. Transcriptionally, YY1 and SRF can up-regulate ING5 mRNA expression by the interaction of YY1-SRF-p53-ING5 complex with ING5 promoter. ING5 inhibits proliferation, migration, invasion and tumor growth of cancer cells *via* IL-6/STAT3, EGFR/Akt/NF-κB/MMP-9 or IL-6/CXCL12 pathways. ING5-mediated lipogenesis might lead to chemoresistance. Target ING5 KO increased the sensitivity of chemically-induced gastric and colorectal carcinogenesis, and suppressed lipogenesis and fat absorption in intestine. The overexpression and nucleocytoplasmic translocation were seen and ING5 expression is inversely associated with aggressive behaviors and poor prognosis of cancers.

In the future, aberrant ING5 expression might be used as a molecular marker for carcinogenesis, progression and prognosis of malignancies. Additionally, ING5 functions as a tumor suppressor, but facilitates lipogenesis and subsequent chemoresistance. If ING5 will be used as a gene therapy target of cancers, the chemoresistant function should be ameliorated. Therefore, ING5 knockout or the suppression of ING5-related signal pathway would also be helpful to the control of body weight and the prevention of obesity-related diseases. If we can develop the anti-ING5 reagents, these drugs would be used to treat cancer, reverse chemoresistance and control the obesity.

Author contributions

Writing -original draft, H-cZ and HX; writing -review and editing, H-mJ and H-cZ; funding acquisition, H-cZ.

Funding

The present study was supported by Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (21377772D), Emphasis Project of Education Department of Hebei Province (ZD2022096) and National Natural Scientific Foundation of China (81672700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Archambeau, J., Blondel, A., and Pedeux, R. (2019). Focus-ING on DNA Integrity: Implication of ING proteins in cell cycle regulation and DNA repair modulation. *Cancers* 12 (1), 58. doi:10.3390/cancers12010058

Barlak, N., Capik, O., Sanli, F., Kilic, A., Aytatli, A., Yazici, A., et al. (2020). ING5 inhibits cancer aggressiveness by inhibiting Akt and activating p53 in prostate cancer. *Cell Biol. Int.* 44 (1), 242–252. doi:10.1002/cbin.11227

Bilgiç, F., Gerçeker, E., Boyacıoğlu, S. Ö., Kasap, E., Demirci, U., Yıldırım, H., et al. (2018). Potential role of chromatin remodeling factor genes in atrophic gastritis/gastric cancer risk. *Turk. J. Gastroenterol.* 29 (4), 427–435. doi:10.5152/tjg.2018.17350

Cao, Y., Chen, J., Wang, D., Peng, H., Tan, X., Xiong, D., et al. (2015). Upregulated in Hepatitis B virus-associated hepatocellular carcinoma cells, miR-331-3p promotes proliferation of hepatocellular carcinoma cells by targeting ING5. *Oncotarget* 6 (35), 38093–38106. doi:10.18632/oncotarget.5642

Cengiz, B., Gunduz, M., Nagatsuka, H., Beder, L., Gunduz, E., Tamamura, R., et al. (2007). Fine deletion mapping of chromosome 2q21-37 shows three preferentially deleted regions in oral cancer. *Oral Oncol.* 43 (3), 241–247. doi:10.1016/j.oraloncology.2006.03.004

Cengiz, B., Gunduz, E., Gunduz, M., Beder, L. B., Tamamura, R., Bagci, C., et al. (2010). Tumor-specific mutation and downregulation of ING5 detected in oral squamous cell carcinoma. *Int. J. Cancer* 127 (9), 2088–2094. doi:10.1002/ijc.25224

Chen, W. T., Yang, Y. J., Zhang, Z. D., An, Q., Li, N., Liu, W., et al. (2017). MiR-1307 promotes ovarian cancer cell chemoresistance by targeting the ING5 expression. *J. Ovarian Res.* 10 (1), 1. doi:10.1186/s13048-016-0301-4

Cui, S., Liao, X., Ye, C., Yin, X., Liu, M., Hong, Y., et al. (2017). ING5 suppresses breast cancer progression and is regulated by miR-24. *Mol. Cancer* 16 (1), 89. doi:10.1186/s12943-017-0658-z

Dantas, A., Al Shueili, B., Yang, Y., Nabbi, A., Fink, D., and Riabowol, K. (2019). Biological functions of the ING proteins. *Cancers (Basel)* 11 (11), 1817. doi:10.3390/cancers11111817

Ding, X. Q., Zhao, S., Yang, L., Zhao, X., Zhao, G. F., Zhao, S. P., et al. (2017). The nucleocytoplasmic translocation and up-regulation of ING5 protein in breast cancer: a potential target for gene therapy. *Oncotarget* 8 (47), 81953–81966. doi:10.18632/oncotarget.17918

Gao, W., and Han, J. (2018). Overexpression of ING5 inhibits HGF-induced proliferation, invasion and EMT in thyroid cancer cells via regulation of the c-Met/PI3K/Akt signaling pathway. *Biomed. Pharmacother.* 98 (33), 265–270. doi:10.1016/j.biopha.2017.12.045

Gou, W. F., Shen, D. F., Yang, X. F., Zhao, S., Liu, Y. P., Sun, H. Z., et al. (2015). ING5 suppresses proliferation, apoptosis, migration and invasion, and induces autophagy and differentiation of gastric cancer cells: a good marker for carcinogenesis and subsequent progression. *Oncotarget* 6 (23), 19552–19579. doi:10.18632/oncotarget.3735

Guan, W., Cui, H., Huang, P., Chun, W. J., Lee, J. W., Kim, H., et al. (2020). miR-200b/200a/429 cluster stimulates ovarian cancer development by targeting ING5. *J. Oncol.* 2020, 3404059. doi:10.1155/2020/3404059

Han, Y., Liu, Y., Yang, C., Gao, C., Guo, X., and Cheng, J. (2020). LncRNA CASC2 inhibits hypoxia- induced pulmonary artery smooth muscle cell proliferation and migration by regulating the miR-222/ING5 axis. *Cell. Mol. Biol. Lett.* 25, 21. doi:10.1186/s11658-020-00215-y

He, L., Yan, R., Yang, Z., Zhang, Y., Liu, X., Yang, J., et al. (2021). SCF JFK is functionally linked to obesity and metabolic syndrome. *EMBO Rep.* 22 (7), e52036. doi:10.15252/embr.202052036

Li, X., Nishida, T., Noguchi, A., Zheng, Y., Takahashi, H., Yang, X., et al. (2010). Decreased nuclear expression and increased cytoplasmic expression of ING5 may be linked to tumorigenesis and progression in human head and neck squamous cell carcinoma. *J. Cancer Res. Clin. Oncol.* 136 (10), 1573–1583. doi:10.1007/s00432-010-0815-x

Li, Y., Deng, H., Lv, L., Zhang, C., Qian, L., Xiao, J., et al. (2015). The miR-193a-3p-regulated ING5 gene activates the DNA damage response pathway and inhibits multi-chemoresistance in bladder cancer. *Oncotarget* 6 (12), 10195–10206. doi:10.18632/oncotarget.3555

Linzen, U., Lilischkis, R., Pandithage, R., Schilling, B., Ullius, A., Lüscher-Firzlaff, J., et al. (2015). ING5 is phosphorylated by CDK2 and controls cell proliferation independently of p53. *PLoS One* 10, e0123736. doi:10.1371/journal.pone.0123736

Liu, N., Wang, J., Wang, J., Wang, R., Liu, Z., Yu, Y., et al. (2013). ING5 is a Tip60 cofactor that acetylates p53 in response to DNA damage. *Cancer Res.* 73 (12), 3749–3760. doi:10.1158/0008-5472.CAN-12-3684

Liu, X. L., Zhang, X. T., Meng, J., Zhang, H. F., Zhao, Y., Li, C., et al. (2017). ING5 knockdown enhances migration and invasion of lung cancer cells by inducing EMT via EGFR/PI3K/Akt and IL-6/STAT3 signaling pathways. *Oncotarget* 8 (33), 54265–54276. doi:10.18632/oncotarget.17346

Liu, X. L., Meng, J., Zhang, X. T., Liang, X. H., Zhang, F., Zhao, G. R., et al. (2019). ING5 inhibits lung cancer invasion and epithelial-mesenchymal transition by inhibiting the WNT/β-catenin pathway. *Thorac. Cancer* 10 (4), 848–855. doi:10.1111/1759-7714.13013

Ma, D., Chen, S., Wang, H., Wei, J., Wu, H., Gao, H., et al. (2021). Baicalein induces apoptosis of pancreatic cancer cells by regulating the expression of miR-139-3p and miR-196b-5p. *Front. Oncol.* 11, 653061. doi:10.3389/fonc.2021.653061

Ormaza, G., Rodríguez, J. A., Ibáñez de Opakua, A., Merino, N., Villate, M., Gorroño, I., et al. (2019). The tumor suppressor ING5 is a dimeric, bivalent recognition molecule of the histone H3K4me3 mark. *J. Mol. Biol.* 431 (12), 2298–2319. doi:10.1016/j.jmb.2019.04.018

Pourdavoud, P., Pakzad, B., Mosallaei, M., Saadatian, Z., Esmaeilzadeh, E., Alimolaie, A., et al. (2020). MiR-196: emerging of a new potential therapeutic target and biomarker in colorectal cancer. *Mol. Biol. Rep.* 47 (12), 9913–9920. doi:10.1007/s11033-020-05949-8

Qi, L., and Zhang, Y. (2014). Truncation of inhibitor of growth family protein 5 effectively induces senescence, but not apoptosis in human tongue squamous cell carcinoma cell line. *Tumour Biol.* 35 (4), 3139–3144. doi:10.1007/s13277-013-1410-y

Qin, X., Guo, H., Wang, X., Zhu, X., Yan, M., Wang, X., et al. (2019). Exosomal miR-196a derived from cancer-associated fibroblasts confers cisplatin resistance in head and neck cancer through targeting CDKN1B and ING5. *Genome Biol.* 20 (1), 12. doi:10.1186/s13059-018-1604-0

Saha, A., Bamidele, A., Murakami, M., and Robertson, E. S. (2011). EBNA3C attenuates the function of p53 through interaction with inhibitor of growth family proteins 4 and 5. *J. Virol.* 85 (5), 2079–2088. doi:10.1128/JVI.02279-10

Shiseki, M., Nagashima, M., Pedeux, R. M., Kitahama-Shiseki, M., Miura, K., Okamura, S., et al. (2003). p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. *Cancer Res.* 63 (10), 2373–2378.

Wang, J., Huang, W., Wu, Y., Hou, J., Nie, Y., Gu, H., et al. (2012). MicroRNA-193 pro-proliferation effects for bone mesenchymal stem cells after low-level laserirradiation treatment through inhibitor of growth family, member 5. *Stem Cells Dev.* 21 (13), 2508–2519. doi:10.1089/scd.2011.0695

Wang, F., Wang, A. Y., Chesnelong, C., Yang, Y., Nabbi, A., Thalappilly, S., et al. (2018). ING5 activity in self-renewal of glioblastoma stem cells via calcium and follicle stimulating hormone pathways. *Oncogene* 37 (3), 286–301. doi:10.1038/onc.2017.324

Wang, Y., Tan, J., Li, J., Chen, H., and Wang, W. (2021). ING5 inhibits migration and invasion of esophageal cancer cells by downregulating the IL-6/CXCL12 signaling pathway. *Technol. Cancer Res. Treat.* 20, 15330338211039940. doi:10.1177/15330338211039940

Wu, J. C., Jiang, H. M., Yang, X. H., and Zheng, H. C. (2018). ING5-mediated antineuroblastoma effects of suberoylanilide hydroxamic acid. *Cancer Med.* 7, 4554–4569. doi:10.1002/cam4.1634

Xie, X., Xu, X., Sun, C., and Yu, Z. (2018). Hepatitis B virus X protein promotes proliferation of hepatocellular carcinoma cells by upregulating miR-181b by targeting ING5. *Biol. Chem.* 399 (6), 611–619. doi:10.1515/hsz-2018-0178

Xin, H., Wang, C., Chi, Y., and Liu, Z. (2020). MicroRNA-196b-5p promotes malignant progression of colorectal cancer by targeting ING5. *Cancer Cell Int.* 20, 119. doi:10.1186/s12935-020-01200-3

Xing, Y. N., Yang, X., Xu, X. Y., Zheng, Y., Xu, H. M., Takano, Y., et al. (2011). The altered expression of ING5 protein is involved in gastric carcinogenesis and subsequent progression. *Hum. Pathol.* 42 (1), 25–35. doi:10.1016/j.humpath.2010.05.024

Xu, J. H., Zhao, J. X., Jiang, M. Y., Yang, L. P., Sun, M. L., and Wang, H. W. (2020). MiR-193 promotes cell proliferation and invasion by ING5/PI3K/AKT pathway of triple-negative breast cancer. *Eur. Rev. Med. Pharmacol. Sci.* 24 (6), 3122–3129. doi:10.26355/eurrev_202003_20679

Xuan, Y., Shi, S., Xue, H., Liu, K., Han, W. L., Li, B. C., et al. (2022). Lamp2a facilitates lipid droplet production and chemoresistance of colorectal cancer cells by maintaining ING5 protein stability. *BMC Cancer*. in press. doi:10.21203/rs.3.rs-1373107/v1

Yang, X. F., Shen, D. F., Zhao, S., Ren, T. R., Gao, Y., Shi, S., et al. (2019). Expression pattern and level of ING5 protein in normal and cancer tissues. *Oncol. Lett.* 17 (1), 63–68. doi:10.3892/ol.2018.9581

Ye, P., Ke, X., Zang, X., Sun, H., Dong, Z., Lin, J., et al. (2018). Up-regulated miR-27-3p promotes the G1-S phase transition by targeting inhibitor of growth family member 5 in osteosarcoma. *Biomed. Pharmacother.* 101, 219–227. doi:10.1016/j.biopha.2018.02.066

Yu, Q., Xue, H., Sun, H. Z., Ha, M. W., Yu, D. Y., Zhang, R., et al. (2022). The suppressive effects of ING5 on tumorigenesis and aggressiveness of colorectal cancer via inhibition of NF-κB and β-catenin pathways. *Tumor Biol.* in press.

Zhang, F., Bäumer, N., Rode, M., Ji, P., Zhang, T., Berdel, W. E., et al. (2011). The inhibitor of growth protein 5 (ING5) depends on INCA1 as a co-factor for its antiproliferative effects. *PLoS One* 6, e21505. doi:10.1371/journal.pone.0021505

Zhang, F., Zhang, X., Meng, J., Zhao, Y., Liu, X., Liu, Y., et al. (2015). ING5 inhibits cancer aggressiveness via preventing EMT and is a potential prognostic biomarker for lung cancer. *Oncotarget* 6 (18), 16239–16252. doi:10.18632/oncotarget.3842

Zhang, H., Zhou, J., Zhang, M., Yi, Y., and He, B. (2019). Upregulation of miR-376c-3p alleviates oxygen-glucose deprivation-induced cell injury by targeting ING5. *Cell. Mol. Biol. Lett.* 24, 67. doi:10.1186/s11658-019-0189-2

Zhang, G. J., Zhao, J., Jiang, M. L., and Zhang, L. C. (2018b). ING5 inhibits cell proliferation and invasion in esophageal squamous cell carcinoma through regulation of the Akt/NF-κB/MMP-9 signaling pathway. *Biochem. Biophys. Res. Commun.* 496 (2), 387–393. doi:10.1016/j.bbrc.2018.01.045

Zhang, R., Jin, J., Shi, J., and Hou, Y. (2017a). INGs are potential drug targets for cancer. *J. Cancer Res. Clin. Oncol.* 143 (2), 189–197. doi:10.1007/s00432-016-2219-z

Zhang, T., Meng, J., Liu, X., Zhang, X., Peng, X., Cheng, Z., et al. (2017b). ING5 differentially regulates protein lysine acetylation and promotes p300 autoacetylation. *Oncotarget* 9 (2), 1617–1629. doi:10.18632/oncotarget.22176

Zhang, X., Xu, Z. H., Xie, H., Sun, Y. W., Liu, J., and Zhao, Y. B. (2018a). ING5 is a potential target for osteosarcoma therapy. *Technol. Cancer Res. Treat.* 17, 1533033818762680. doi:10.1177/1533033818762680

Zhao, Q. Y., Ju, F., Wang, Z. H., Ma, X. Z., and Zhao, H. (2015). ING5 inhibits epithelial-mesenchymal transition in breast cancer by suppressing PI3K/Akt pathway. *Int. J. Clin. Exp. Med.* 8 (9), 15498–15505.

Zhao, S., Yang, X. F., Shen, D. F., Gao, Y., Shi, S., Wu, J. C., et al. (2016). The down-regulated ING5 expression in lung cancer: a potential target of gene therapy. *Oncotarget* 7 (34), 54596–54615. doi:10.18632/oncotarget.10519

Zhao, S., Zhao, Z. J., He, H. Y., Wu, J. C., Ding, X. Q., Yang, L., et al. (2017). The roles of ING5 in gliomas: a good marker for tumorigenesis and a potential target for gene therapy. *Oncotarget* 8 (34), 56558–56568. doi:10.18632/oncotarget.17802

Zheng, H. C., Xia, P., Xu, X. Y., Takahashi, H., and Takano, Y. (2011). The nuclear to cytoplasmic shift of ING5 protein during colorectal carcinogenesis with their distinct links to pathologic behaviors of carcinomas. *Hum. Pathol.* 42 (3), 424–433. doi:10.1016/j.humpath.2009.12.018

Zheng, H. C., Zhao, S., Song, Y., and Ding, X. Q. (2017). The roles of ING5 expression in ovarian carcinogenesis and subsequent progression: a target of gene therapy. *Oncotarget* 8 (61), 103449–103464. doi:10.18632/oncotarget.21968

Zheng, H. C., Xue, H., Wu, X., Xu, H. L., Zhao, E. H., and Cui, Z. G. (2022). Transcriptional regulation of ING5 and its suppressive effects on gastric cancer. *Front. Oncol.* 12, 918954. doi:10.3389/fonc.2022.918954

frontiers | Frontiers in Cell and Developmental Biology

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fcell.2022.1035033

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Xuguang Yang,
Shanghai University of Traditional Chinese Medicine, China
Peng-Fei Zhang,
Fudan University, China

*CORRESPONDENCE
Feng Jiao,
jiaofeng@renji.com
Ming Quan,
mquan@tongji.edu.cn

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 02 September 2022
ACCEPTED 01 December 2022
PUBLISHED 15 December 2022

CITATION
Li N, Chen Z, Huang M, Zhang D, Hu M, Jiao F and Quan M (2022), Detection of *ROS1* gene fusions using next-generation sequencing for patients with malignancy in China.
Front. Cell Dev. Biol. 10:1035033.
doi: 10.3389/fcell.2022.1035033

Detection of *ROS1* gene fusions using next-generation sequencing for patients with malignancy in China

Ning Li[1], Zhiqin Chen[2], Mei Huang[3], Ding Zhang[4], Mengna Hu[4], Feng Jiao[5]* and Ming Quan[2]*

[1]Department of Oncology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Oncology, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai, China, [3]Department of Oncology, Yancheng Third People's Hospital, Yancheng, China, [4]The Medical Department, 3D Medicines Co., Ltd., Shanghai, China, [5]Department of Oncology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Objective: This study aimed to identify *ROS1* fusion partners in Chinese patients with solid tumors.

Methods: Next-generation sequencing (NGS) analysis was used to detect *ROS1* rearrangement in 45,438 Chinese patients with solid tumors between 2015 and 2020, and the clinical characteristics and genetic features of gene fusion were evaluated. H&E staining of the excised tumor tissues was conducted. Samples with a tumor cell content ≥ 20% were included for subsequent DNA extraction and sequencing analysis.

Results: A total of 92 patients with *ROS1* rearrangements were identified using next-generation sequencing, and the most common histological type lung cancer. From the 92 *ROS1* fusion cases, 24 *ROS1* fusion partners had been identified, including 14 novel partners and 10 reported partners. Of these, *CD74*, *EZR*, *SDC4*, and *TPM3* were the four most frequently occurring partners. Fourteen novel *ROS1* fusion partners were detected in 16 patients, including *DCBLD1-ROS1*, *FRK-ROS1*, and *VGLL2-ROS1*. In many patients, the *ROS1* breakpoint was located between exons 32 and 34.

Conclusion: This study describes 14 novel *ROS1* fusion partners based on the largest *ROS1* fusion cohort, and the *ROS1* breakpoint was mostly located between exons 32 and 34. Additionally, next-generation sequencing is an optional method for identifying novel *ROS1* fusions.

KEYWORDS

solid tumor, next-generation sequencing, ROS1 fusions partners, ROS1 breakpoint, lung cancer

Highlights

1. This study detected *ROS1* fusion partners and the *ROS1* fusion breakpoint in solid tumors of Chinese patients in the largest *ROS1* fusion cohort to date.

2. A total of 14 novel *ROS1* fusions in solid tumor of Chinese patients were identified.
3. The majority of patients had a *ROS1* breakpoint between exons 32 and 34.

Introduction

Several gene-targeting therapies have been developed to treat human malignancies. With an in-depth study of the biological mechanism of tumor cells, the status of target genes in gene targeting therapy, such as *EGFR*, *MET*, and *ROS1*, has gradually become prominent (Cao et al., 2020). ROS1 is a receptor tyrosine kinase whose activation is reported to be linked to the growth and proliferation of malignant tumors. Previous studies have reported that *ROS1* undergoes gene rearrangement in many malignant tumors, such as lung cancer and liver cancer. In addition, *ROS1* rearrangements rarely overlap with alterations in EGFR, KRAS, or other targeted oncogenes (Zhu et al., 2018). *ROS1* was originally identified in 1986 as a viral proto-oncogene with unique oncogenic effects in the UR2 avian sarcoma virus. Hybridization analysis revealed that *ROS1* was located in the human chromosome region 6ql6–q22 and was further positioned on chromosome 6q22.1 (Birchmeier et al., 1986; Matsushime et al., 1986).

ROS1 gene fusion expression can drive cell proliferation and induce malignant transformation, which is common in many tumor cells, such as malignant gliomas (Sievers et al., 2021). ROS1 fusions occur in 1%–2% of NSCLC, and the prevalence is higher among patients in Asian countries (including China) than in Western countries (Cai and Su, 2013; Davies and Doebele, 2013; Melosky et al., 2021). A synthetic study has introduced the role of *ROS1* in various cancers. It has been reported that 26 genes fuse with *ROS1*. With the advances in sequencing technology, many new genes have been reported to fuse with *ROS1*. Natural *ROS1* rearrangement was first found in the human brain glioblastoma cell line U118MG (Birchmeier et al., 1987; Charest et al., 2003). The deletion of chromosome six led to the fusion of the *ROS1* gene into the *FIG* gene, which was observed in samples from patients with hepatobiliary carcinoma and ovarian cancer (Birch et al., 2011; Suehara et al., 2012). Numerous studies have shown that crizotinib achieves good results in NSCLC patients with positive *ROS1* rearrangement (Shaw et al., 2011; Camidge et al., 2012). AP26113, a specific inhibitor of *ROS1*, exhibits additional inhibitory activity against oncogenic *ROS1* fusions that are involved in the clinical treatment of patients with advanced solid tumors (Anjum et al., 2012). Thus, it is important to detect and evaluate *ROS1* rearrangements and gene fusion in patients with malignancies.

Currently, there are many methods for detecting *ROS1* fusion genes, including qRT-PCR, FISH, IHC, and NGS (Sakai et al., 2019). These methods have advantages and drawbacks. For example, owing to the continuous increase in *ROS1* gene fusion partners, some positive cases may be missed by RT-PCR (Shaw et al., 2011). Fluorescence *in situ* hybridization (FISH) requires advanced technology, which limits its use. Immunohistochemistry (IHC) is suitable for screening existing fusion expression, whereas only NGS can detect novel *ROS1* fusion partners (Zhang et al., 2019). More than 30 *ROS1* fusion gene partners have been reported in lung cancer, glioma, and hepatic angiosarcoma, containing *CD74*, *SLC34A2*, *GOPC* (Zhu et al., 2019). NGS is regarded as a powerful tool for detecting *ROS1* rearrangements because of its accuracy, sensitivity, and specificity.

Although some studies have reported fusion partners of *ROS1*, more novel fusions are still being reported. In addition, the distribution of *ROS1* fusions varies among different types of cancer. Exploring *ROS1* fusions in different cancers may help identify more precise therapies. Therefore, this study aimed to detect the *ROS1* partners of Chinese patients with solid tumors using NGS. This study may offer a novel understanding of the treatment of solid tumors based on the *ROS1* fusion profile.

Material and methods

Patient information

A total of 45,438 Chinese patients with solid tumor were treated in the 3Dmed Lab (Shanghai, China) between 2015 and 2020. *ROS1* gene fusion was confirmed by NGS. This study was approved by Shanghai East Hospital, Tongji University School of Medicine (No. 2020-093).

Sample collection and DNA extraction

Pathological results were used to screen the samples for subsequent analysis. Hematoxylin and eosin (H&E) staining of the removed tumor tissues was conducted. Samples with tumor cell content ≥ 20% were included for subsequent analysis. Genomic DNA (gDNA) was extracted using the ReliaPrep™ FFPE gDNA Miniprep System (Promega) and quantified using a Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific). The cfDNA in plasma was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen), and the gDNA in white blood cells was extracted using the QIAamp DNA Mini Kit (Qiagen).

Library preparation and targeted capture

Sequencing libraries were established as described previously (Shu et al., 2017). Probe-based hybridization was carried out on

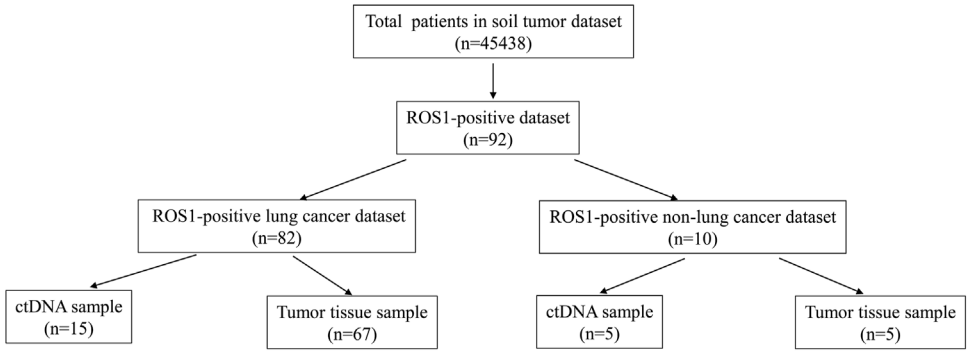


FIGURE 1
Study population. Schematic representing the population of patients in this study.

TABLE 1 Clinicopathologic features of patients with *ROS1* rearrangement (*n* = 92).

	Total (*n* = 92)	Lung cancer (*n* = 82)	CD74-ROS1 lung cancer (*n* = 35)	Non-CD74-ROS1 lung cancer (*n* = 47)
Age				
≥60	41	37	12	25
<60	51	45	23	22
Sex				
Male	34	27	13	14
Female	58	55	22	33
Histological types				
Lung adenocarcinoma	68	68	26	42
Lung Squamous Carcinoma	1	1	1	0
NA	23	13	8	5
TMB				
NA	52	46	20	26
≥10	6	6	3	3
<10	34	30	12	18
TPS				
NA	60	54	23	31
<1%	9	7	2	5
1%–49%	16	14	8	6
≥50%	7	7	2	5
CPS				
NA	74	67	29	38
<1%	2	2	0	2
1%–49%	14	11	6	5
≥50%	2	2	0	2

the index library with a custom NGS panel, in which the probe bait was a single compound 5 biotinylated 120 bp DNA oligonucleotide. All these contain introns of *ROS1* for fusion detection. Repetitive elements were screened and removed from the baits of introns, as previously described (Karolchik et al., 2004).

DNA sequencing and data processing

The extracted DNA was analyzed using a NovaSeq 6,000 platform (Illumina) to screen for targeted gene rearrangements (Xia et al., 2020). The detection approach for variants was based on a binomial test, and an R package was

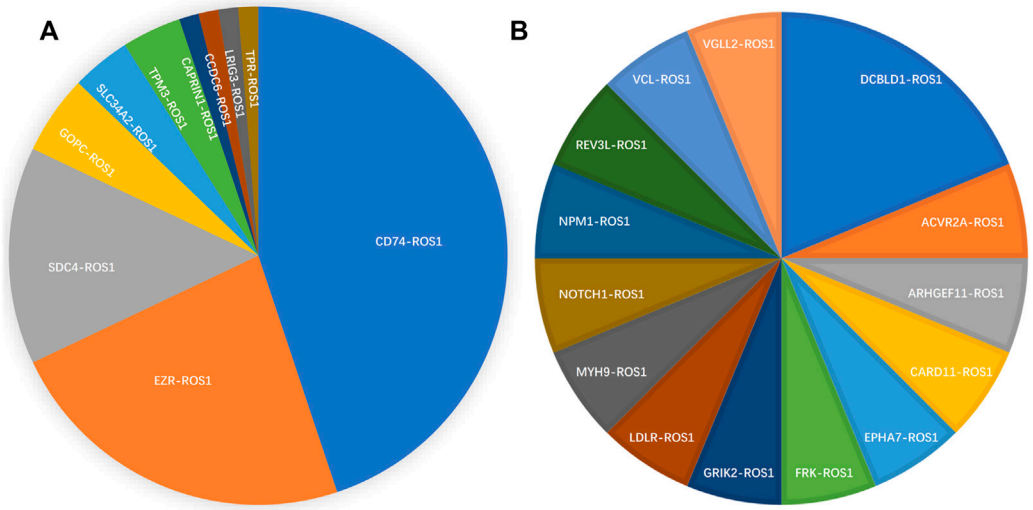


FIGURE 2
Spectrum of *ROS1* fusion partners. **(A)** 10 reported *ROS1* fusion partners. **(B)** 14 novel *ROS1* fusion partners.

developed to analyze these variants through the analysis of unique supporting read depth, strand bias, and base quality (Su et al., 2017). The variants were analyzed using an automatic false-positive filtering pipeline to ensure specificity and sensitivity when the allele frequency was ≥ 5%. ANNOVAR was performed against dbSNP (v138), 1000Genome, and ESP6500 to annotate the SNPs, insertions, and deletions. Only missense, stop-gain, frameshift, and non-frameshift indel mutations were retained for gene rearrangement analysis.

Results

Patient characteristics

A total of 45,438 solid tumor patients from the 3Dmed lab were evaluated, of which 92 patients were identified to have *ROS1* rearrangements in the blood or tumor tissues (Figure 1). The basic features of the 92 patients were analyzed. Their average age was 57 years (range, 16–82 years) and 63.0% (58/92) of the patients were female (Table 1). As shown in Supplementary Table S1, there were 82 cases of lung cancer, two cases of gastric cancer, two cases of retroperitoneal neoplasm, and one case of liver cancer, epithelioid hemangioendothelioma, liposarcoma, schwannoma, colorectal cancer, and squamous cell carcinoma. Lung adenocarcinoma was diagnosed in 78.9% (68/92) of the patients (Table 1).

ROS1 fusion partners

From 92 *ROS1* fusion cases, 24 *ROS1* fusion partners were identified, including 14 novel partners and 10 reported partners. Cluster of differentiation 74 (*CD74*), ezrin (*EZR*), syndecan 4 (*SDC4*), and tropomyosin 3 (*TPM3*) were the four most frequently occurring fusion partners. The reported *ROS1* fusion partners were identified in 75 cases (81.5%, 75/92) and their distribution was as follows: *CD74-ROS1* (38.0%, 35/92), *EZR-ROS1* (19.6%, 18/92), *SDC4-ROS1* (12.0%, 11/92), *GOPC-ROS1* (4.3%, 4/92), *SLC34A2-ROS1* (3.2%, 3/92), *TPM3-ROS1* (3.2%, 3/92), *CAPRIN1-ROS1* (1.1%, 1/92), *CCDC6-ROS1*(1.1%, 1/92), *LRIG3-ROS1*(1.1%, 1/92), and *TPR-ROS1* (1.1%, 1/92) (Figure 2A).

Fourteen novel *ROS1* fusion partners were identified in the 16 patients. Of these, novel *DCBLD1-ROS1* was observed three times in three patients with lung cancer, and other novel fusion partners were observed only once in six lung cancer cases. Moreover, seven novel partners were identified in two retroperitoneal neoplasm patients (*FRK-ROS1* and *VGLL2-ROS1*), one gastric cancer patient (*ARHGEF11-ROS1*), one liver cancer patient (*REV3L-ROS1*), one liposarcoma patient (*EPHA7-ROS1*), one patient with epithelioid hemangioendothelioma (*NOTCH1-ROS1*), and one squamous cell carcinoma patient (*CARD11-ROS1*). In addition, two patients with lung cancer harbored two *ROS1* fusions (Figure 2B). One patient had *CD74-ROS1* and *SLC34A2-ROS1* fusions, and the other patient had novel *DCBLD1-ROS1* and *GOPC-ROS1* fusions.

ROS1 fusion breakpoints

The distribution of *ROS1* fusion partners and *ROS1* breakpoints was investigated. The results showed that the *ROS1* breakpoint was mostly located between exons 32 and

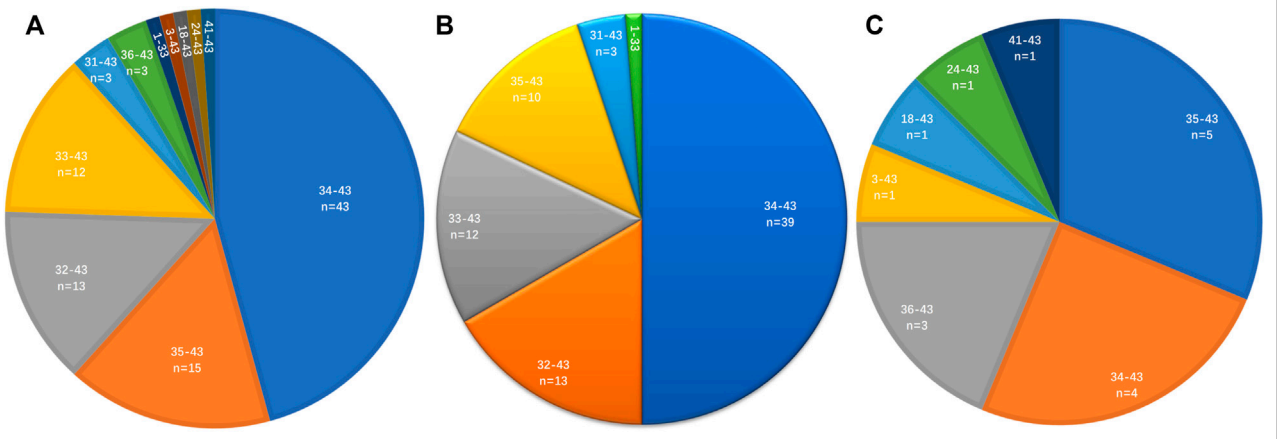


FIGURE 3
Distribution of *ROS1* breakpoints. **(A)** *ROS1* breakpoints of *ROS1* fusion patients. **(B)** *ROS1* breakpoints of reported *ROS1* fusion types. **(C)** *ROS1* breakpoints of novel *ROS1* fusion types in patients.

34, which was more apparently in reported *ROS1* fusion types (Figures 3A–C), without affecting the transmembrane and tyrosine kinase domains of the ROS1 protein. A breakpoint at exon 35 was found in 15 patients, which disrupted the transmembrane domain (Tables 2, 3).

Discussion

Gene fusion is an ideal target for cancer therapy; therefore, reliable detection of gene fusion is necessary. Many methods have been developed for detecting gene fusion, such as FISH, IHC, qRT-PCR, and NGS, which have different characteristics and applications. In clinical practice, FISH is considered the gold standard for detecting *ROS1* fusions (Chiari et al., 2020; Shen et al., 2020). Since FISH is a confirmation test, it may be applied to detect existing *ROS1* fusions with high accuracy. However, it is challenging to identify novel ROS1 partners using FISH and simultaneous detection of *ROS1* fusion is limited. Due to this facile manipulation, IHC seems to be user-friendly, cost-effective, and highly sensitive. However, it is mostly a complementary tool to other methods, since the results should be observed and analyzed by skilled personnel. Similar to FISH and IHC, qRT-PCR can be used to detect reported fusions, and it exhibits better throughput, sensitivity, and specificity. NGS is an emerging tool for screening *ROS1* fusions. A unique advantage of NGS is that it is an innovative and optimal assay for identifying novel fusions. Woo et al. detected glioblastomas harboring gene fusions using NGS in 356 diffuse gliomas. They identified 53 patients harboring various oncogenic gene fusions, including MET, EGFR, and FGFR. They also identified two patients with novel CCDC6-RET fusions (Woo et al., 2020). NGS is a promising approach for obtaining the distribution information of fusions. Thus, our study also used NGS fusion assays to detect the profiles of *ROS1* rearrangement in 45,438 patients with malignancies in China, which is the largest *ROS1* fusion cohort screened to date, which led to reliable results. It is important to screen for new *ROS1* fusions, which can be further confirmed and applied as therapeutic targets.

Another advantage of NGS for detecting *ROS1* fusion is that blood samples can be used, including plasma and white blood cells. In this study, positive *ROS1* fusion was observed in the blood samples of patients with lung cancer (n = 15) and other cancers (n = 5). This indicated that *ROS1* fusions could also be detected in plasma samples, even for the detection of novel fusion types, which was the same as that in tissue samples.

ROS1 fusions were identified in 92 patients. A total of 10 reported and 14 novel *ROS1* fusions were found in the solid tumors of Chinese patients with various malignancies. *CD74*, *EZR*, *SDC4*, and *TPM3* were 4 most frequently occurring fusion partners. Matsuura et al. (2013) detected common fusion genes in 114 NSCLCs using RT-PCR. They found that the *CD74-ROS1* fusion was involved in the carcinogenesis of a subpopulation of NSCLC, which may assist in clarifying the features of tumors and guiding treatment. The *CD74-ROS1* fusion gene was reported for the first time in an inflammatory breast cancer patient by NGS (Hu et al., 2021). Cui et al. found that *CD74-ROS1* was the most frequently occurring fusion protein in NSCLC (Cui et al., 2020), which is in line with the results of this study. Tatjana et al. (2022) first reported the *EZR-ROS1* fusion identified in renal cell carcinoma by molecular sequencing. This new fusion was significant, as crizotinib may be effective in future treatment. The existing *ROS1* fusions can be detected in samples from patients, and the positive results may provide information for subsequent therapies, such as the use of kinase inhibitors.

TABLE 2 ROS1 fusion variants described in reported *ROS1* fusion patients.

Reported partner	Number	5′ gene exon	Number	3′ gene exon	Number
CD74-ROS1	35	1–6	16	34–43	15
				33–43	1
		1–4	15	34–43	12
				33–43	2
				32–43	1
		1–2	1	34–43	1
		1–7	1	34–43	1
		1–8	1	33–43	1
		7–8	1	34–43	1
EZR-ROS1	18	1–9	17	32–43	7
				33–43	7
				32–43	2
				31–43	1
		1–10	1	34–43	1
SDC4-ROS1	11	1–2	8	32–43	8
		1–4	2	32–43	2
		1–5	1	34–43	1
GOPC-ROS1	4	1–8	3	35–43	3
		1–4	1	36–43	1
SLC34A2-ROS1	3	1–12	1	1–33	1
		1–13	1	33–43	1
		1–4	1	31–43	1
TPM3-ROS1	3	1–10	2	31–43	1
				35–43	1
		1–7	1	35–43	1
CAPRIN1-ROS1	1	1–7	1	35–43	1
CCDC6-ROS1	1	1–5	1	35–43	1
LRIG3-ROS1	1	1–17	1	35–43	1
TPR-ROS1	1	1–4	1	35–43	1

Different *ROS1* fusions were observed in one case. Our results indicated that two patients with lung cancer harbored two *ROS1* fusions. One patient had *CD74-ROS1* and *SLC34A2-ROS1* fusions, and the other patient had novel *DCBLD1-ROS1* and *GOPC-ROS1* fusions. This finding has also been reported in previous studies. Using NGS, Xu et al. found two *ROS1* fusions [*SDC4-ROS1* (EX2:EX32) and *ROS1-GK* (EX31: EX13)] in a patient with lung adenocarcinoma. These results indicated that the patient might be sensitive to *ROS1* inhibitors (Xu et al., 2021). Patients harboring *ROS1* fusions may benefit from crizotinib if their tumors are metastatic (Cai et al., 2019).

With the large number of patients included in our study, we revealed that 16 patients harbored 14 novel *ROS1* fusion partners, all of whom had lung cancer. Of these, a novel *DCBLD1-ROS1* fusion was observed in three cases, while other fusions occurred once in six cases, and seven novel partners occurred in two retroperitoneal neoplasm patients (*FRK-ROS1*, *VGLL2-ROS1*), one gastric cancer

TABLE 3 ROS1 fusion variants described in novel *ROS1* fusion patients.

Novel partner	Number	5′ gene exon	Number	3′ gene exon	Number
DCBLD1-ROS1	3	1–14	1	35–43	1
		15–15	1	35–43	1
		1–9	1	35–43	1
ACVR2A-ROS1	1	6–12	1	41–43	1
ARHGEF11-ROS1	1	1–41	1	34–43	1
CARD11-ROS1	1	26–25	1	3–43	1
EPHA7-ROS1	1	6–17	1	36–53	1
FRK-ROS1	1	1–7	1	18–43	1
GRIK2-ROS1	1	1–10	1	34–43	1
LDLR-ROS1	1	1–14	1	34–43	1
MYH9-ROS1	1	1–40	1	35–43	1
NOTCH1-ROS1	1	1–30	1	34–43	1
NPM1-ROS1	1	1–4	1	35–43	1
REV3L-ROS1	1	23–33	1	24–43	1
VCL-ROS1	1	1–16	1	36–43	1
VGLL2-ROS1	1	4–4	1	36–43	1

patient (*ARHGEF11-ROS1*), one liver cancer patient (*REV3L-ROS1*), one liposarcoma patient (*EPHA7-ROS1*), one patient with epithelioid hemangioendothelioma (*NOTCH1-ROS1*), and one squamous cell carcinoma patient (*CARD11-ROS1*). These novel ROS*1* fusion partners should be further explored and may also be linked to promising therapies. For example, FPK is a Fyn-related kinase that functions as a tumor suppressor. It has been reported to inhibit glioma progression by suppressing ITGB1/FAK signaling. The effects of specific kinase inhibitors in *FRK-ROS1* fusion-positive cases should be explored further.

Variable genomic breakpoints have been identified in Chinese patients through GNS, most location was between exons 32 and 34, and exon 35was also a common. Most of canonical ROS1 fusions were sensitive to crizotinib, especially CD74-ROS1 fusion. Many novel uncommon ROS1 fusions have been found using NGS, most of which were reported to be sensitive to matched targeted therapy, similar to the canonical fusions (Hung et al., 2022). Clinical significance of some genomic breakpoints remained unclear. Simultaneously, more in-depth studies should be conducted to confirm and explore the mechanism underlying these fusions.

This study had numerous limitations. First, the response of the novel *ROS1* fusion types to the ROS1 inhibitors was not clear, and even the same *ROS1* fusion types with different fusion breakpoints of the partners or the *ROS1* gene should be collected, which will be beneficial for clinical therapy. In addition, the specific biological function of gene fusion has not been experimentally determined, and technical errors in NGS analyses cannot be completely excluded. Finally, although NGS is an established and powerful tool, there are barriers to its extensive application. It is expensive for patients and requires complicated equipment and skilled personnel to perform sequencing and subsequent bioinformatics analysis.

Conclusion

In summary, this study was performed to detect *ROS1* fusion partners and *ROS1* fusion breakpoints in solid tumors of Chinese patients in the largest *ROS1* fusion cohort to date. Fourteen novel *ROS1* fusions were identified in the solid tumors of Chinese patients, and the *ROS1* breakpoint was located between exons 32 and 34 in many patients. Moreover, this study showed that NGS fusion assays can be used on plasma and tissue samples. NGS is a potent tool for reliably identifying novel *ROS1* fusions and for detecting molecular alterations.

Data availability statement

The data presented in the study are deposited in the Genome Sequence Archive for human (GSA-human) repository, accession number HRA003505, under the BioProject PRJCA013094.

Ethics statement

The study was approved by the Institutional Review Boards of Shanghai East Hospital (Approved Number: 2020-093). Each patient included in this study has signed the informed consent form.

Author contributions

NL and MH contributed to the data collection. NL and MQ contributed to paper assessment. DZ, MH, and NL contributed to manuscript writing. FJ and MQ contributed to manuscript revising. All authors contributed to read and approved the final version.

Funding

This work was supported by grants 81972280 (to MQ) from National Natural Science Foundation of China, and by Shanghai "Rising Stars of Medical Talent" Youth Development Program (to MQ), and by the Top-level Clinical Discipline Project of Shanghai Pudong (PWYgf 2021-07).

Conflict of interest

Authors DZ and MH were employed by the company 3D Medicines Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.1035033/full#supplementary-material

References

Anjum, R., Zhang, S., Squillace, R., Clackson, T., Garner, A. P., and Rivera, V. M. (2012). 164 the dual ALK/EGFR inhibitor AP26113 also potently inhibits. *Eur. J. Cancer* 48, 50. doi:10.1016/s0959-8049(12)71962-7

Birch, A. H., Arcand, S. L., Oros, K. K., Rahimi, K., Watters, A. K., Provencher, D., et al. (2011). Chromosome 3 anomalies investigated by genome wide SNP analysis of benign, low malignant potential and low grade ovarian serous tumours. *PloS one* 6 (12), e28250. doi:10.1371/journal.pone.0028250

Birchmeier, C., Birnbaum, D., Waitches, G., Fasano, O., and Wigler, M. (1986). Characterization of an activated human ros gene. *Mol. Cell. Biol.* 6 (9), 3109–3116. doi:10.1128/mcb.6.9.3109

Birchmeier, C., Sharma, S., and Wigler, M. (1987). Expression and rearrangement of the ROS1 gene in human glioblastoma cells. *Proc. Natl. Acad. Sci. U. S. A.* 84 (24), 9270–9274. doi:10.1073/pnas.84.24.9270

Cai, C. L., Zhang, M. Q., Guo, J., Wang, L. W., Zhao, J. Q., Guo, W. J., et al. (2019). Response to crizotinib in a patient with metastatic lung spindle cell carcinoma harboring TPM3-ROS1 fusion. *Chin. Med. J.* 132 (24), 3003–3005. doi:10.1097/CM9.0000000000000556

Cai, W. J., and Su, B. (2013). A new target in non-small cell lung cancer: ROS1 fusion gene. *Zhonghua Zhong Liu Za Zhi* 35 (1), 1–4. doi:10.3760/cma.j.issn.0253-3766.2013.01.001

Camidge, D. R., Bang, Y. J., Kwak, E. L., Iafrate, A. J., Varella-Garcia, M., Fox, S. B., et al. (2012). Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: Updated results from a phase 1 study. *Lancet. Oncol.* 13 (10), 1011–1019. doi:10.1016/S1470-2045(12)70344-3

Cao, Z., Wu, W., Zhang, W., Li, Z., Gao, C., Huang, Y., et al. (2020). ALK and ROS1 rearrangement tested by ARMS-PCR in non-small-cell lung cancer patients via cytology specimens: The experience of Shanghai Pulmonary Hospital. *Diagn. Cytopathol.* 48 (6), 524–530. doi:10.1002/dc.24404

Charest, A., Lane, K., McMahon, K., Park, J., Preisinger, E., Conroy, H., et al. (2003). Fusion of FIG to the receptor tyrosine kinase ROS in a glioblastoma with an interstitial del(6)(q21q21). *Genes Chromosom. Cancer* 37 (1), 58–71. doi:10.1002/gcc.10207

Chiari, R., Ricciuti, B., Landi, L., Morelli, A. M., Delmonte, A., Spitaleri, G., et al. (2020). ROS1-rearranged non-small-cell lung cancer is associated with a high rate of venous thromboembolism: Analysis from a phase II, prospective, multicenter, two-arms trial (METROS). *Clin. Lung Cancer* 21 (1), 15–20. doi:10.1016/j.cllc.2019.06.012

Cui, M., Han, Y., Li, P., Zhang, J., Ou, Q., Tong, X., et al. (2020). Molecular and clinicopathological characteristics of ROS1-rearranged non-small-cell lung cancers identified by next-generation sequencing. *Mol. Oncol.* 14 (11), 2787–2795. doi:10.1002/1878-0261.12789

Davies, K. D., and Doebele, R. C. (2013). Molecular pathways: ROS1 fusion proteins in cancer. *Clin. Cancer Res.* 19 (15), 4040–4045. doi:10.1158/1078-0432.CCR-12-2851

Hu, H., Ding, N., Zhou, H., Wang, S., Tang, L., Xiao, Z., et al. (2021). Association of subjective cognitive decline with risk of cognitive impairment and dementia: A systematic review and meta-analysis of prospective longitudinal studies. *J. Prev. Alzheimers Dis.* 15 (1), 277–285. doi:10.14283/jpad.2021.27

Hung, M. S., Lin, Y. C., Chen, F. F., Jiang, Y. Y., Fang, Y. H., Lu, M. S., et al. (2022). The potential and limitation of targeted chromosomal breakpoint sequencing for the ROS1 fusion gene identification in lung cancer. *Am. J. Cancer Res.* 12 (5), 2376–2386. eCollection 2022.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., et al. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496. doi:10.1093/nar/gkh103

Matsushime, H., Wang, L. H., and Shibuya, M. (1986). Human c-ros-1 gene homologous to the v-ros sequence of UR2 sarcoma virus encodes for a transmembrane receptorlike molecule. *Mol. Cell. Biol.* 6 (8), 3000–3004. doi:10.1128/mcb.6.8.3000

Matsuura, S., Shinmura, K., Kamo, T., Igarashi, H., Maruyama, K., Tajima, M., et al. (2013). CD74-ROS1 fusion transcripts in resected non-small cell lung carcinoma. *Oncol. Rep.* 30 (4), 1675–1680. doi:10.3892/or.2013.2630

Melosky, B., Wheatley-Price, P., Juergens, R. A., Sacher, A., Leighl, N. B., Tsao, M. S., et al. (2021). The rapidly evolving landscape of novel targeted therapies in advanced non-small cell lung cancer. *Lung cancer (Amsterdam, Neth.* 160, 136–151. doi:10.1016/j.lungcan.2021.06.002

Sakai, K., Ohira, T., Matsubayashi, J., Yoneshige, A., Ito, A., Mitsudomi, T., et al. (2019). Performance of Oncomine Fusion Transcript kit for formalin-fixed,

paraffin-embedded lung cancer specimens. *Cancer Sci.* 110 (6), 2044–2049. doi:10.1111/cas.14016

Shaw, A. T., Yeap, B. Y., Solomon, B. J., Riely, G. J., Gainor, J., Engelman, J. A., et al. (2011). Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: A retrospective analysis. *Lancet. Oncol.* 12 (11), 1004–1012. doi:10.1016/S1470-2045(11)70232-7

Shen, L., Qiang, T., Li, Z., Ding, D., Yu, Y., and Lu, S. (2020). First-line crizotinib versus platinum-pemetrexed chemotherapy in patients with advanced ROS1-rearranged non-small-cell lung cancer. *Cancer Med.* 9 (10), 3310–3318. doi:10.1002/cam4.2972

Shu, Y., Wu, X., Tong, X., Wang, X., Chang, Z., Mao, Y., et al. (2017). Circulating tumor DNA mutation profiling by targeted next generation sequencing provides guidance for personalized treatments in multiple cancer types. *Sci. Rep.* 7 (1), 583. doi:10.1038/s41598-017-00520-1

Sievers, P., Stichel, D., Sill, M., Schrimpf, D., Sturm, D., Selt, F., et al. (2021). GOPC:ROS1 and other ROS1 fusions represent a rare but recurrent drug target in a variety of glioma types. *Acta Neuropathol.* 142 (6), 1065–1069. doi:10.1007/s00401-021-02369-1

Su, D., Zhang, D., Chen, K., Lu, J., Wu, J., Cao, X., et al. (2017). High performance of targeted next generation sequencing on variance detection in clinical tumor specimens in comparison with current conventional methods. *J. Exp. Clin. Cancer Res.* 36 (1), 121. doi:10.1186/s13046-017-0591-4

Suehara, Y., Arcila, M., Wang, L., Hasanovic, A., Ang, D., Ito, T., et al. (2012). Identification of KIF5B-RET and GOPC-ROS1 fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. *Clin. Cancer Res.* 18 (24), 6599–6608. doi:10.1158/1078-0432.CCR-12-0838

Tatjana, A., Tjota, M. Y., O'Donnell, P. H., Eggener, S. E., Agarwal, P. K., Haridas, R., et al. (2022). EZR-ROS1 fusion renal cell carcinoma mimicking urothelial carcinoma: Report of a previously undescribed gene fusion in renal cell carcinoma. *Virchows Arch.* 480 (2), 487–492. doi:10.1007/s00428-021-03138-x

Woo, H. Y., Na, K., Yoo, J., Chang, J. H., Park, Y. N., Shim, H. S., et al. (2020). Glioblastomas harboring gene fusions detected by next-generation sequencing. *Brain Tumor Pathol.* 37 (4), 136–144. doi:10.1007/s10014-020-00377-9

Xia, H., Xue, X., Ding, H., Ou, Q., Wu, X., Nagasaka, M., et al. (2020). Evidence of NTRK1 fusion as resistance mechanism to EGFR TKI in EGFR+ NSCLC: Results from a large-scale survey of NTRK1 fusions in Chinese patients with lung cancer. *Clin. Lung Cancer* 21 (3), 247–254. doi:10.1016/j.cllc.2019.09.004

Xu, L., Chen, X., Huo, H., Liu, Y., Yang, X., Gu, D., et al. (2021). Case report: Detection of double ROS1 translocations, SDC4-ROS1 and ROS1-GK, in a lung adenocarcinoma patient and response to crizotinib. *Front. Med.* 8, 649177. doi:10.3389/fmed.2021.649177

Zhang, L., Wang, Y., Zhao, C., Shi, J., Zhao, S., Liu, X., et al. (2019). High feasibility of cytological specimens for detection of ROS1 fusion by reverse transcriptase PCR in Chinese patients with advanced non-small-cell lung cancer. *Onco. Targets. Ther.* 12, 3305–3311. doi:10.2147/OTT.S198827

Zhu, Y. C., Lin, X. P., Li, X. F., Wu, L. X., Chen, H. F., Wang, W. X., et al. (2018). Concurrent ROS1 gene rearrangement and KRAS mutation in lung adenocarcinoma: A case report and literature review. *Thorac. Cancer* 9 (1), 159–163. doi:10.1111/1759-7714.12518

Zhu, Y. C., Zhang, X. G., Lin, X. P., Wang, W. X., Li, X. F., Wu, L. X., et al. (2019). Clinicopathological features and clinical efficacy of crizotinib in Chinese patients with ROS1-positive non-small cell lung cancer. *Oncol. Lett.* 17 (3), 3466–3474. doi:10.3892/ol.2019.9949

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fgene.2022.1095428

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Faming Zhao,
Huazhong University of Science and Technology, China
Dong Zhang,
Qilu Hospital, Shandong University, China

*CORRESPONDENCE
Feng Cao,
fengcao8828@163.com

SPECIALTY SECTION
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

RECEIVED 11 November 2022
ACCEPTED 27 December 2022
PUBLISHED 10 January 2023

CITATION
Qin C, Wang Y, Zhang Y, Zhu Y, Wang Y and Cao F (2023), Transcriptome-wide analysis reveals the molecular mechanisms of cannabinoid type II receptor agonists in cardiac injury induced by chronic psychological stress.
Front. Genet. 13:1095428.
doi: 10.3389/fgene.2022.1095428

Transcriptome-wide analysis reveals the molecular mechanisms of cannabinoid type II receptor agonists in cardiac injury induced by chronic psychological stress

Cheng Qin[1], Yujia Wang[1], Yang Zhang[1], Yan Zhu[2], Yabin Wang[1] and Feng Cao[1,3]*

[1]Department of Cardiology, National Clinical Research Center for Geriatric Diseases and Second Medical Center of Chinese PLA General Hospital, Beijing, China, [2]Nankai University School of Medicine, Nankai University, Tianjin, China, [3]Beijing Key Laboratory of Research on Aging and Related Diseases, Beijing, China

Background: Growing evidence has supported that chronic psychological stress would cause heart damage, However the mechanisms involved are not clear and effective interventions are insufficient. Cannabinoid type 2 receptor (CB2R) can be a potential treatment for cardiac injury. This study is aimed to investigate the protective mechanism of CB2R agonist against chronic psychological stress-induced cardiac injury.

Methods: A mouse chronic psychological stress model was constructed based on a chronic unpredictable stress pattern. Mice were performed a three-week psychological stress procedure, and cardiac tissues of them were collected for whole-transcriptome sequencing. Overlap analysis was performed on differentially expressed mRNAs (DE-mRNAs) and ER stress-related genes (ERSRGs), and bioinformatic methods were used to predict the ceRNA networks and conduct pathway analysis. The expressions of the DE-ERSRGs were validated by RT-qPCR.

Results: In the comparison of DE mRNA in Case group, Control group and Treatment group, three groups of ceRNA networks and ceRNA (circ) networks were constructed. The DE-mRNAs were mainly enriched in chromatid-relevant terms and Hematopoietic cell lineage pathway. Additionally, 13 DE-ERSRGs were obtained by the overlap analysis, which were utilized to establish a ceRNA network with 15 nodes and 14 edges and a ceRNA (circ) network with 23 nodes and 28 edges. Furthermore, four DE-ERSRGs (Cdkn1a, Atf3, Fkbp5, Gabarapl1) in the networks were key, which were mainly enriched in response to extracellular stimulus, response to nutrient levels, cellular response to external stimulus, and FoxO signaling pathway. Finally, the RT-qPCR results showed almost consistent expression patterns of 13 DE-ERSRGs between the transcriptome and tissue samples.

Conclusion: The findings of this study provide novel insights into the molecular mechanisms of chronic psychological stress-induced cardiac diseases and reveal novel targets for the cardioprotective effects of CB2R agonists.

KEYWORDS

chronic psychological stress, endoplasmic reticulum stress, cardiac injuries, CB2R, ceRNA network, functional enrichment analysis

1 Introduction

Psychological stress is a maladaptation in response to exposure to an environmental threat or severe external stimuli. Traumatic events, such as attacks, natural disasters, or military combat, can result in psychological stress, which is detrimental to the health of an individual (Glaser and Kiecolt-Glaser, 2005). A potential link might exist between psychological stress and cardiovascular diseases (CVD), a topic that has received increasing attention from scholars (Gold et al., 2020). Findings from multiple clinical studies have shown that psychological stress can increase the risk of adverse cardiovascular events, such as myocardial ischemia (Ebrahimi et al., 2021; Vaccarino et al., 2021; Shah et al., 2022), sudden death (Meisel et al., 1991; Leor et al., 1996), arrhythmia (Steinberg et al., 2004; Hourani et al., 2020), stroke (Jordan et al., 2013; Chen et al., 2015), and the aggravation of heart failure (Roy et al., 2015).

However, despite the epidemiological evidence, the pathophysiological mechanisms underlying these associations remain incomplete. Findings from recent studies suggest that autonomic dysfunction and inflammatory response promote plaque vulnerability, which is the cause of cardiovascular events induced by acute stress (Hinterdobler et al., 2021). However, the long-term effects of chronic stress on the cardiovascular system are not completely understood. Therefore, specific and effective interventions are urgently needed. Moreover, whether healthy individuals exposed to stress elicit similar pathophysiological responses as individuals with pre-existing cardiovascular diseases remains doubtful.

The endoplasmic reticulum (ER) is the primary site for protein synthesis, folding, and transport (Wang and Kaufman, 2016). The exposure of cells to oxidative stress, hypoxia, nutrient deprivation, and pathological infection will cause the accumulation of misfolded or unfolded proteins in the ER and initiate the unfolded protein response (UPR) (Ochoa et al., 2018). If stimuli persist, the UPR will lead to apoptosis and autophagy. ATF6, PERK, and IRE1 are the representative molecules of UPR signaling (Ron and Walter, 2007). Under usual conditions, these molecules bind to their partner molecule glucose-regulated protein 78 (GRP78). Under ER stress, unfolded or misfolded proteins recruit GRP78/BiP to dissociate from the complex composed of IRE1, ATF6, and PERK. This process marks the initiation of ER stress (Bertolotti et al., 2000; Frakes and Dillin, 2017). The expression of the abovementioned ER stress-related genes (ESRGs) was found to be related to the occurrence and development of various cardiovascular diseases, including myocardial ischemia (Bi et al., 2018; Zhang et al., 2019), ventricular remodeling (Blackwood et al., 2019), atherosclerosis (Tang et al., 2019), and vascular calcification (Fu et al., 2019). Likewise, the expression of ESRGs was found to be upregulated in the brain tissues of mice subjected to psychological stress (Gold et al., 2013; Mondal et al., 2015; Mao et al., 2019). Thus far, small molecule drugs targeting ESRGs have been developed, but they have not been adopted widely in clinical practice owing to their relative toxicity and lack of specificity (Atkins et al., 2013; Almanza et al., 2019).

Cannabinoid receptor type 2 (CB2R) is primarily localized in immune cells as well as in endothelial cells and cardiac tissues. It plays an important role in the regulation of the immune response and inflammation. CB2R agonists can exhibit a cardioprotective role in myocardial infarction, cardiomyopathy, arrhythmia, and other CVDs by reducing reactive oxygen species (ROS) production, inhibiting necroptosis and apoptosis, promoting efferocytosis, upregulating endothelial NO synthase, and reducing macrophage infiltration (Molica et al., 2012; Li et al., 2016; Yu et al., 2019; Liu X. et al., 2020). Owing to the anti-oxidative stress effect of CB2R agonists and the regulation of autophagy, CB2R agonists may affect the cellular process underlying ER stress.

Here, we aimed to elucidate the mechanism underlying cardiac injury induced by mental stress and the therapeutic targets of CB2R agonists. We constructed a mouse model induced by psychological stress (case group), stressed mice treated with CB2R agonists (treatment group), and normal mice (control group) and collected samples for whole transcriptome sequencing analysis to analyze the differences in gene expression profiles between the different groups.

2 Methods

2.1 Animal models

All mice were obtained from SPF Biotechnology and allowed 1 week of acclimation to the specific-pathogen-free animal facilities before the experiments were conducted. Six-week-old male mice were used in this study and were randomly divided into three groups: control group (three normal mice were maintained with free access to food and water on a 12 h light/dark cycle), case group (four mice with psychological stress-induced heart disease), and treatment group (four mice with psychological stress-induced heart disease treated with the CB2R agonist JWH133). JWH133 (ApexBio, B7941) was injected twice daily intraperitoneally at 10 mg/kg from the beginning of the stress induction procedure for 21 days. The stress induction process is shown inSupplementary Table S1. Behavioral tests were performed 24 h after stress induction.

2.2 Data sources

After stress induction, the hearts of the mice were removed, and the tissue was processed for whole-transcriptome sequencing analysis. After RNA extraction, purification, and library construction, these libraries were double-ended (paired-end, PE) sequenced by next-generation sequencing (NGS) using the Illumina HiSeq sequencing platform. The sequencing depth for the transcriptome-wide analysis was showed in Supplementary Figure S1.

In addition, 787 ERS-related genes with a relevance score of ≥7 were extracted from a report by Zhang et al. (2021). Since these genes were from a human gene set, the biomaRt package in R software was used to identify the homologous genes of mice. We identified 828 ERS-related genes in mice (Supplementary Table S2). GSE210252 was downloaded from the Gene Expression Omnibus (GEO) dataset (https://www.ncbi.nlm.nih.gov/), which contained four pairs samples from mice received restraint stress or no stress.

2.3 Behavioral test

2.3.1 Open-field test

The open-field test can be used to observe voluntary movement and assess the severity of stress. The open-field system comprised an open-field box (40 × 40 × 40 cm^3), an imaging system, and analysis software. During the experiment, the mice were placed in a specific

position in the open-field, the camera system was used to monitor the activities of the animals in the open-field, and an analysis software was used to track the trajectory of the animals and collect data on their length of stay in the central area. The lesser time the mice remained in the central area, the more severe stress they experienced (Yan et al., 2022).

2.3.2 Elevated plus maze (EPM)

EPM is an experimental tool used to evaluate the stress response of rodents. In this study, EPM was a cross-shaped platform composed of two open arms and two closed arms (arms 35 cm long × 5 cm wide) positioned approximately 55 cm above the ground. A monitoring camera for animal behavior experiments was installed above the EPM. The residence time in the open arm (open arm time, OT) was recorded using the Labmaze 3.0 animal behavior analysis software (Zhongshidichuang Sci. and Tech). For the EPM test, animals were placed in the center of the maze facing an open arm (Cho et al., 2022).

2.3.3 Forced swimming test

The forced swimming test was conducted in a cylindrical water tank (50 cm in height and 10 cm in diameter). The water temperature was 23 ~ 25 °C, and the water had a depth that was sufficient to position the tail of the animal at a certain distance from the bottom. The swimming time was 6 min, and the immobility time in the last 4 min of the swimming duration was recorded. Moreover, the standard of immobility was that the animal stopped struggling in the water and started floating, with only slight limb movement to keep the head afloat (Gu et al., 2021). A longer immobility time represented more severe stress.

2.3.4 Tail suspension test

This experiment was conducted in a dark and quiet environment. The tail of the mouse was fixed and hung in a 25 cm × 25 cm × 30 cm box, and the head of the mouse was positioned approximately 5 cm from the bottom. Furthermore, the activity of the mouse was recorded within 6 min. The immobility time in the last 4 min was recorded. Small movements of the forelimbs rather than the hindlimbs were considered to represent immobility (Gu et al., 2021).

2.4 Differential analysis

The Limma package (version 3.44.3) was used to screen differentially expressed mRNAs (DE-mRNAs), miRNAs (DE-miRNAs), lncRNAs (DE-lncRNAs), and circRNAs (DE-circRNAs) in the three comparison groups (Control vs. Case; Case vs. Treatment; Control vs. Treatment). In detail, the count values obtained by comparing the mRNA and lncRNA data were converted to the fragments per kilobase per million mapped reads (FPKM) values. The count values used by miRNA and circRNA were converted to the counts per million mapped (CPM) value, and the selection standard was set at $p < .05$ and $|Log_2FC|>1$. Moreover, the distribution of DE-mRNAs, DE-miRNAs, DE-circRNAs, and DE-lncRNAs in the three comparison groups was demonstrated at the chromosomal level using Circos in R package.

2.5 Construction of ceRNA networks

To predict target miRNAs that could potentially bind to the DE-mRNAs in the three comparison groups, mirwalk 3.0 was used with a threshold score >.95. Simultaneously, only the DE-miRNAs with opposite expression trends to DE-mRNAs were retained. After the target miRNA was obtained, starBase 2.0 was used to predict the miRNA-lncRNA pairs, and only the relationship pairs with opposite differential expression trends were retained. Thus, an mRNA-miRNA-lncRNA binding network was obtained from each comparison group, which was considered the ceRNA network.

Similarly, in each comparison group, mirwalk was used to predict the binding of DE-mRNA and DE-miRNA with a threshold score >.95, and only the pairs with contrasting trends of differential expression were retained. To predict the circRNAs that might be related to the miRNAs after obtaining miRNAs, the circRNA sequences were extracted from assembled gene files, miRNA sequences of mice were retrieved from miRbase, and the miRanda software was used to predict the binding with the selection criteria as binding score = 140, and only the opposite relationship pairs of differential expression trends were retained. Thus, an mRNA-miRNA-circRNA network was constructed for each group, known as the ceRNA (circ) network.

2.6 Screening and function annotation of differentially expressed ER stress-related genes (DE-ERSRGs)

Overlap analysis was performed on the DE-mRNAs from the Case vs. Control and Treatment vs. Case comparison groups and the 828 mouse ER stress-related genes. The intersected genes were considered to be DE-ERSRGs.

2.7 Construction of the DE-ERSRGs-based ceRNA network

Initially, mirwalk was used to predict the target miRNAs that can potentially bind to the DE-ERSRGs with a threshold score >.95. Simultaneously, only miRNAs in the pairs with contrasting trends of differential expression were retained in the Treatment vs. Case group. After obtaining the target miRNA, starBase was applied to predict the miRNA-lncRNA pairs, and the relationship pairs with opposite differential expression trends were retained as well. Thus, an mRNA-miRNA-lncRNA binding network was obtained for the Treatment vs. Case group.

Likewise, mirwalk was used to predict the target miRNAs of DE-ERSRG, and only miRNAs in the miRNAs-DE-ERSRG pairs with opposite trends of differential expressions were retained. After the miRNAs were obtained, the target circRNAs of the miRNAs were predicted using miRanda, and the pairs with the contrasting relationship of differential expression trends were selected. An mRNA-miRNA-circRNA network was established using the data.

2.8 Functional enrichment analyses of DE-ERSRGs in the ceRNA network

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed with DE-mRNAs, mRNAs in the ceRNA and ceRNA (circ) networks as well as DE-ERSRGs from the three comparison groups individually using clusterProfiler (version4.0.2), with $p < .05$ as the selection standard.

2.9 Real-time qPCR

Eleven frozen samples were segregated into three groups, with three samples each in the control and case groups and five samples in the treatment group. Then, 50 mg of tissue was extracted from each sample and lysed using TRIzol Reagent (Life Technologies, CA, United States), and total RNA was isolated by following the manufacturer's instructions. After the concentration and purity of RNA were determined, the RNA was reverse-transcribed to cDNA using the SureScript-First-strand-cDNA-synthesis-kit (Genecopoeia, Guangzhou, China) before qRT-PCR. The qRT-PCR mixture was composed of 3 μL of cDNA, 5 μL of 2x Universal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China), and 1 μL each of forward and reverse primers. PCR was performed in a BIO-RAD CFX96 Touch TM PCR detection system (Bio-Rad Laboratories, Inc., United States) under the following thermal cycling conditions: 40 cycles at 95°C for 60 s, 95°C for 20 s, 55°C for 20 s, and 72°C for 30 s. The $2^{-\triangle\triangle Ct}$ method was used to calculate gene expression levels, and GraphPad Prism 5 was applied to plot and calculate the statistical significance. The primer sequences used in this study are listed in Supplementary Table S3.

3 Results

3.1 Successful establishment of the chronic psychological stress model

Behavioral tests were performed 24 h after the end of stress induction in the three groups of mice. As shown in Supplementary Figure S2A, in the open-field experiment, mice from the case group spent significantly less time in the central zone than mice from the control and CB2R agonist treatment groups. Additionally, in the elevated plus maze test, the time of stay in the open arm in the case group was significantly lesser than that in the control group, but no difference was observed between the case and treatment groups (Supplementary Figure S2B). Furthermore, in the forced swimming test, the immobility time in the case group was significantly greater than that in the control and treatment groups (Supplementary Figure S2C). Finally, in the tail suspension test, the immobility time in the case group was significantly greater than that in the control group, but no statistically significant differences were observed compared with the treatment group (Supplementary Figure S2D). The above results indicated that the protocol could induce significant stress in the mice.

3.2 Identification of DE-mRNAs, DE-miRNAs, DE-lncRNAs, and DE-circRNAs

In the Control vs. Case group, 701 DE-mRNAs (357 upregulated, 344 downregulated), 80 DE-miRNAs (19 upregulated, 61 downregulated), 530 DE-lncRNAs (268 upregulated, 262 downregulated), and 95 DE-circRNAs (26 upregulated, 69 downregulated) were screened between four case mice and three normal case mice (Figure 1A). In the Treatment vs. Case group, 466 DE-mRNAs (230 upregulated, 236 downregulated), 18 DE-miRNAs (eight upregulated, ten downregulated), 500 DE-lncRNAs (290 upregulated, 210 downregulated), and 83 DE-circRNAs (45 upregulated, 38 downregulated) were obtained (Figure 1B). In the Control vs. Treatment group, 475 DE-mRNAs (231 upregulated, 245 downregulated), 24 DE-miRNAs (four upregulated, 20 downregulated), 349 DE-lncRNAs (196 upregulated, 153 downregulated), and 67 DE-circRNAs (27 upregulated, 40 downregulated) were eventually selected (Figure 1C).

In the Control vs. Case group, the majority of DE-mRNAs were located in the chromosomes, except the Y chromosome (only one downregulated DE-mRNA was located in Y chromosome). The DE-lncRNAs were almost evenly situated in each pair of chromosomes, and DE-miRNAs were situated in chromosomes 2, 3, 4, 7, 11, 12, 15, 17, 19, and X. DE-circRNAs were located in the euchromosomes and chromosome X (Supplementary Figure S3A).

In the Treatment vs. Case group, the DE-mRNAs, DE-lncRNAs, and DE-circRNAs were situated in all chromosomes except the Y chromosome. The DE-miRNAs were located in the 3, 5, 6, 7, 8, 11, 12, 13, 15, and X chromosomes (Supplementary Figure S3B).

In the Treatment vs. Control group, the locations of DE-mRNAs and DE-lncRNAs were similar to those in the Treatment vs. Case group, but the DE-circRNAs were located in the euchromosomes and Y chromosomes, and DE-miRNAs were located in the 1, 3, 5, 6, 7, 8, 10, 11, 12, 14, 15, 19, and X chromosomes (Supplementary Figure S3C).

3.3 Functional enrichment analyses of DE-mRNAs in the three groups

The functional enrichment results of DE-mRNAs in the Control vs. Case group illustrated that 1261 GO terms and 98 KEGG pathways were enriched by 701 DE-mRNAs. For example, chromosome separation, nuclear division, nuclear chromosome segregation, organelle fission, and mitotic nuclear segregation were the major enriched GO BP terms. Chromosomes, centromeric region, spindle, and chromosomal region, among others, were the primarily enriched terms in CC. In MF, microtubule binding, tubulin binding, metal ion transmembrane transporter activity, microtubule motor activity, among others, were the major common terms. Furthermore, systemic lupus erythematosus, neutrophil extracellular trap formation, and alcoholism were the primarily enriched KEGG pathways (Figure 2A).

With respect to DE-mRNAs in the Treatment vs. Case group, 928 GO terms and 33 KEGG pathways were enriched by the 466 DE-mRNAs. The common enriched GO terms primarily included regulation of metal ion transport, anatomical structure homeostasis, negative regulation of

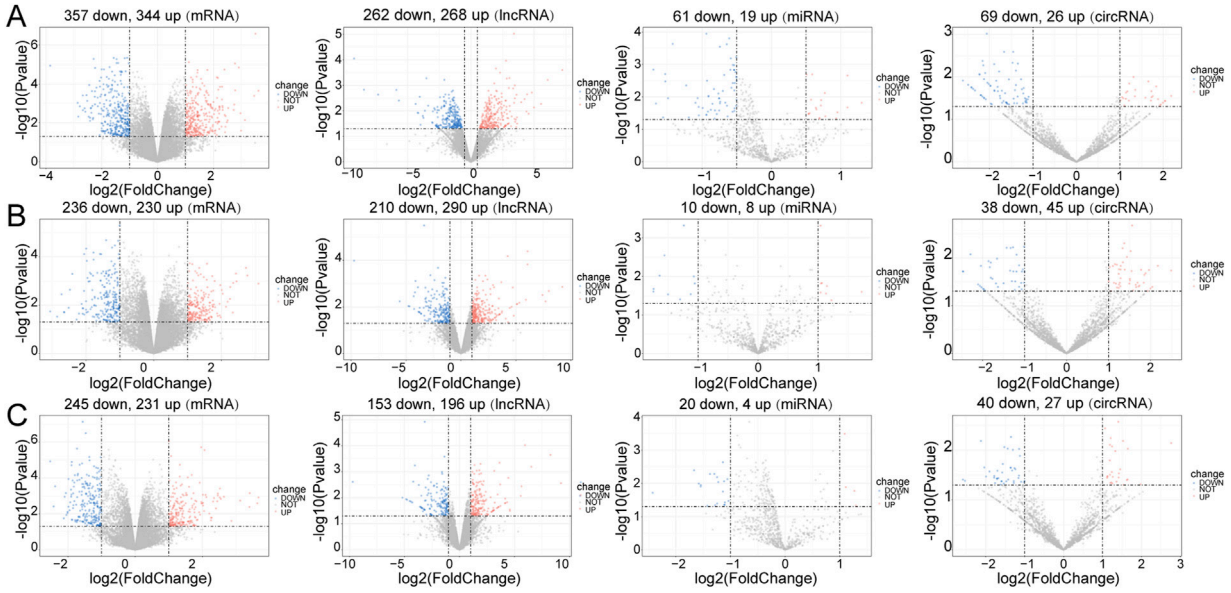


FIGURE 1
Differential expression analysis. Volcano plot for differentially expressed mRNAs (DE-mRNAs), lncRNAs (DE-lncRNAs), miRNAs (DE-miRNAs) and circRNAs (DE-circRNAs) in the three comparison groups. **(A)** Control vs. Case; **(B)** Case vs. Treatment; **(C)** Control vs. Treatment. Blue: downregulation, red: upregulation.

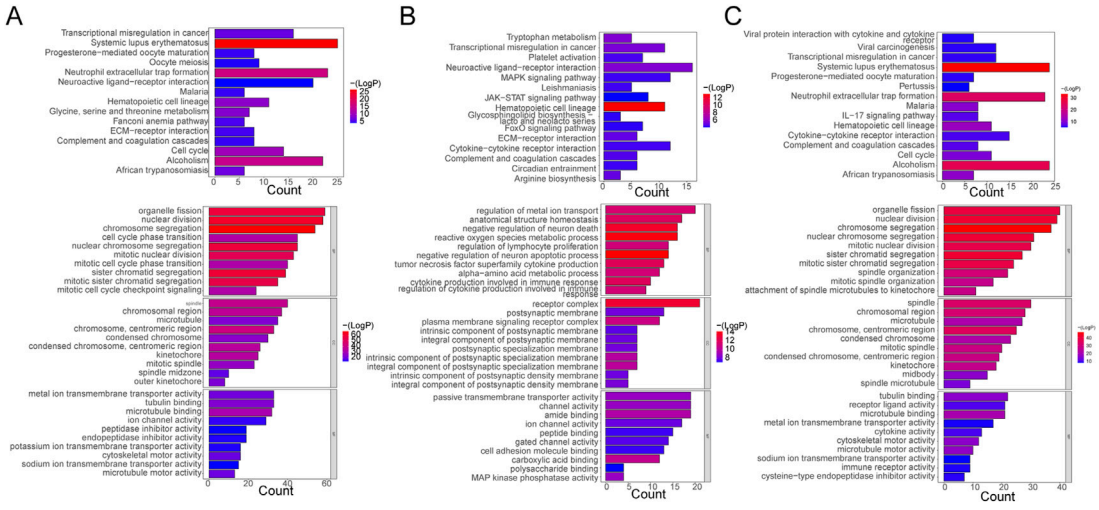


FIGURE 2
Functional Enrichment Analysis. **(A)** Bar plots of KEGG pathway and GO enrichment analyses in the control and case groups. **(B)** Bar plots of KEGG pathway and GO enrichment analyses in the case and treatment groups. **(C)** Bar plots of KEGG pathway and GO enrichment analyses in the control and case groups. Each GO map presents the top 10 enriched terms in BP, CC, and MF. Each KEGG map shows the top 15 enriched KEGG pathways.

neuron death, and others in BP; receptor complex, plasma membrane signaling receptor complex, integral component, and others in CC; and carboxylic acid binding, amide binding, kinase phosphatase activity, and others in MF. Additionally, hematopoietic cell lineage, neuroactive ligand-receptor interaction, transcriptional dysregulation in cancer, tryptophan metabolism, and others were the primarily enriched pathways (Figure 2B).

Furthermore, in the Treatment vs. Control group, the 476 DE-mRNAs enriched 952 GO terms. The major BP terms were chromosome segregation, nuclear division, sister chromatid segregation, nuclear chromosome segregation, and others. The major CC terms were chromosome, centromeric region, microtubule, chromosomal region, spindle, and others. The major MF terms were microtubule binding, receptor-ligand activity, tubulin binding, and others. Moreover, 31 KEGG pathways were enriched, and the top fifteen most significant pathways were shown in the chart, which primarily included alcoholism, neutrophil extracellular trap formation, and systemic lupus erythematosus, among others (Figure 2C).

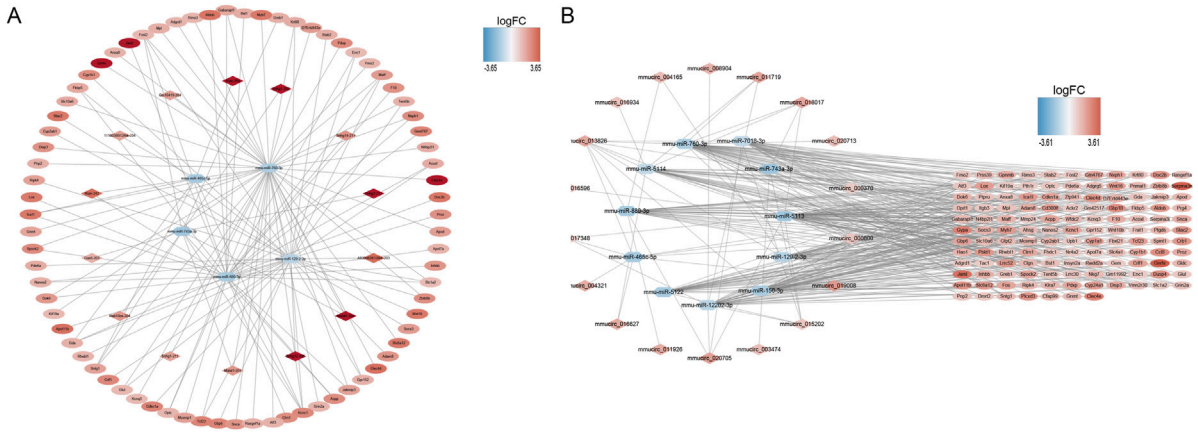


FIGURE 3
ceRNA network construction. **(A)** mRNA-miRNA-lncRNA network in control and case groups. **(B)** mRNA-miRNA-circRNA network in control and case groups. In the mRNA-miRNA-lncRNA (circRNA) network, the hexagon represents miRNA, the rhombus represents lncRNA (circRNA), the oval represents mRNA, and the color represents logFC value. The redder the color, the more upregulated the expression of the gene, and the bluer the color, the more downregulated the expression of the gene.

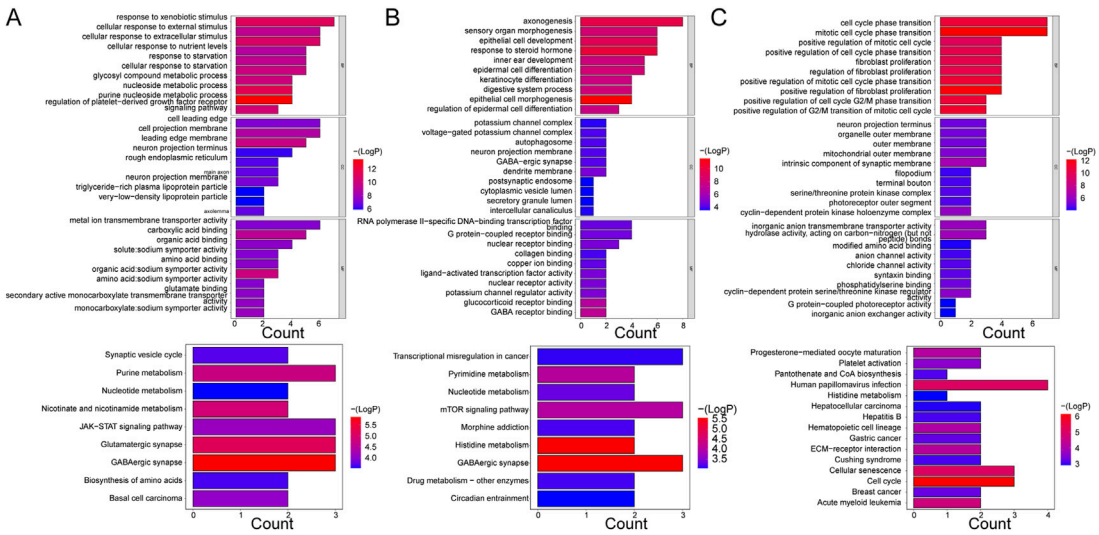


FIGURE 4
GO enrichment and KEGG pathway analyses of mRNAs enrolled in ceRNA networks. **(A)** KEGG pathway and GO enrichment analysis of mRNAs in the ceRNA (lncRNA) network in control and case groups. **(B)** KEGG pathway and GO enrichment analysis of mRNAs in the ceRNA (lncRNA) network in the case and treatment groups. **(C)** KEGG pathway and GO enrichment analysis of mRNAs in the ceRNA (lncRNA) network in the control and treatment groups. Each GO map shows the top 10 enriched phrases in BP, CC, and MF. On **(A, B)**, the top 9 enriched KEGG pathways are shown, while **(C)** shows the top 15 enriched KEGG pathways.

3.4 ceRNA and ceRNA (circ) networks of the three groups

A ceRNA network containing 99 nodes and 127 edges and a ceRNA (circ) network containing 157 nodes and 354 edges were established using DE-mRNAs in the Case vs. Control group (Figure 3A, B). Based on the mRNAs in the Treatment vs. Case group, a ceRNA network with 104 nodes and 112 edges and a ceRNA (circ) network with 293 nodes and 492 edges were established (Supplementary Figures S4A, B). In the Treatment vs. Control group, a ceRNA network with 68 nodes and 66 edges and a ceRNA (circ) network with 226 nodes and 424 edges were established (Supplementary Figures S4C, D).

3.5 Functional enrichment analyses of mRNAs in the networks

Five hundred and sixty-seven GO terms and nine KEGG pathways were enriched by the mRNAs in the ceRNA network of the Case vs. Control group. For instance, the purine nucleoside metabolic process,

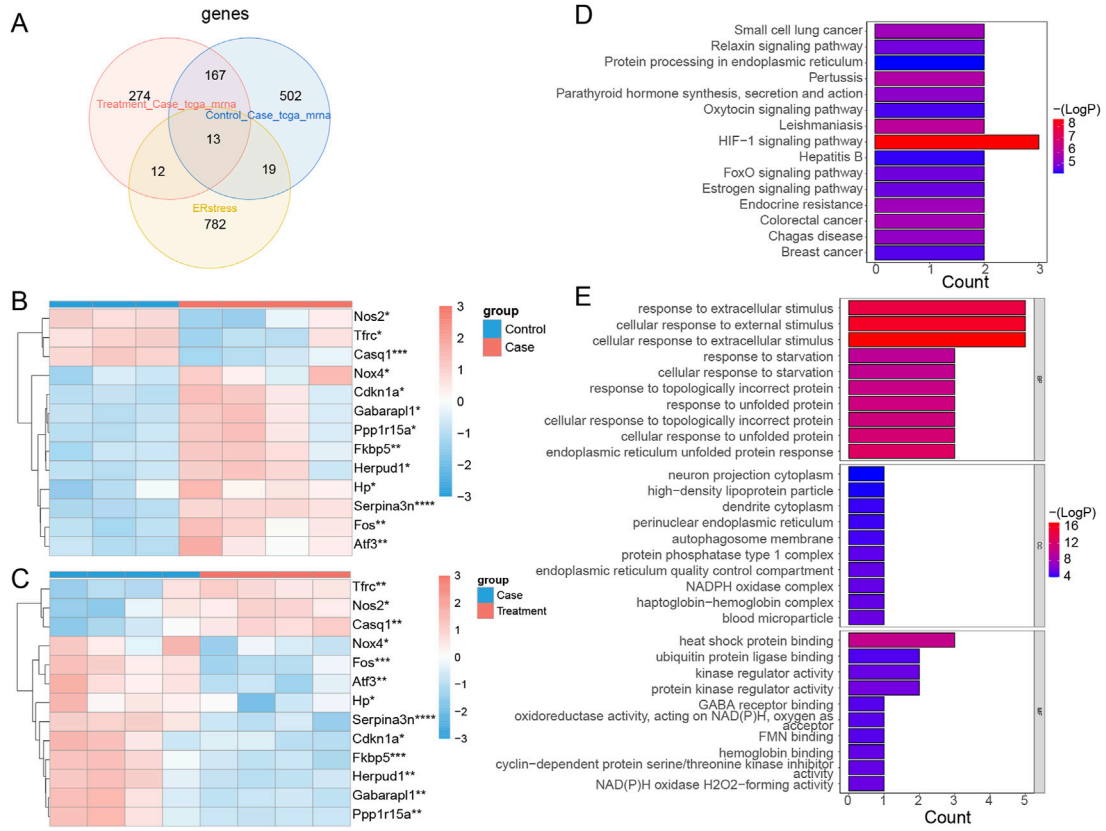


FIGURE 5
Identification of DE-ERSRGs and functional enrichement analysis. **(A)** Venn diagram of 13 DE-ERSRGs by overlapping DEGs in control vs. case group, DEGs in case group vs. treatment, with 826 ER stress-related genes. **(B)** Heat map of expression of 13 DE-ERSRGs in control and case groups. **(C)** Heat map of expression of 13 DE-ERSRGs in the case and treatment groups. **(D)** KEGG pathway analysis of 13 DE-ERSRGs. **(E)** GO enrichment analysis of 13 DE-ERSRGs.

nucleoside metabolic process, response to xenobiotic stimulus, leading-edge membrane, cell projection membrane, cell leading edge, carboxylic acid binding, metal ion transmembrane transporter activity, and organic acid:sodium symporter activity were the primary enriched terms. The nine KEGG pathways were basal cell carcinoma, biosynthesis of amino acids, GABAergic synapse, glutamatergic synapse, JAK-STAT signaling pathway, nicotinate and nicotinamide metabolism, nucleotide metabolism, purine metabolism, and synaptic vesicle cycle (Figure 4A). Additionally, the mRNAs in the ceRNA (circ) network of the Case vs. Control group enriched 694 GO terms and 17 KEGG pathways, such as response to xenobiotic stimulus, small molecule catabolic process, cell projection membrane, and carboxylic acid binding in GO. Parathyroid hormone synthesis, secretion, and action, glyoxylate and dicarboxylate metabolism, breast cancer, and basal cell carcinoma were the major KEGG pathways (Supplementary Figure S5A).

With respect to the Case vs. Treatment group, the mRNAs in ceRNA network enriched axonogenesis, response to steroid hormone, epithelial cell development, dendrite membrane, G protein-coupled receptor binding, RNA polymerase II-specific DNA-binding transcription factor binding, and others among 426 GO terms. The nine enriched KEGG pathways primarily included GABAergic synapse, histidine metabolism, and mTOR signaling pathway (Figure 4B). In addition, the mRNAs in the ceRNA (circ) network commonly enriched 684 GO terms and 23 KEGG pathways, such as regulation of metal ion transport, glutamatergic synapse, carbohydrate binding, transcriptional dysregulation in cancer, and FoxO signaling pathway (Supplementary Figure S5B).

Finally, in the Treatment vs. Control group, the mRNAs from the ceRNA network enriched 314 GO terms and 14 KEGG pathways. The enriched BP terms were more significant than the CC and MF terms, such as mitotic cell cycle phase transition, cell cycle phase transition, and positive regulation of fibroblast proliferation. The most significantly enriched pathways included cell cycle, cellular senescence, and human papillomavirus infection, among others (Figure 4C). Besides, the enrichment results of mRNAs in the ceRNA (circ) network were similar in that the enriched BP terms were more significant than those in CC and MF as well, such as regulation of body fluid levels, G2/M transition of mitotic cell cycle, and cell cycle G2/M phase transition. Hematopoietic cell lineage, cytokine-cytokine receptor interaction, *Staphylococcus aureus* infection, and others were some of the primarily enriched KEGG pathways (Supplementary Figure S5C).

3.6 Identification and functional annotation of DE-ERSRGs

Thirteen intersected mRNAs (Fos, Hp, Serpina3n, Cdkn1a, Atf3, Nos2, Fkbp5, Tfrc, Nox4, Gabarapl1, Herpud1, Casq1, and Ppp1r15a)

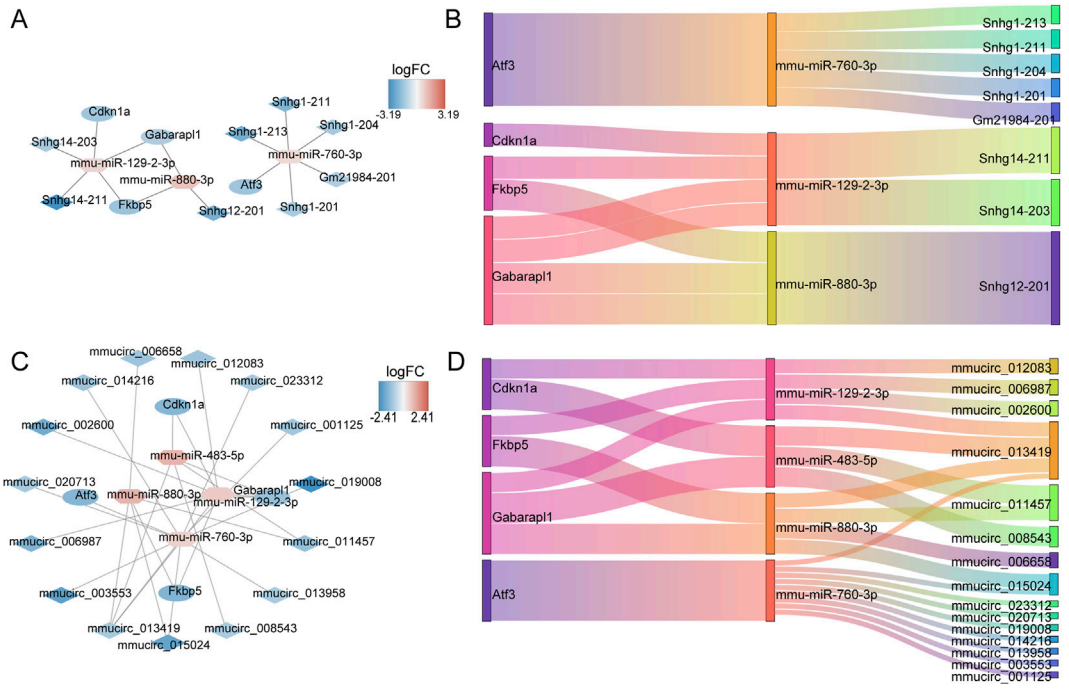


FIGURE 6
Construction of ceRNA networks for 13 DE-ERSRGs. **(A)** mRNA-miRNA-lncRNA regulatory network. **(B)** Sankey diagram of the mRNA-miRNA-lncRNA regulatory. **(C)** mRNA-miRNA-circRNA regulatory network. **(D)** Sankey diagram of mRNA-miRNA-circRNA regulatory. In the mRNA-miRNA-lncRNA (circRNA) network, the hexagon represents miRNA, the rhombus represents lncRNA (circRNA), the oval represents mRNA, and the color represents logFC value. The closer the color is to red, the more upregulated the expression is, and the closer the color is to blue, the downregulated the expression is.

were identified from the overlap analysis; these were considered DE-ERSRGs (Supplementary Table S4; Figure 5A). The expression levels of DE-ERSRGs in the Case group were significantly different from those in the Control groups (Supplementary Figure S6A). Of note, the expression of Nos2, Tfrc, and Casq1 was higher in the control group (Figure 5B). Furthermore, these DE-ERSRGs were expressed differently between the Treatment and Case groups as well (Supplementary Figure S6B), and the expression levels of Tfrc, Nos2, and Casq1 were significantly higher in the treatment group (Figure 5C).

Moreover, the enrichment results indicated that 24 KEGG and 429 GO term pathways were enriched by the 13 DE-ERSRGs. HIF-1 signaling was the most significantly enriched KEGG pathway, followed by Leishmaniasis, Pertussis, and others (Figure 5D). Furthermore, cellular response to extracellular stimulus, cellular response to external stimulus, response to extracellular stimulus, and other terms, were the primary BP terms. Meanwhile, the CC terms included dendrite cytoplasm, high-density lipoprotein particle, neuron projection cytoplasm, and others. Heat shock protein binding, ubiquitin protein ligase binding, kinase regulator activity, and others were the primary terms in MF (Figure 5E).

3.7 Establishment of DE-ERSRG-based ceRNA and ceRNA (circ) networks

The ceRNA network of 13 DE-ERSRGs illustrated that 15 nodes and 14 edges were present in the network, which only contained four of the 13 DE-ERSRGs (Cdkn1a, Atf3, Fkbp5, and Gabarapl1) (Figures 6A, B). In addition, a ceRNA (circ) network with 23 nodes and 28 edges was established based on the 13 DE-ERSRGs, which only contained the same four DE-ERSRGs (Figures 6C, D). Therefore, these four genes were regarded as key DE-ERSRGs. Similarly, according to the online dataset from GEO database (GSE210252), the expression patterns of four key DE-ERSRGs between mice received restraint stress or no stress were further explored and exhibited in Supplementary Figure S7, supporting that Cdkn1a, Atf3, and Gabarapl1 expressed higher in case group than that in control group, which were consistent to the results of our study above that shown in Supplementary Figure S6A.

3.8 RT-qPCR validation of 13 DE-ERSRGs

To gain deeper insights on the expression levels of the 13 DE-ERSRGs (Fos, Hp, Serpina3n, Cdkn1a, Atf3, Nos2, Fkbp5, Tfrc, Nox4, Gabarapl1, Herpud1, Casq1, and Ppp1r15a), RT-qPCR was performed using eleven tissue samples. The expression levels of the DE-ERSRGs were distinctly different between the control and case samples (all $p < .05$), which was consistent with the results of previous bioinformatics analysis. Conversely, the expression levels in the case and treatment groups were significantly different, except for the levels of Ppp1r15a and Nos2. Although the expression levels of these two genes were not statistically significant, they were altered, as confirmed by the sequencing results (Figure 7).

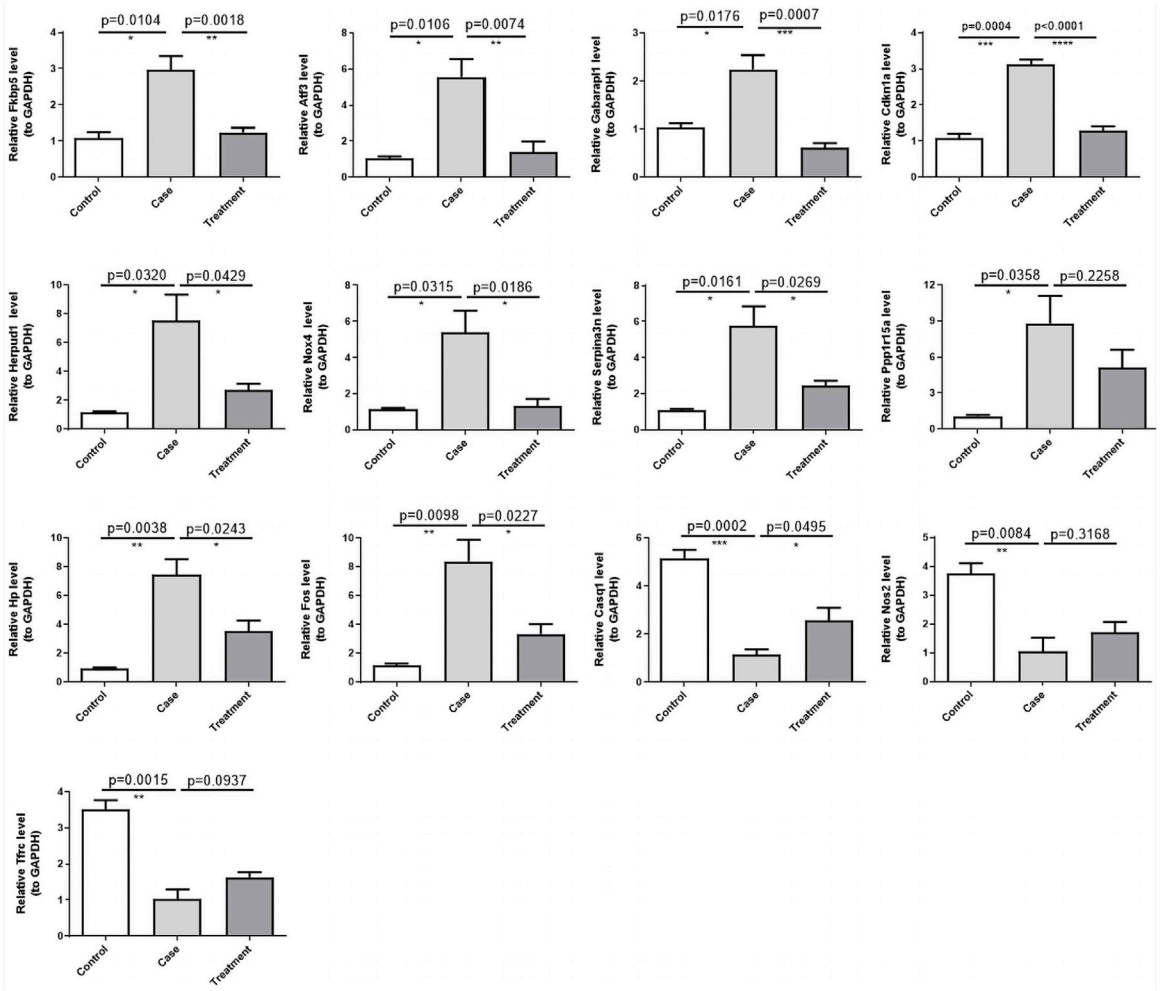


FIGURE 7
Validation of the expression of 13 DE-ERSRGs by RT-qPCR. Validation of the expression levels of 13 DE-ERSRGs in mice tissue samples by quantitative RT-PCR. Error bars show mean ± SD. Data are representative of three independent experiments. *$p < .05$, **$p < .01$, ***$p < .001$.

4 Discussion

Numerous studies, including epidemiological and experimental studies in humans and animals, have yielded compelling evidence favoring the strong association between psychosocial factors and cardiovascular morbidity. In current European guidelines for the prevention of CVDs, psychological stress is recognized as a potential contributing factor to both CVD development and progression (Piepoli et al., 2016). Yet, the limited understanding of the molecular mechanisms underpinning chronic stress-induced cardiac injury has led to the lack of effective treatment options. Several studies have shown that psychological stress is a determinant of the prognosis and outcome of cardiovascular events, rather than a cause of CVD (Sara et al., 2021). On one hand, psychological stress can contribute to some adverse lifestyle patterns in humans, such as the abuse of tobacco and alcohol, poor eating habits, or sleep disorders. These confounding factors complicate the assessment of the effects of mental factors on the cardiovascular system in a given context. On the other hand, in animal experiments, subjects suffering from CVDs were simultaneously exposed to psychological stress to investigate the effect of stress on the progression of pre-existing CVDs, but limited attention was paid to the potential molecular processes induced by chronic psychological stress that induce cardiac tissue damage in healthy individuals. Cho et al. evaluated the molecular basis of cardiac damage in healthy persons exposed to psychological stress using a transcriptome sequencing approach, but their results failed to adequately account for the irreversible damage caused to the heart by chronic stress in American soldiers (Cho et al., 2014). This might be related to the methods (social defeat) they used, because the recurrence of the same stressor usually leads to adaptation and failure to reflect the unpredictable, uncontrollable, complex, and variable properties of naturally occurring stressors. In our study, we addressed and resolved some of the limitations of their study. The chronic psychological stress model was slightly modified from the chronic unpredictability stress model by increasing the stimulus intensity to reduce adaptation and better simulate chronic stress. The results suggested that mice that were subjected to the stress stimuli exhibited significant differences in all four behavioral experiments compared to control mice. In this modified model, we observed the abnormal expression of some ERSRGs, and the occurrence of ER stress may be a cause of permanent cardiac

damage in healthy individuals suffering from chronic psychological stress.

In recent years, a large body of evidence from research has shown that ER stress is linked to the pathogenesis and progression of multiple CVDs and is an important target for the treatment of diverse CVDs. In this study, we identified 13 DE-ERSRGs. Among them, Fos was previously reported to be a downstream signaling molecule of ER stress in obesity models (Levi et al., 2018), and GRP78 was shown to interact with the transcription factor Fos to promote ER stress (Gopal and Pizzo, 2021). Fu et al. found that the overexpression of sarcoplasmic ER calcium ATPase (SERCA2a) reduced Hp expression while alleviating ER stress, thereby improving cardiac function in animals with heart failure. This suggests that Hp expression may be positively correlated with the ER stress levels in cardiac tissues (Fu et al., 2012). Serpina3 has been detected as a marker of ER stress in astrocytes in immune system diseases (Masvekar et al., 2022), and the role played by its Serpina3n (a protein from the same family) in ER stress has never been described. The role of NOX4, a member of the NADPH oxidase family, in ER stress in CVDs has garnered greater attention, as NOX4 participates in two UPR signaling pathways (Wu et al., 2010). In vascular smooth muscle cells and cardiomyocytes ontogenically related to VSMCs, high NOX4 expression induced by ER stress reduced cell death but promoted cell proliferation, leading to arterial remodeling and myocardial hypertrophy, and exacerbated the deterioration of cardiac function (Mittal et al., 2007; Kuroda et al., 2010). The PP1-Ppp1r15a phosphatase complex dephosphorylates eIF2α, and the anti-hypertensive compound guanabenz disrupts this complex and thereby attenuates ER stress in cardiomyocytes (Neuber et al., 2014). We also detected the upregulation of this complex in cardiac tissues post stress, which was reduced upon treatment with CB2R agonists. Herpud1 is an ER-resident membrane protein that plays a part in ER degradation, and its upregulation can exert a cytoprotective role, which is why we assumed that the upregulation of Herpud1 in this experiment is a compensatory protective mechanism (Navarro-Marquez et al., 2018). Furthermore, three DE-ERSRGs in our results showed a trend of downregulation in response to psychological stress and upregulation after treatment. Among them, Tfrc and Casq1 usually play a protective role and lead to the generation of ER stress when external factors cause their lowering (Jung et al., 2015; Hanna et al., 2021), which is also consistent with our results. Payne et al. suggested that Nos2 could protect against the development of ER stress (Payne et al., 2001), whereas in other studies, increased Nos2 expression has been suggested to exacerbate ER stress (Deslauriers et al., 2011; Zanotto et al., 2017). Whether the tendency of Nos2 downregulation in response to psychological stress results from a compensatory mechanism or some other mechanism remains elusive. Cdkn1a has been reported in ER stress (Lopez et al., 2015; Inoue et al., 2017). In this model, Cdkn1a probably acted as a negative regulator of the cell cycle, allowing cells sufficient time for repair of damaged DNA after external stimuli and various metabolic abnormalities. Moreover, Gabapal1 is involved in the formation of autophagosomal vacuoles (Chakrama et al., 2010). We hypothesized that it may participate in ER stress-induced autophagy. Atf3 is involved in ER stress through multiple pathways. In the most common pathway, Atf3 acts as an Atf4 downstream molecule in PERK-ATF4-ATF3-CHOP-induced apoptosis (Tang et al., 2020). In addition, Atf3 can also act as a transcription factor for NOX4 and upregulate NOX4-induced ROS production (Liu Z. et al., 2020). This connection can also explain the increased Nox4 expression observed in our results. Additionally, Fkbp5 is a well-known stress response gene that plays a central role in psychological stress-induced diseases (Le-Niculescu et al., 2020). The increase in Fkbp5 expression also corroborates the validity of our established model of chronic psychological stress. Targeted drugs of FKBP5 are being developed actively, and CB2R agonists may be one of the promising candidates.

Among the 13 DE-ERSRGs, four key gene could constitute the ceRNA regulative network. They have been reported to be strongly associated with the onset and progression of CVDs. The protein p21 encoded by cdkn1a is considered to be a marker of cardiomyocyte senescence in addition to inhibiting the cell cycle (Bernard et al., 2020). By increasing the levels of p21 intracellularly in cardiac myocytes, phenylalanine causes premature cardiomyocyte senescence (Czibik et al., 2021). Additionally, p21 overexpression in cardiac fibroblasts induces differentiation of primary cardiac fibroblasts into myofibroblasts and promotes cardiac remodeling (Roy et al., 2007). In a pressure overload-induced cardiac remodeling paradigm, cdkn1a knockdown mice showed reduced myocardial hypertrophy and fibrosis and improved cardiac function when compared to wild type mice (Xu and Tang, 2016). It is evident that cdkn1a is crucial in cardiac injury. Acute myocardial infarction (AMI) can induce the release of gabarapl1 from circulating endothelial cells, which can stimulate an increase in monocytes and neutrophils by recruiting NLRP3 inflammasomes, finally causing vascular inflammation during AMI (Zhang et al., 2022). ATF3, which functions as a "hub" for numerous adaptive cellular responses, is important for the progression and development of CVDs. ATF3 overexpression in cardiac tissues can promote myocardial hypertrophy and fibrosis (Okamoto et al., 2001). Elevated ATF3 expression due to different stimuli can damage endothelial cells to promote atherosclerosis (Cai et al., 2000). ATF3 was increased in carotid artery of hypertensive rats, and downregulation of ATF3 was associated with the protective effect of enalapril against hypertension (Liu et al., 2015). However, ATF3 plays a paradoxical role in the development of heart failure (Lin et al., 2014; Brooks et al., 2015). Thus, the role of ATF3 in CVDs is complicated. FKBP5 is a key molecule in the triggering of CVDs by psychological stress. FKBP5 can increase plaque destabilization by promoting local inflammation *via* the NF-κB pathway, according to previous research (Zannas et al., 2019). High levels of FKBP5 expression were also detected in platelets after AMI (Eicher et al., 2016). FKBP5 single nucleotide polymorphisms (SNPs) are closely related with depression and coronary heart disease co-morbidity (Wang et al., 2020).

Cho et al. assessed the dynamic course of gene expression in the mouse heart in the transcriptional level analysis after 1–10 days of exposure to social defeat (Cho et al., 2014). They identified extracellular matrix remodeling, immune responses (e.g., complement activation), and cell proliferation as the primary molecular-level alterations in cardiac tissues after psychological stress. Despite the differences in the models (social defeat vs. modified CMS), sequencing methods (microarray vs. whole transcriptome sequencing), and durations of stress (10 days vs. 3 weeks), we observed similar changes in gene expression patterns by analyzing the KEGG and GO enrichment results, suggesting that the same pathophysiological changes mentioned above were the basic reactions to psychological stress affecting the heart. In addition, we observed some notable changes, such as metabolic alterations in

tissues (including glycine, serine, and threonine metabolism, purine metabolism, nucleotide metabolism, and nicotinate and nicotinamide metabolism.) and ER stress, which was the focus of this study.

Notably, the use of CB2R agonists diminishes the severity of stress in mice, as seen in the behavioral results, although there was no statistically significant difference between the results of the elevated plus maze and tail suspension tests, and only a trend toward improvement was observed. Possibly, ER stress occurs in neurons in regions such as the hippocampus, amygdala, and striatum, which are engaged in the development of psychological stress (Pavlovsky et al., 2013; Jangra et al., 2016; Jangra et al., 2017; Liu et al., 2019). CB2R may attenuate the symptoms of psychological stress by reducing the expression of ERSRGs and other pathways. CB2R agonists have been reported to exert neuroprotective effects to counteract the symptoms of psychological stress through anti-inflammatory functions (Zoppi et al., 2014).

However, there remain some limitations in our study. First, the sample size was small, which could have led to the low concordance between the qRT-PCR and sequencing results. Second, the experimental subjects used were male mice, which might have introduced gender bias. Third, the sequencing of whole heart tissues may have masked specific pathological changes in a particular region. We intend to follow up with experimental validation of the regulatory network of the four key genes identified and the role of CB2R agonists. We will also focus on the impact of metabolite changes in the results. In summary, our findings provide a comprehensive overview of the molecular changes in post-transcriptional regulation underlying chronic psychological stress-induced cardiac injury, with ER stress as one of the important mechanisms. CB2R agonists can alleviate ER stress through a competitive endogenous RNA mechanism and thus can be used as a therapeutic target in chronic psychological stress-induced cardiac injury.

5 Conclusion

Our findings revealed, for the first time, the expression profiles of circRNA, lncRNA, miRNA, and mRNA in cardiac tissues exposed to chronic psychological stress. We identified four key DE-ERSRGs-related genes and established relevant ceRNA networks. Our findings may provide novel insights on the molecular mechanisms underlying the effects of CB2R agonists in the treatment of chronic psychological stress-induced cardiac diseases.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE210820; https://www.ncbi.nlm.nih.gov/geo/, GSE210252.

Ethics statement

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of Chinese PLA General Hospital.

Author contributions

FC and CQ conceived and designed the experiments. CQ, YZ (3rd author), and YZ (4th author) performed the experiments. CQ, YuW, and YaW analyzed the data. CQ wrote the manuscript. FC edited the manuscript.

Funding

This study was supported by the National Key Research and Development Projects (2022YFC3602400); the Key Health Care Projects of National Health Commission (2020ZD05); Basic Research Reinforcement Project (2022-JCJQ-ZD-079-00); Project of National Clinical Research Center for Geriatric Diseases (NCRCG-PLAGH-2022007).

Acknowledgments

We thank Prof. Jiyan Zhang at Beijing Institute of Basic Medical Sciences for providing the behavioral equipment.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1095428/full#supplementary-material

References

Almanza, A., Carlesso, A., Chintha, C., Creedican, S., Doultsinos, D., Leuzzi, B., et al. (2019). Endoplasmic reticulum stress signalling - from basic mechanisms to clinical applications. *FEBS J.* 286 (2), 241–278. doi:10.1111/febs.14608

Atkins, C., Liu, Q., Minthorn, E., Zhang, S. Y., Figueroa, D. J., Moss, K., et al. (2013). Characterization of a novel PERK kinase inhibitor with antitumor and antiangiogenic activity. *Cancer Res.* 73 (6), 1993–2002. doi:10.1158/0008-5472.CAN-12-3109

Bernard, M., Yang, B., Migneault, F., Turgeon, J., Dieudé, M., Olivier, M. A., et al. (2020). Autophagy drives fibroblast senescence through MTORC2 regulation. *Autophagy* 16 (11), 2004–2016. doi:10.1080/15548627.2020.1713640

Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P., and Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat. Cell Biol.* 2 (6), 326–332. doi:10.1038/35014014

Bi, X., Zhang, G., Wang, X., Nguyen, C., May, H. I., Li, X., et al. (2018). Endoplasmic reticulum chaperone GRP78 protects heart from ischemia/reperfusion injury through akt activation. *Circ. Res.* 122 (11), 1545–1554. doi:10.1161/CIRCRESAHA.117.312641

Blackwood, E. A., Hofmann, C., Santo Domingo, M., Bilal, A. S., Sarakki, A., Stauffer, W., et al. (2019). ATF6 regulates cardiac hypertrophy by transcriptional induction of the mTORC1 activator, rheb. *Circ. Res.* 124 (1), 79–93. doi:10.1161/CIRCRESAHA.118.313854

Brooks, A. C., DeMartino, A. M., Brainard, R. E., Brittian, K. R., Bhatnagar, A., and Jones, S. P. (2015). Induction of activating transcription factor 3 limits survival following infarct-induced heart failure in mice. *Am. J. Physiol. Heart Circ. Physiol.* 309 (8), H1326–H1335. doi:10.1152/ajpheart.00513.2015

Cai, Y., Zhang, C., Nawa, T., Aso, T., Tanaka, M., Oshiro, S., et al. (2000). Homocysteine-responsive ATF3 gene expression in human vascular endothelial cells: Activation of c-jun NH(2)-terminal kinase and promoter response element. *Blood* 96 (6), 2140–2148. doi:10.1182/blood.v96.6.2140.h8002140_2140_2148

Chakrama, F. Z., Seguin-Py, S., Le Grand, J. N., Fraichard, A., Delage-Mourroux, R., Despouy, G., et al. (2010). GABARAPL1 (GEC1) associates with autophagic vesicles. *Autophagy* 6 (4), 495–505. doi:10.4161/auto.6.4.11819

Chen, M. H., Pan, T. L., Li, C. T., Lin, W. C., Chen, Y. S., Lee, Y. C., et al. (2015). Risk of stroke among patients with post-traumatic stress disorder: Nationwide longitudinal study. *Br. J. Psychiatry* 206 (4), 302–307. doi:10.1192/bjp.bp.113.143610

Cho, J. H., Lee, I., Hammamieh, R., Wang, K., Baxter, D., Scherler, K., et al. (2014). Molecular evidence of stress-induced acute heart injury in a mouse model simulating posttraumatic stress disorder. *Proc. Natl. Acad. Sci. U. S. A.* 111 (8), 3188–3193. doi:10.1073/pnas.1400113111

Cho, W. H., Noh, K., Lee, B. H., Barcelon, E., Jun, S. B., Park, H. Y., et al. (2022). Hippocampal astrocytes modulate anxiety-like behavior. *Nat. Commun.* 13 (1), 6536. doi:10.1038/s41467-022-34201-z

Czibik, G., Mezdari, Z., Murat Altintas, D., Bréhat, J., Pini, M., d'Humières, T., et al. (2021). Dysregulated phenylalanine catabolism plays a key role in the trajectory of cardiac aging. *Circulation* 144 (7), 559–574. doi:10.1161/circulationaha.121.054204

Deslauriers, A. M., Afkhami-Goli, A., Paul, A. M., Bhat, R. K., Acharjee, S., Ellestad, K. K., et al. (2011). Neuroinflammation and endoplasmic reticulum stress are coregulated by crocin to prevent demyelination and neurodegeneration. *J. Immunol.* 187 (9), 4788–4799. doi:10.4049/jimmunol.1004111

Ebrahimi, R., Lynch, K. E., Beckham, J. C., Dennis, P. A., Viernes, B., Tseng, C. H., et al. (2021). Association of posttraumatic stress disorder and incident ischemic heart disease in women veterans. *JAMA Cardiol.* 6 (6), 642–651. doi:10.1001/jamacardio.2021.0227

Eicher, J. D., Wakabayashi, Y., Vitseva, O., Esa, N., Yang, Y., Zhu, J., et al. (2016). Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets* 27 (3), 230–239. doi:10.3109/09537104.2015.1083543

Frakes, A. E., and Dillin, A. (2017). The UPR(ER): Sensor and coordinator of organismal homeostasis. *Mol. Cell* 66 (6), 761–771. doi:10.1016/j.molcel.2017.05.031

Fu, Z., Li, F., Jia, L., Su, S., Wang, Y., Cai, Z., et al. (2019). Histone deacetylase 6 reduction promotes aortic valve calcification via an endoplasmic reticulum stress-mediated osteogenic pathway. *J. Thorac. Cardiovasc Surg.* 158 (2), 408–417. doi:10.1016/j.jtcvs.2018.10.136

Fu, Z. Q., Li, X. Y., Lu, X. C., Mi, Y. F., Liu, T., and Ye, W. H. (2012). Ameliorated stress related proteins are associated with improved cardiac function by sarcoplasmic reticulum calcium ATPase gene transfer in heart failure. *J. Geriatr. Cardiol.* 9 (3), 269–277. doi:10.3724/SP.J.1263.2012.05299

Glaser, R., and Kiecolt-Glaser, J. K. (2005). Stress-induced immune dysfunction: Implications for health. *Nat. Rev. Immunol.* 5 (3), 243–251. doi:10.1038/nri1571

Gold, P. W., Licinio, J., and Pavlatou, M. G. (2013). Pathological parainflammation and endoplasmic reticulum stress in depression: Potential translational targets through the CNS insulin, klotho and PPAR-gamma systems. *Mol. Psychiatry* 18 (2), 154–165. doi:10.1038/mp.2012.167

Gold, S. M., Kohler-Forsberg, O., Moss-Morris, R., Mehnert, A., Miranda, J. J., Bullinger, M., et al. (2020). Comorbid depression in medical diseases. *Nat. Rev. Dis. Prim.* 6 (1), 69. doi:10.1038/s41572-020-0200-2

Gopal, U., and Pizzo, S. V. (2021). Cell surface GRP78 signaling: An emerging role as a transcriptional modulator in cancer. *J. Cell Physiol.* 236 (4), 2352–2363. doi:10.1002/jcp.30030

Gu, Y., Ye, T., Tan, P., Tong, L., Ji, J., Gu, Y., et al. (2021). Tolerance-inducing effect and properties of innate immune stimulation on chronic stress-induced behavioral abnormalities in mice. *Brain Behav. Immun.* 91, 451–471. doi:10.1016/j.bbi.2020.11.002

Hanna, A. D., Lee, C. S., Babcock, L., Wang, H., Recio, J., and Hamilton, S. L. (2021). Pathological mechanisms of vacuolar aggregate myopathy arising from a Casq1 mutation. *FASEB J.* 35 (5), e21349. doi:10.1096/fj.202001653RR

Hinterdobler, J., Schott, S., Jin, H., Meesmann, A., Steinsiek, A. L., Zimmermann, A. S., et al. (2021). Acute mental stress drives vascular inflammation and promotes plaque destabilization in mouse atherosclerosis. *Eur. Heart J.* 42 (39), 4077–4088. doi:10.1093/eurheartj/ehab371

Hourani, L. L., Davila, M. I., Morgan, J., Meleth, S., Ramirez, D., Lewis, G., et al. (2020). Mental health, stress, and resilience correlates of heart rate variability among military reservists, guardsmen, and first responders. *Physiol. Behav.* 214, 112734. doi:10.1016/j.physbeh.2019.112734

Inoue, Y., Kawachi, S., Ohkubo, T., Nagasaka, M., Ito, S., Fukuura, K., et al. (2017). The CDK inhibitor p21 is a novel target gene of ATF4 and contributes to cell survival under ER stress. *FEBS Lett.* 591 (21), 3682–3691. doi:10.1002/1873-3468.12869

Jangra, A., Dwivedi, S., Sriram, C. S., Gurjar, S. S., Kwatra, M., Sulakhiya, K., et al. (2016). Honokiol abrogates chronic restraint stress-induced cognitive impairment and depressive-like behaviour by blocking endoplasmic reticulum stress in the hippocampus of mice. *Eur. J. Pharmacol.* 770, 25–32. doi:10.1016/j.ejphar.2015.11.047

Jangra, A., Sriram, C. S., Dwivedi, S., Gurjar, S. S., Hussain, M. I., Borah, P., et al. (2017). Sodium phenylbutyrate and edaravone abrogate chronic restraint stress-induced behavioral deficits: Implication of oxido-nitrosative, endoplasmic reticulum stress cascade, and neuroinflammation. *Cell Mol. Neurobiol.* 37 (1), 65–81. doi:10.1007/s10571-016-0344-5

Jordan, H. T., Stellman, S. D., Morabia, A., Miller-Archie, S. A., Alper, H., Laskaris, Z., et al. (2013). Cardiovascular disease hospitalizations in relation to exposure to the September 11, 2001 World Trade Center disaster and posttraumatic stress disorder. *J. Am. Heart Assoc.* 2 (5), e000431. doi:10.1161/JAHA.113.000431

Jung, I. R., Choi, S. E., Jung, J. G., Lee, S. A., Han, S. J., Kim, H. J., et al. (2015). Involvement of iron depletion in palmitate-induced lipotoxicity of beta cells. *Mol. Cell Endocrinol.* 407, 74–84. doi:10.1016/j.mce.2015.03.007

Kuroda, J., Ago, T., Matsushima, S., Zhai, P., Schneider, M. D., and Sadoshima, J. (2010). NADPH oxidase 4 (Nox4) is a major source of oxidative stress in the failing heart. *Proc. Natl. Acad. Sci. U. S. A.* 107 (35), 15565–15570. doi:10.1073/pnas.1002178107

Le-Niculescu, H., Roseberry, K., Levey, D. F., Rogers, J., Kosary, K., Prabha, S., et al. (2020). Towards precision medicine for stress disorders: Diagnostic biomarkers and targeted drugs. *Mol. Psychiatry* 25 (5), 918–938. doi:10.1038/s41380-019-0370-z

Leor, J., Poole, W. K., and Kloner, R. A. (1996). Sudden cardiac death triggered by an earthquake. *N. Engl. J. Med.* 334 (7), 413–419. doi:10.1056/NEJM199602153340701

Levi, N. J., Wilson, C. W., Redweik, G. A. J., Gray, N. W., Grzybowski, C. W., Lenkey, J. A., et al. (2018). Obesity-related cellular stressors regulate gonadotropin releasing hormone gene expression via c-Fos/AP-1. *Mol. Cell Endocrinol.* 478, 97–105. doi:10.1016/j.mce.2018.07.011

Li, X., Han, D., Tian, Z., Gao, B., Fan, M., Li, C., et al. (2016). Activation of cannabinoid receptor type II by AM1241 ameliorates myocardial fibrosis via nrf2-mediated inhibition of TGF-β1/smad3 pathway in myocardial infarction mice. *Cell Physiol. Biochem.* 39 (4), 1521–1536. doi:10.1159/000447855

Lin, H., Li, H. F., Chen, H. H., Lai, P. F., Juan, S. H., Chen, J. J., et al. (2014). Activating transcription factor 3 protects against pressure-overload heart failure via the autophagy molecule Beclin-1 pathway. *Mol. Pharmacol.* 85 (5), 682–691. doi:10.1124/mol.113.090092

Liu, L., Liu, J., Huang, Z., Yu, X., Zhang, X., Dou, D., et al. (2015). Berberine improves endothelial function by inhibiting endoplasmic reticulum stress in the carotid arteries of spontaneously hypertensive rats. *Biochem. Biophys. Res. Commun.* 458 (4), 796–801. doi:10.1016/j.bbrc.2015.02.028

Liu, L., Zhao, Z., Lu, L., Liu, J., Sun, J., Wu, X., et al. (2019). Icariin and icaritin ameliorated hippocampus neuroinflammation via inhibiting HMGB1-related pro-inflammatory signals in lipopolysaccharide-induced inflammation model in C57BL/6J mice. *Int. Immunopharmacol.* 68, 95–105. doi:10.1016/j.intimp.2018.12.055

Liu, X., Zhang, D., Dong, X., Zhu, R., Ye, Y., Li, L., et al. (2020a). Pharmacological activation of CB2 receptor protects against ethanol-induced myocardial injury related to RIP1/RIP3/MLKL-mediated necroptosis. *Mol. Cell Biochem.* 474 (1-2), 1–14. doi:10.1007/s11010-020-03828-1

Liu, Z., Gu, S., Lu, T., Wu, K., Li, L., Dong, C., et al. (2020b). IFI6 depletion inhibits esophageal squamous cell carcinoma progression through reactive oxygen species accumulation via mitochondrial dysfunction and endoplasmic reticulum stress. *J. Exp. Clin. Cancer Res.* 39 (1), 144. doi:10.1186/s13046-020-01646-3

Lopez, I., Tournillon, A. S., Nylander, K., and Fahraeus, R. (2015). p53-mediated control of gene expression via mRNA translation during Endoplasmic Reticulum stress. *Cell Cycle* 14 (21), 3373–3378. doi:10.1080/15384101.2015.1090066

Mao, J., Hu, Y., Ruan, L., Ji, Y., and Lou, Z. (2019). Role of endoplasmic reticulum stress in depression (Review). *Mol. Med. Rep.* 20 (6), 4774–4780. doi:10.3892/mmr.2019.10789

Masvekar, R., Kosa, P., Barbour, C., Milstein, J. L., and Bielekova, B. (2022). Drug library screen identifies inhibitors of toxic astrogliosis. *Mult. Scler. Relat. Disord.* 58, 103499. doi:10.1016/j.msard.2022.103499

Meisel, S. R., Kutz, I., Dayan, K. I., Pauzner, H., Chetboun, I., Arbel, Y., et al. (1991). Effect of Iraqi missile war on incidence of acute myocardial infarction and sudden death in Israeli civilians. *Lancet* 338 (8768), 660–661. doi:10.1016/0140-6736(91)91234-l

Mittal, M., Roth, M., Konig, P., Hofmann, S., Dony, E., Goyal, P., et al. (2007). Hypoxia-dependent regulation of nonphagocytic NADPH oxidase subunit NOX4 in the pulmonary vasculature. *Circ. Res.* 101 (3), 258–267. doi:10.1161/CIRCRESAHA.107.148015

Molica, F., Matter, C. M., Burger, F., Pelli, G., Lenglet, S., Zimmer, A., et al. (2012). Cannabinoid receptor CB2 protects against balloon-induced neointima formation. *Am. J. Physiol. Heart Circ. Physiol.* 302 (5), H1064–H1074. doi:10.1152/ajpheart.00444.2011

Mondal, T. K., Emeny, R. T., Gao, D., Ault, J. G., Kasten-Jolly, J., and Lawrence, D. A. (2015). A physical/psychological and biological stress combine to enhance endoplasmic reticulum stress. *Toxicol. Appl. Pharmacol.* 289 (2), 313–322. doi:10.1016/j.taap.2015.09.013

Navarro-Marquez, M., Torrealba, N., Troncoso, R., Vasquez-Trincado, C., Rodriguez, M., Morales, P. E., et al. (2018). Herpud1 impacts insulin-dependent glucose uptake in skeletal muscle cells by controlling the Ca(2+)-calcineurin-Akt axis. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864 (5), 1653–1662. doi:10.1016/j.bbadis.2018.02.018

Neuber, C., Uebeler, J., Schulze, T., Sotoud, H., El-Armouche, A., and Eschenhagen, T. (2014). Guanabenz interferes with ER stress and exerts protective effects in cardiac myocytes. *PLoS One* 9 (6), e98893. doi:10.1371/journal.pone.0098893

Ochoa, C. D., Wu, R. F., and Terada, L. S. (2018). ROS signaling and ER stress in cardiovascular disease. *Mol. Asp. Med.* 63, 18–29. doi:10.1016/j.mam.2018.03.002

Okamoto, Y., Chaves, A., Chen, J., Kelley, R., Jones, K., Weed, H. G., et al. (2001). Transgenic mice with cardiac-specific expression of activating transcription factor 3, a stress-inducible gene, have conduction abnormalities and contractile dysfunction. *Am. J. Pathol.* 159 (2), 639–650. doi:10.1016/s0002-9440(10)61735-x

Pavlovsky, A. A., Boehning, D., Li, D., Zhang, Y., Fan, X., and Green, T. A. (2013). Psychological stress, cocaine and natural reward each induce endoplasmic reticulum stress genes in rat brain. *Neuroscience* 246, 160–169. doi:10.1016/j.neuroscience.2013.04.057

Payne, C. M., Bernstein, H., Bernstein, C., Kunke, K., and Garewal, H. (2001). The specific NOS2 inhibitor, 1400W, sensitizes HepG2 cells to genotoxic, oxidative, xenobiotic, and endoplasmic reticulum stresses. *Antioxid. Redox Signal* 3 (5), 931–936. doi:10.1089/15230860152665082

Piepoli, M. F., Hoes, A. W., Agewall, S., Albus, C., Brotons, C., Catapano, A. L., et al. (2016). 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *Rev. Esp. Cardiol. Engl. Ed.* 69 (10), 939. doi:10.1016/j.rec.2016.09.009

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* 8 (7), 519–529. doi:10.1038/nrm2199

Roy, S., Khanna, S., Rink, T., Radtke, J., Williams, W. T., Biswas, S., et al. (2007). P21waf1/cip1/sdi1 as a central regulator of inducible smooth muscle actin expression and differentiation of cardiac fibroblasts to myofibroblasts. *Mol. Biol. Cell* 18 (12), 4837–4846. doi:10.1091/mbc.e07-03-0270

Roy, S. S., Foraker, R. E., Girton, R. A., and Mansfield, A. J. (2015). Posttraumatic stress disorder and incident heart failure among a community-based sample of US veterans. *Am. J. Public Health* 105 (4), 757–763. doi:10.2105/AJPH.2014.302342

Sara, J. D. S., Lerman, L. O., and Lerman, A. (2021). The endothelium is a key player in the vascular response to acute mental stress. *Eur. Heart J.* 42 (39), 4089–4091. doi:10.1093/eurheartj/ehab510

Shah, A. J., Weeks, V., Lampert, R., Bremner, J. D., Kutner, M., Raggi, P., et al. (2022). Early life trauma is associated with increased microvolt T-wave alternans during mental stress challenge: A substudy of mental stress ischemia: Prognosis and genetic influences. *J. Am. Heart Assoc.* 11 (5), e021582. doi:10.1161/JAHA.121.021582

Steinberg, J. S., Arshad, A., Kowalski, M., Kukar, A., Suma, V., Vloka, M., et al. (2004). Increased incidence of life-threatening ventricular arrhythmias in implantable defibrillator patients after the World Trade Center attack. *J. Am. Coll. Cardiol.* 44 (6), 1261–1264. doi:10.1016/j.jacc.2004.06.032

Tang, Q., Ren, L., Liu, J., Li, W., Zheng, X., Wang, J., et al. (2020). Withaferin A triggers G2/M arrest and intrinsic apoptosis in glioblastoma cells via ATF4-ATF3-CHOP axis. *Cell Prolif.* 53 (1), e12706. doi:10.1111/cpr.12706

Tang, V., Fu, S., Rayner, B. S., and Hawkins, C. L. (2019). 8-Chloroadenosine induces apoptosis in human coronary artery endothelial cells through the activation of the unfolded protein response. *Redox Biol.* 26, 101274. doi:10.1016/j.redox.2019.101274

Vaccarino, V., Almuwaqqat, Z., Kim, J. H., Hammadah, M., Shah, A. J., Ko, Y. A., et al. (2021). Association of mental stress-induced myocardial ischemia with cardiovascular events in patients with coronary heart disease. *JAMA* 326 (18), 1818–1828. doi:10.1001/jama.2021.17649

Wang, H., Wang, C., Song, X., Liu, H., Zhang, Y., and Jiang, P. (2020). Association of FKBP5 polymorphisms with patient susceptibility to coronary artery disease comorbid with depression. *PeerJ* 8, e9286. doi:10.7717/peerj.9286

Wang, M., and Kaufman, R. J. (2016). Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature* 529 (7586), 326–335. doi:10.1038/nature17041

Wu, R. F., Ma, Z., Liu, Z., and Terada, L. S. (2010). Nox4-derived H2O2 mediates endoplasmic reticulum signaling through local Ras activation. *Mol. Cell Biol.* 30 (14), 3553–3568. doi:10.1128/MCB.01445-09

Xu, M., and Tang, Q. (2016). GW27-e0439 Cdkn1α knockout alleviates pressure overload induced cardiac remodeling. *J. Am. Coll. Cardiol.*, 68(16), C58–C58. doi:10.1016/j.jacc.2016.07.212

Yan, J. J., Ding, X. J., He, T., Chen, A. X., Zhang, W., Yu, Z. X., et al. (2022). A circuit from the ventral subiculum to anterior hypothalamic nucleus GABAergic neurons essential for anxiety-like behavioral avoidance. *Nat. Commun.* 13 (1), 7464. doi:10.1038/s41467-022-35211-7

Yu, W., Jin, G., Zhang, J., and Wei, W. (2019). Selective activation of cannabinoid receptor 2 attenuates myocardial infarction via suppressing NLRP3 inflammasome. *Inflammation* 42 (3), 904–914. doi:10.1007/s10753-018-0945-x

Zannas, A. S., Jia, M., Hafner, K., Baumert, J., Wiechmann, T., Pape, J. C., et al. (2019). Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB-driven inflammation and cardiovascular risk. *Proc. Natl. Acad. Sci. U. S. A.* 116 (23), 11370–11379. doi:10.1073/pnas.1816847116

Zanotto, T. M., Quaresma, P. G. F., Guadagnini, D., Weissmann, L., Santos, A. C., Vecina, J. F., et al. (2017). Blocking iNOS and endoplasmic reticulum stress synergistically improves insulin resistance in mice. *Mol. Metab.* 6 (2), 206–218. doi:10.1016/j.molmet.2016.12.005

Zhang, B. F., Jiang, H., Chen, J., Guo, X., Li, Y., Hu, Q., et al. (2019). Nobiletin ameliorates myocardial ischemia and reperfusion injury by attenuating endoplasmic reticulum stress-associated apoptosis through regulation of the PI3K/AKT signal pathway. *Int. Immunopharmacol.* 73, 98–107. doi:10.1016/j.intimp.2019.04.060

Zhang, Q., Guan, G., Cheng, P., Cheng, W., Yang, L., and Wu, A. (2021). Characterization of an endoplasmic reticulum stress-related signature to evaluate immune features and predict prognosis in glioma. *J. Cell Mol. Med.* 25 (8), 3870–3884. doi:10.1111/jcmm.16321

Zhang, T., Hou, D., He, J., Zeng, X., Liu, R., Liu, L., et al. (2022). Oxidative-damaged mitochondria activate GABARAPL1-induced NLRP3 inflammasomes in an autophagic-exosome manner after acute myocardial ischemia. *Oxid. Med. Cell Longev.* 2022, 7958542. doi:10.1155/2022/7958542

Zoppi, S., Madrigal, J. L., Caso, J. R., Garcia-Gutierrez, M. S., Manzanares, J., Leza, J. C., et al. (2014). Regulatory role of the cannabinoid CB2 receptor in stress-induced neuroinflammation in mice. *Br. J. Pharmacol.* 171 (11), 2814–2826. doi:10.1111/bph.12607

frontiers | Frontiers in Cell and Developmental Biology

TYPE Original Research
PUBLISHED 13 April 2023
DOI 10.3389/fcell.2023.1094588

OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean Universiity, China

REVIEWED BY
Dongbo Jiang,
Air Force Medical University, China
Yuzhen Gao,
Zhejiang University, China

*CORRESPONDENCE
Jun Zhao,
junzhao@suda.edu.cn
Xin Tong,
txin@suda.edu.cn
Chun Xu,
xuchun@suda.edu.cn

†These authors have contributed equally to this work

RECEIVED 10 November 2022
ACCEPTED 03 April 2023
PUBLISHED 13 April 2023

CITATION
Chen Z, Yang J, Tang L, Sun X, Li Y, Sheng Z, Ding H, Xu C, Tong X and Zhao J (2023), SUMOylation patterns and signature characterize the tumor microenvironment and predict prognosis in lung adenocarcinoma.
Front. Cell Dev. Biol. 11:1094588.
doi: 10.3389/fcell.2023.1094588

SUMOylation patterns and signature characterize the tumor microenvironment and predict prognosis in lung adenocarcinoma

Zhike Chen[1,2†], Jian Yang[1,2†], Lijuan Tang[3,4†], Xue Sun[5], Yu Li[1,2], Ziqing Sheng[1,2], Hao Ding[1,2], Chun Xu[1,2*], Xin Tong[1,2*] and Jun Zhao[1,2*]

[1]Institute of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China, [2]Department of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China, [3]Dalian Medical University, Dalian, China, [4]Department of Pathology, Affiliated Hospital of Nantong University, Nantong, China, [5]Department of Pathology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

Background: Recent studies have revealed that SUMOylation modifications are involved in various biological processes, including cancer development and progression. However, the precise role of SUMOylation in lung adenocarcinoma (LUAD), especially in the tumor immune microenvironment, is not yet clear.

Methods: We identified SUMOylation patterns by unsupervised consensus clustering based on the expression of SUMOylation regulatory genes. The tumor microenvironment in lung adenocarcinoma was analyzed using algorithms such as GSVA and ssGSEA. Key genes of SUMOylation patterns were screened for developing a SUMOylation scoring model to assess immunotherapy and chemotherapy responses in lung adenocarcinoma patients. Experiments were conducted to validate the differential expression of model genes in lung adenocarcinoma. Finally, we constructed a nomogram based on the SUMOylation score to assess the prognosis of individual lung adenocarcinoma patients.

Results: Two patterns of SUMOylation were identified, namely, SUMO-C1, which showed anti-tumor immune phenotype, and SUMO-C2, which showed immunosuppressive phenotype. Different genomic subtypes were also identified; subtype gene-T1 exhibited a reciprocal restriction between the immune microenvironment and stromal microenvironment. High SUMOylation scores were indicative of poor lung adenocarcinoma prognosis. SUMOylation score was remarkably negatively correlated with the infiltration of anti-tumor immune cells, and significantly positively correlated with immune cells promoting immune escape and immune suppression. In addition, patients with low scores responded better to immunotherapy. Therefore, the developed nomogram has a high prognostic predictive value.

Conclusion: The SUMOylation patterns can well discriminate the tumor microenvironment features of lung adenocarcinoma, especially the immune cell infiltration status. The SUMOylation score can further assess the

relationship between SUMOylation and immune cell crosstalk and has significant prognostic value and can be used to predict immunotherapy and chemotherapy response in patients with lung adenocarcinoma.

KEYWORDS

SUMOylation, tumor microenvironment, immune therapy, prognosis, lung adenocarcinoma

1 Introduction

Lung cancer is among the most commonly diagnosed cancers that pose a critical threat to public health (Siegel et al., 2022) and is a leading cause of cancer-related deaths worldwide (Minguet et al., 2016; Horn et al., 2017), approximately 85% of lung cancers are Non-small cell lung cancer with lung adenocarcinoma being the predominant pathological type. Increasing evidence suggests that the tumor microenvironment (TME) plays a critical role in local drug resistance, immune escape, cancer metastasis, and recurrence. It is well known that macrophage M2 type polarization, T regulatory (Treg) cells and MDSC cells can facilitate the advancement of lung adenocarcinoma, and Th2 helper cells can help the tumor immune escape process, however, $CD8^+$ T cells and cytotoxic T cells, etc. are the main force of anti-tumor immunity (Anderson and Simon, 2020). Understanding the tumor microenvironment has profound relevance in preventing metastasis, surmounting acquired drug resistance, and boosting therapeutic efficacy.

SUMOylation is one of the post-translational protein modifications and participates in the regulation of the cell cycle, DNA replication and repair, immunomodulation, and other biological processes (Eifler and Vertegaal, 2015; Chang and Yeh, 2020). Five categories of regulatory genes, including SUMO isoforms, SUMO-Activating Enzyme, SUMO-Conjugating Enzyme, SUMO E3 ligases, and SUMO proteases, regulate the homeostasis of SUMOylation and deSUMOylation (Bawa-Khalfe and Yeh, 2010; Chang and Yeh, 2020). SUMOylation protein modification is associated with Myc-driven tumor heterogeneity (Kessler et al., 2012). SENP7 has been reported to sustain the metabolic fitness and effector functions of tumor-infiltrating $CD8^+$ T cells (Wu et al., 2022). Overexpression of *SENP1* has also been shown to be involved in lung adenocarcinoma progression (Wang et al., 2013), and *SAE1*, *UBC9*, and *SENP3* are remarkably upregulated in lung adenocarcinoma and are associated with poor prognosis (Han et al., 2010). Recent studies suggest that the dynamic regulatory processes of SUMOylation and deSUMOylation may modulate antitumor immune processes in the TME by targeting multiple immunocytes. PIAS3 inhibits the activated STAT3 signaling pathway to engage in antitumor immune regulation (Zou et al., 2020). The overexpression of SUMO2 increased IL6 production by T cells, which are potent killers of tumor cells (Won et al., 2015). Decreased UBC9 expression suppressed Treg activation and the production of immune inhibitory molecules including IL10, CTLA4, PD-1, and ICOS (Shevach, 2009). AK-981, an inhibitor of SUMOylation, can directly stimulate T cells and enhance T cell sensitivity as well as response to antigens *in vitro* and *in vivo* (Lightcap et al., 2021). In this study, we explored the critical roles of SUMOylation in lung adenocarcinoma TME.

Several studies have revealed that SUMOylation regulatory genes are involved in the process of TME regulation (Chang and Ding, 2018). However, these studies are limited in that they study individual SUMOylation regulators, and a comprehensive analysis of the effect of SUMOylation on the tumor immune microenvironment has not been conducted before. By identifying different SUMO-associated TME types modulated by multiple SUMOylation regulators, we can gain a more comprehensive understanding of the regulation of the lung adenocarcinoma tumor microenvironment by SUMOylation. Therefore, we developed and validated SUMOylation patterns and signature to predict tumor microenvironmental phenotype and prognostic risk in lung adenocarcinoma.

2 Materials and methods

2.1 Data source and processing

Lung adenocarcinoma datasets GSE31210 (N = 226) (Okayama et al., 2012), GSE37745 (N = 106) (Botling et al., 2013), GSE50081 (N = 180) (Der et al., 2014), and GSE72094 (N = 398) (Schabath et al., 2016) were downloaded from the NCBI gene expression omnibus (GEO) database. The samples with missing survival data were removed and all data were log2 (x+1) normalized and corrected. The ComBat method in the "SVA" package was used to eliminate batch effects among the four GEO datasets and to integrate the four GEO datasets for subsequent analyses. All datasets contained a total of 857 lung adenocarcinoma sample data containing clinical information such as survival data, gender, age, pathological stage, and smoking history. The lung adenocarcinoma transcriptome expression dataset was downloaded through the cancer genome atlas (TCGA) database containing a total of 515 lung adenocarcinoma samples (different samples from the same patient were homogenized by taking the mean value) and 59 normal tissue samples; all expression data were downloaded in the fragments per kilobase of exon per million mapped fragments format and converted to the transcripts per kilobase million (TPM) format with log2 (TPM+1) data normalization operation. The survival data of 500 tumor samples was known, of which the data of samples with a survival time of 0 was removed and all others were rounded to survival time in years. Mutation data and copy number variation (CNV) data for lung adenocarcinoma were downloaded from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/).

2.2 Unsupervised consensus clustering for the identification of SUMOylation patterns

By reviewing the literature related to SUMOylation, we identified 33 genes recognized as critical regulators in SUMOylation (Chang and Yeh, 2020). Subsequently, we performed unsupervised consensus clustering using the "ConsensusClusterPlus" package based on the mRNA expression of the 33 SUMOylation regulators. We use the "km" clustering algorithm, the distance calculation algorithm is "euclidean," and the random seed is set to "123,456". The optimal clusters were determined under the cumulative distribution curve (CDF), and the rationality of the clusters was further verified using principal component analysis (PCA). Consensus clustering was adopted with 1,000,000 replicates to optimize clustering results. The clustering method for unsupervised consensus clustering analysis based on the expression of differentially expressed key SUMOylation genes is consistent with the above.

2.3 Gene set variation analysis and biological pathways

We used the R package GSVA (Hanzelmann et al., 2013) to study the biological processes associated with different SUMOylation modification patterns. Hallmark gene sets (Liberzon et al., 2015) and well-defined gene sets of biological features were downloaded from MsigDB (http://www.gsea-msigdb.org/gsea/msigdb/). The gene set of tumor-associated biological pathways was were obtained from the compilation by Mariathasan et al. (Mariathasan et al., 2018) and included the following: angiogenesis, immune checkpoint, cell cycle regulators, Pan fibroblast transforming growth factor beta (Pan F TBRs), epithelial-mesenchymal transition processes involved in EMT1, EMT2, EMT3, and the cell cycle. The immune checkpoint genes (including immune co-stimulatory and immune co-inhibitory molecules) and MHC molecules were derived from the immune-related gene set compiled by Rooney et al. (Rooney et al., 2015).

2.4 Gene ontology (GO) annotation, GO enrichment analysis, and KEGG enrichment analysis

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases are used to annotate genes and their RNA or protein products to determine their unique biological properties. GO annotation and KEGG annotation of SUMOylation-related genes was performed using the R package "clusterProfiler" with a cutoff value of false discovery rate (FDR<0.05).

2.5 Estimation of immune cell infiltration by ssGSEA and deconvolution algorithm

The gene set for each immune cell-specific marker was compiled from a recent study, and Single Sample Gene Set Enrichment Analysis (ssGSEA) (Barbie et al., 2009) was used to quantify the relative abundance of 24 immune cell types in the TME and was expressed as an enrichment fraction (Bindea et al., 2013). Subsequently, the EPIC immune infiltration algorithm (Racle et al., 2017) and the TIMER immune algorithm (Li et al., 2017) were used to calculate the extent of infiltration of each immune cell in lung adenocarcinoma tissue. Tumor tissue with abundant immune cell infiltration is consistent with a higher immune score and lower tumor purity.

2.6 Identification of differentially expressed genes (DEGs) associated with SUMOylation patterns and the construction of their protein-protein interaction (PPI) networks

The R package "limma" was used to measure DEGs associated with different SUMOylation patterns in lung adenocarcinoma samples, with the significance filtering criteria for DEGs of $|FC| > 1.5$ and FDR <0.05. The interaction of DEGs was investigated using the STRING database and used to build PPI networks, and the screening network type was a physical subnetwork with a confidence score ≥0.7. We used the "CytoHubba" plugin (Chin et al., 2014) in "Cytoscape" software (Shannon et al., 2003) to calculate the number of neighboring nodes in the PPI network, and filtered genes with neighboring nodes ≥5 as SUMOylation-related key genes.

2.7 Calculation of SUMOylation scores

We performed a univariate cox regression analysis using DEGs associated with SUMOylation modification patterns. Next, we screened the significant prognostic correlates ($p < 0.05$) among them for the optimal variables for constructing SUMOylation scores using the least absolute shrinkage and selection operator (LASSO) regression method. The SUMO score was calculated as follows: RiskScore = Exp (Gene1)*β1+Exp (Gene2)*β2+. . . + Exp (Genen) *βn. Where Exp is the expression of the genes and β is the regression coefficient calculated by LASSO. Patients were then divided into the high-risk and low-risk groups according to the optimal cut-off value calculated using the surv-cutpoint function in the R package "survival". Prognostic analyses were conducted using the R package "survminer". Receiver operating characteristic (ROC) curves were used to evaluate the prognostic performance of the SUMOylation score and the nomogram model. The area under the curve (AUC) was calculated using the R package "timeROC," while multiple regression analysis was used to verify the validity of the score as an independent prognostic factor for lung adenocarcinoma.

2.8 Response to immunotherapy and sensitivity to chemotherapeutic substances

The immunotherapy cohort GSE126044 (Cho et al., 2020) contained a total of 16 lung adenocarcinoma recipients of nivolumab (anti-PD1 drug) treatment, five of which were responsive to (CR and PR) and 11 non-responsive to

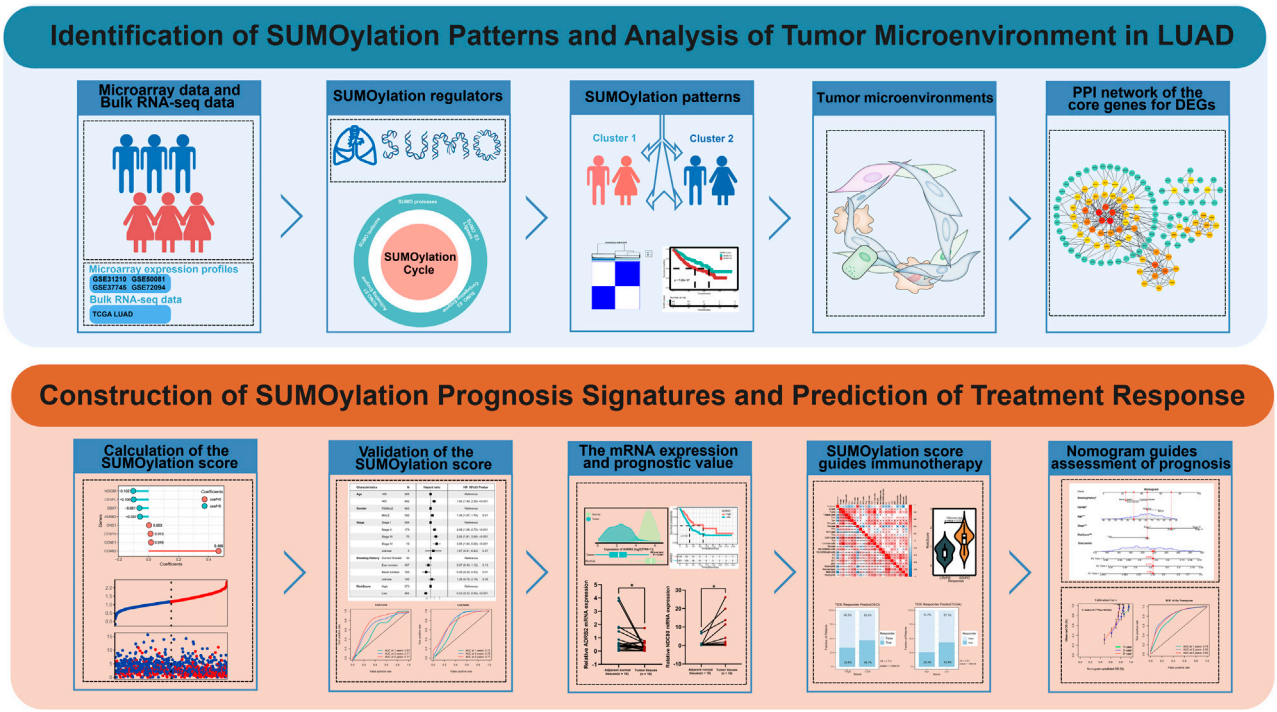


FIGURE 1
Flow chart of our study.

immunotherapy (PD and SD). (Tumor Immune Dysfunction and Exclusion) TIDE (http://tide.dfci.harvard.edu/) is applied to predict immunotherapy response for lung adenocarcinoma patients. Gene expression data from pretreatment tissue samples were collated and transformed into TPM format with log2 (TPM+1) normalization for further analysis. Chemotherapy drug sensitivity analysis was performed using the R package "pRRophetic" (Geeleher et al., 2014) to predict clinical response to chemotherapy.

2.9 Clinical samples, RNA extraction, and quantitative real-time PCR

Eighteen pairs of LUAD tissues and adjacent non-cancerous lung tissues were obtained from the First Affiliated Hospital of Soochow University after informed consent from patients. Pathological diagnostics for patients with LUAD were assessed according to the Revised International System for Staging Lung Cancer. All patients included in this study did not receive any radiation therapy, chemotherapy, or immunotherapy. The experiment was approved by the Academic Advisory Board of Soochow University.

Total RNA was isolated using Trizol reagent (Beyotime, China) and reverse transcribed into cDNA by Hiscript III Reverse Transcriptase (Vazyme, China). Real-Time PCR was performed using ChamQ SYBR qPCR Master Mix (Vazyme, China) on a LightCycler96 real-time PCR system (Roche, Switzerland). The specific primers we used are listed in Supplementary Table S1.

2.10 Statistical analysis

All statistical analyses in this study were performed using the R software 4.1.2 or GraphPad Prism 9. For quantitative data, normally distributed variables were analyzed using the Student's t-test, and non-normally distributed variables were analyzed using the Wilcoxon rank-sum test. Statistical significant differences between three and more data sets were analyzed using the Kruskal–Wallis method for non-parametric statistical tests and one-way ANOVA for parametric statistical tests. We calculated correlation coefficients using Spearman and distance correlation analysis. Survival analyses were conducted using the Kaplan–Meier method and log-rank tests, while Cox proportional risk regression models were used to analyze the relationship between SUMOylation patterns and regulatory genes and prognosis. All statistical comparisons in this study were two-sided with $\alpha = 0.05$ and the Benjamini-Hochberg method was employed to control the false discovery rate (FDR) for multiple hypothesis testing. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

3 Results

3.1 Genetic variation in SUMOylation regulatory genes in lung adenocarcinoma

The flow chart shows our research procedure (Figure 1). We investigated the role of 33 SUMOylation-modified regulatory genes

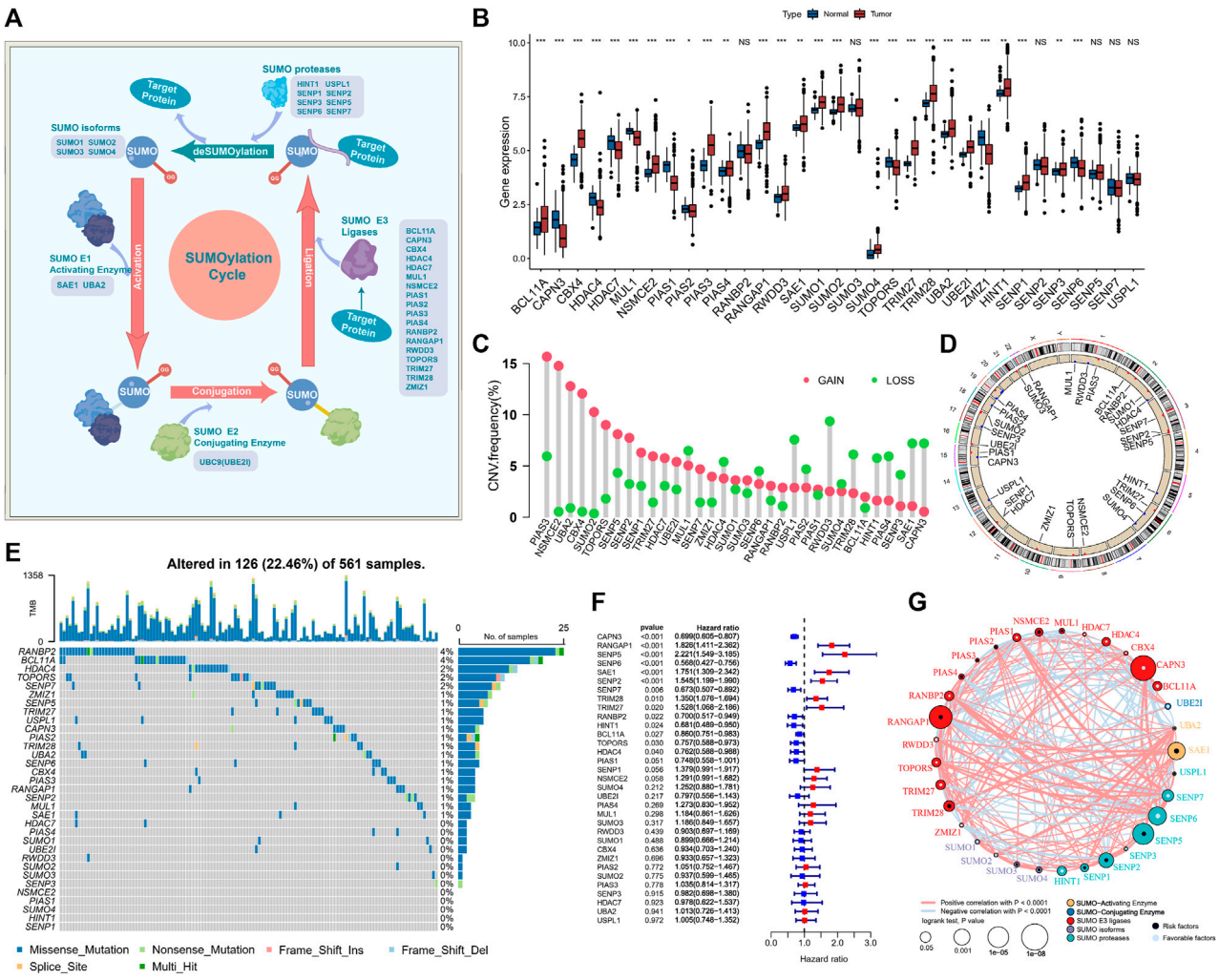


FIGURE 2
SUMOylation regulatory gene landscape in lung adenocarcinoma. **(A)** The reversible process regulation of SUMOylation and deSUMOylation. **(B)** Differential expression of SUMOylation regulatory genes among lung adenocarcinoma and normal sample tissues. **(C)** Frequency of CNV (copy number variation) of SUMOylation regulatory genes in lung adenocarcinoma. **(D)** Chromosomal localization of the CNV of SUMOylation regulatory genes. **(E)** The landscape of tumor somatic mutations in SUMOylation regulatory genes in lung adenocarcinoma. **(F)** Univariate cox prognostic analysis of SUMOylation regulatory genes in lung adenocarcinoma. **(G)** Interaction of SUMOylation regulatory genes in lung adenocarcinoma.

in lung adenocarcinoma, namely, SUMO isoforms (*SUMO1*, *SUMO2*, *SUMO3*, *SUMO4*), SUMO-Activating Enzyme (*SAE1*, *UBA2*), SUMO-Conjugating Enzyme (*UBE2I*), SUMO E3 Ligases (*BCL11A*, *CAPN3*, *CBX4*, *HDAC4*, *HDAC7*, *MUL1*, NSMCE2, PIAS1, PIAS2, PIAS3, PIAS4, RANBP2, *RANGAP1*, *RNF212*, *RWDD3*, *TOPORS*, *TRIM27*, *TRIM28*, *ZMIZ1*), SUMO proteases (*HINT1*, *SENP1*, *SENP2*, *SENP3*, *SENP6*, *SENP5*, *SENP7*, *USPL1*). The reversible process of SUMOylation is regulated by the aforementioned regulatory genes (Figure 2A). We first performed GO and KEGG enrichment analysis of the 33 SUMOylation regulatory genes. GO enrichment revealed that the 33 SUMOylation-regulated genes were remarkably enriched in biological processes such as lysine modification, protein SUMOylation, regulation of small protein covalent protein modification (Supplementary Figure S1A), and molecular functions such as ubiquitinated protein convertase activity, SUMO convertase activity, among others (Supplementary Figure S1B). KEGG enrichment analysis revealed other SUMOylation-related pathways (Supplementary Figure S1C).

Subsequently, for the comprehensive understanding of the SUMOylation regulatory genes, we performed the Spearman correlation analysis of the 33 SUMOylation regulatory genes to assess their pattern of co-expression (Supplementary Figure S1D) and found a broad significant negative correlation between the expression of *CAPN3* and *PIAS1* and other regulatory genes, as well as a significant positive correlation between *CAPN3* and *PIAS1* expression. Notably, most of the SUMOylation regulatory genes were remarkably upregulated in tumor tissues, and overall, the expression of SUMOylation regulatory genes was remarkably different between the tumor and normal samples (Figure 2B). PCA based on paired tumor and normal samples identified the SUMOylation regulatory genes whose expression was significantly different between normal and tumor samples (Supplementary Figure S1E).

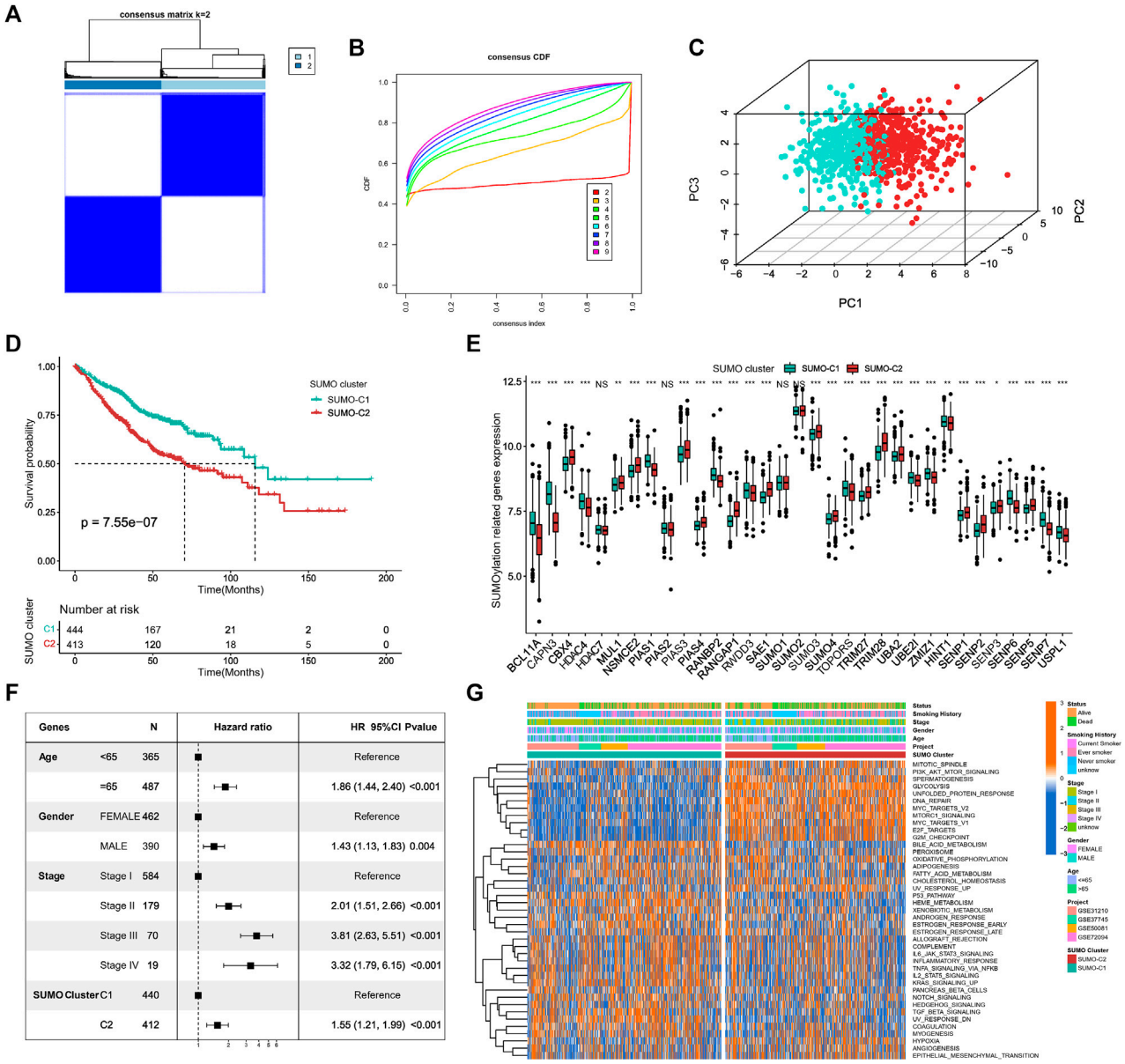


FIGURE 3
Identification of SUMOylation patterns and their associated biological functions. **(A)** Unsupervised cluster analysis based on 33 SUMOylation regulatory genes, K = 2. **(B)** The cumulative distribution curve (CDF) suggests that the optimal K value for unsupervised clustering is 2. **(C)** Principal component analysis (PCA) based on mRNA expression data of SUMOylation regulatory genes. **(D)** Kaplan-Meier curves for overall survival OS of two SUMOylation patterns for a total of 857 lung adenocarcinoma samples (including 444 SUMO-C1 and 413 SUMO-C2 cases). (Log-Rank test p = 7.55e-07). **(E)** SUMOylation regulatory genes were significantly differentially expressed between the two SUMOylation models. **(F)** SUMOylation patterns can be an independent prognostic factor for lung adenocarcinoma. **(G)** Heatmap showing the GSVA scores of the biological pathways of the HALLMARK gene set among two SUMOylation patterns in lung adenocarcinoma. Annotated with Status, Smoking History, Stage, Gender, Age, GEO cohort, SUMOylation patterns. Orange is high expression and blue is low expression.

Moreover, CNVs were very common in the 33 regulatory genes; CNV amplifications were prevalent in *PIAS3*, *NSMCE2*, *UBA2*, *CBX4*, and *SUMO2*, while CNV deletions were prevalent in *RWDD3*, *USPL1*, *SAE1*, and *CAPN3* (Figure 2C). The CNV occurred widely in genes on a variety of chromosomes (Figure 2D); however, they were concentrated mainly on chromosomes 1, 3, 4, 6, 12, 15, and 19. We further evaluated the prevalence of somatic mutations in SUMOylation regulatory genes. Of the 561 lung adenocarcinoma samples, 126 (22.46%) had alterations, mainly missense mutations and non-sense mutations, in SUMOylation regulatory genes. *RANBP2* and *BCL11A* had the highest mutation frequency of 4%, followed by *HDAC4*, *TOPORS*, and *SENP7* with 2% mutation frequency; most mutations occurred in the regulatory genes SUMO activator and E3 ligase, and less in SUMO isoforms (Figure 2E). Subsequently, we analyzed the mutational co-occurrence of the SUMOylation regulatory genes and found that SUMO2 co-occurred with mutations in both RANGAP1 and PIAS4 (Supplementary Figure S1F).

Furthermore, to determine the relationship between the 33 SUMOylation regulatory genes and the prognosis of lung adenocarcinoma patients, we performed a cox risk proportional regression model was used to determine the relationship between the 33 SUMOylation regulatory genes and the prognosis of lung adenocarcinoma patients, and the forest plot revealed that *CAPN3*, *SENP6*, *SENP7*, *RANBP2*, *HINT1*, *BCL11A*, *TOPORS*, and *HDAC4* were protective factors for lung adenocarcinoma and were generally downregulated, while *RANGAP1*, *SENP5*, *SAE1*, *SENP2*, *TRIM28*, and *TRIM27* were risk factors for lung adenocarcinoma and were upregulated in lung adenocarcinoma (Figure 2F). The multivariate Cox proportional Hazards regression analysis further established that SUMOylation regulatory genes were significantly associated with lung adenocarcinoma prognosis (Supplementary Figure S1G). In conclusion, we mapped the relationship between SUMOylation regulatory genes and lung adenocarcinoma prognosis (Figure 2G). In summary, we determined that SUMOylation regulatory genes differed significantly between normal and lung adenocarcinoma tissues and were also markedly correlated with the frequency of mutations and CNVs in lung adenocarcinoma. In addition, we demonstrated that altered expression and genetic variation of specific SUMOylation regulatory genes have a critical role in the development and prognosis of lung adenocarcinoma.

3.2 Identification of two different SUMOylation patterns based on the expression of 33 SUMOylation-regulated genes

For exploring different SUMOylation patterns, we collated 857 lung adenocarcinoma patients in four datasets of the GEO database for unsupervised consistency clustering based on 33 SUMOylation regulatory genes. Based on the variation of the area under the CDF curve (Figures 3A, B; Supplementary Figure S2A), we determined the optimal number of clusters to be 2 (K = 2). We subsequently performed PCA of the two SUMOylation patterns based on mRNA expression data of the regulatory genes (Figure 3C) and found that the two different modification patterns could be completely distinguished by SUMOylation regulatory genes. Meanwhile, Kaplan–Meier survival analysis revealed that patients with the SUMO-C2 pattern showed a worse prognostic survival (log-Rank test p = 7.55e-07) (Figure 3D). The occurrence of the SUMO-C2 pattern was consistent with the downregulation of BCL11A, CAPN3, HDAC4, PIAS4, UBA2, UBE2I, SENP6, SENP7, and USPL1, and the upregulation of CBX4, NSMCE2, PIAS3, RANGAP1, TRIM28 and SENP2 (Figure 3E). The Heatmap further confirmed that the expression of the 33 SUMOylation-regulated genes was significantly different in the two SUMOylation patterns (Supplementary Figure S2B). The multivariate Cox regression results indicated, in agreement with Kaplan–Meier results, that the SUMOylation patterns can be an independent prognostic factor, and compared with SUMO-C1, SUMO-C2 was associated with poor prognostic survival, HR = 1.55, p < 0.001 (Figure 3F).

To further explore the differences in key biological pathways associated with the two SUMOylation patterns. The GSVA algorithm was used to calculate the Hallmark gene set scores and convert them into a scoring matrix of the gene set. We found that there were significant differences in GSVA scores of Hallmark gene sets consistent with SUMO-C1 and SUMO-C2 (Figure 3G), and SUMO-C2 were mainly enriched in biological pathways related to the cell cycle, glucose metabolism, and genetic material replication and repair, such as G2M checkpoint, E2F targets, DNA repair, Glycolysis, and PI3K-AKT-mTOR signaling pathways and was highly active in biological processes related to cancer progression such as hypoxia, angiogenesis, and epithelial-mesenchymal transition (Supplementary Figure S2C). In contrast, SUMO-C1 was mainly enriched in tumor immune-related processes, such as TNFA signaling via NF-κB, IL6-STAT3 signaling, complement, inflammatory response, and IL2-STAT5 signaling, allograft rejection signaling pathway. This suggests that the SUMOylation pattern based on SUMOylation regulatory genes may influence the immune microenvironment status. Biological mechanisms for the poor prognosis associated with SUMO-C2 maybe because of the active immune microenvironment of the SUMO-C1 and the active stromal microenvironment of the SUMO-C2. SUMO-C1 may present better antitumor effects and better prognostic survival in lung adenocarcinoma through high activation of antitumor immunity and inhibition of the pro-cancer progressive stromal microenvironment.

3.3 Different SUMOylation patterns are associated with varying immune microenvironments

To explore the potential mechanisms by which SUMOylation patterns regulate the immune microenvironment, we used the ssGSEA algorithm to calculate the abundance of 24 different immune cells. A heatmap was used for visualizing the difference in immune cell infiltration (Figure 4A), and the Wilcoxon rank-sum test was used for comparing the discrepancy in immune cell infiltration consistent with SUMO-C1 and SUMO-C2 patterns (Figure 4B). Notably, SUMO-C1 showed a widespread active state of immune cells, while SUMO-C2 showed an inactive state of immune cells. The infiltration of B cells, T cells, T helper cells, Central Memory T cell (Tcm), Effector memory T cells (Tem), Follicular helper T cell (TFH), Th17 cells, CD8 T cells, cytotoxic cells, dendritic cells (DCs), and mast cells were significantly higher in the SUMO-C1 than SUMO-C2 group. In contrast, immune cells such as Th2 cells and Treg cells were significantly activated in SUMO-C2. Thereafter, we calculated immune cell infiltration levels in lung adenocarcinoma using the EPIC algorithm and TIMER algorithm to further support the differences in immune cell infiltration levels and compared the significance of the differences between the SUMO-C1 and SUMO-C2 groups using the Wilcoxon rank-sum test. We found that the SUMO-C1 group exhibited higher levels of immune cell infiltration, particularly CD8 T cells, which has been reported to have a killing effect on tumor cells thereby producing protective immunity against tumors (Supplementary Figures S3A, S3B). Furthermore, Spearman correlation between each SUMOylation regulatory gene and immune cell revealed that *CAPN3* was widely associated with various immune cells such as CD8+T cells (r = 0.32), DCs (r = 0.343), iDCs (r = 0.432),

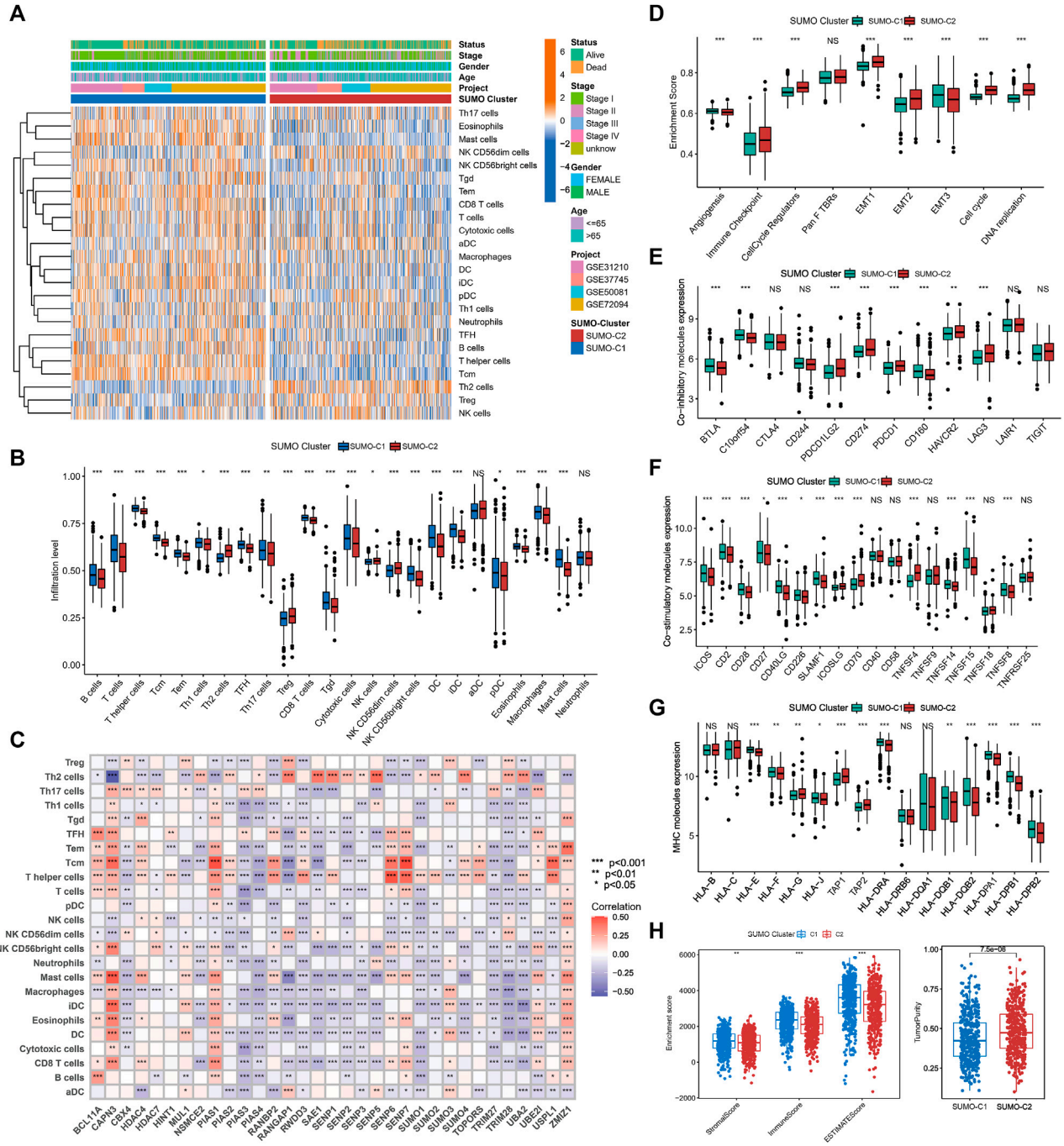


FIGURE 4
The tumor microenvironment differs significantly between SUMOylation patterns. **(A)** The heatmap shows the GSVA scores of 24 immune cells among two SUMOylation patterns in lung adenocarcinoma. Annotated with Status, Smoking History, Stage, Gender, Age, GEO cohort, and SUMOylation patterns. Orange is high expression and blue is low expression. **(B)** Wilcoxon test for differences in GSVA scores of 24 immune cells in two SUMOylation models. **(C)** Correlation between 33 SUMOylation regulatory genes and immune infiltrating cells using Spearman analysis. **(D)** Differences in oncogenic matrix microenvironment between two SUMOylation patterns. **(E)** Significantly differential expression of co-inhibitory molecules between the two SUMOylation patterns. **(F)** Differential expression of co-stimulatory molecules and **(G)** MHC molecules between two SUMOyaltion patterns. **(H)** Wilcoxon test for differences in immune scores, stromal scores, and tumor purity between the two SUMOylation models.

Mast cells (r = 0.468), Tcm (r = 0.421), Tem (r = 0.275), TFH (r = 0.291), Th2 cells (r = 0.563), and Treg (r = 0.115), indicating that *CAPN3* among SUMOylation regulatory genes is closely associated with immune cell infiltration in the tumor microenvironment of lung adenocarcinoma. Overall, the immune microenvironment was regulated by SUMOylation regulatory genes in lung adenocarcinoma mainly by modulating the levels of immune cells such as CD8+T cells,

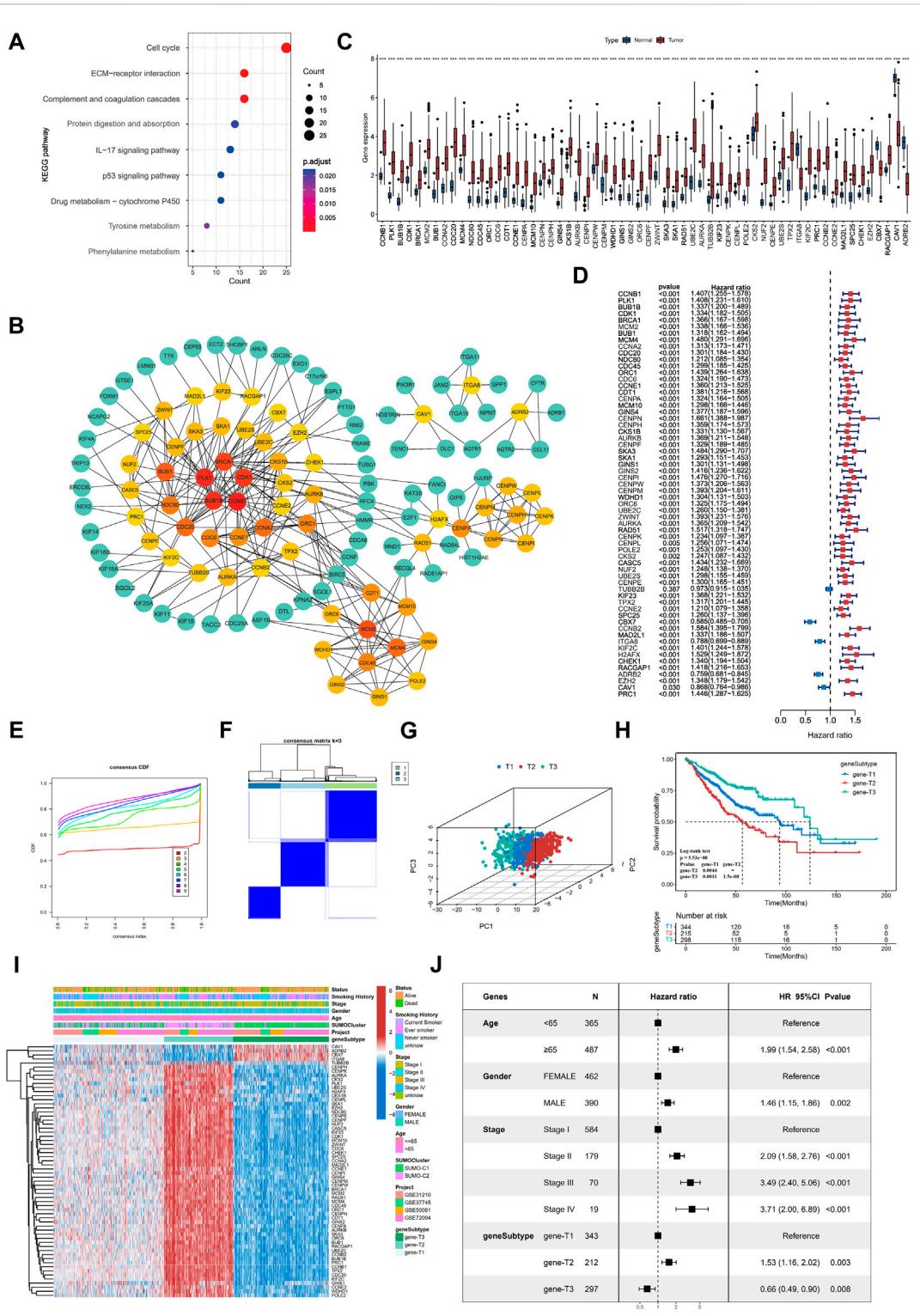


FIGURE 5
Identification of key genes differentially expressed between SUMOylation patterns and characterization of SUMOylation genomic subtypes. **(A)** Functional annotation of SUMOylation differential genes by KEGG enrichment analysis. **(B)** The Protein-Protein Interaction Network (PPI) of key genes associated with SUMOylation. **(C)** Differential expression of SUMOylation key genes between lung adenocarcinoma and normal tissue samples. **(D)** Prognostic value analysis of SUMOylation key genes for lung adenocarcinoma. **(E)** CDF curves for unsupervised clustering with K = 2 to 9. **(F)** Consensus clustering graph with K = 3. **(G)** PCA of SUMOylation genomic subtypes based on key genes transcriptome. **(H)** Kaplan-Meier curves for overall survival OS for the three SUMOylation genomic subtypes. (Log-Rank test p = 5.53e-08, p-value less than 0.001 for both comparisons). **(I)** The heatmap shows the identification of 3 different genomic subtypes by unsupervised clustering based on 62 SUMOylation-associated key genes. Annotated with Status, Smoking History, Stage, Gender, Age, SUMOylation patterns, GEO cohort, and genomic subtypes. Red is high expression and blue is low expression. **(J)** Multivariate cox analysis indicates that SUMOylation genomic subtype can be an independent prognostic factor for lung adenocarcinoma.

mast cells, T helper cells, Tcm, Tem, and Th2 cells (Figure 4C; Supplementary Figures S3A).

The above analysis supports that SUMO-C1 represents a type of TME with an immune characteristic of tumor suppression. Nevertheless, SUMO-C2 exhibited predominantly pro-cancer biological processes, as well as promoted the stromal microenvironment according to our previous GSVA enrichment analysis of the Hallmark gene set. Furthermore, we assessed the enrichment scores for angiogenesis, immune checkpoint, cell cycle regulation, Pan F TBRS, EMT, and DNA replication using the ssGSEA method. The ssGSEA results showed pathways that can promote immune escape such as EMT1, EMT2, and immune checkpoint pathways were remarkably activated in the SUMO-C2 group (Figure 4D). Moreover, the SUMO-C1 group showed a higher expression of co-stimulatory and MHC molecules and lower expression of co-inhibitory molecules (Figures 4E–G), such as CD274, PDCD1, and LAG3, which are now commonly used as immunotherapeutic checkpoints in lung adenocarcinoma. Finally, we assessed the immune score, stroma score, and tumor purity score using the Estimate package for lung adenocarcinoma expression data and found that SUMO-C1 had a better immune score as well as a higher stroma score (Figure 4H), while SUMO-C1 had a lower tumor purity score, which further explains its association with better prognosis. Altogether, the above analysis demonstrated that the SUMO-C2 SUMOylation pattern promoted the development of a silent anti-tumor immune microenvironment and an active tumorigenic stromal microenvironment.

3.4 PPI of SUMOylation pattern-associated genes (SPAGs) and identification of SUMOylation genomic subtypes

To further explore the underlying biological processes modulated in different SUMOylation patterns, we identified 906 significant DEGs (|FoldChange|>1.5, FDR<0.05) as SUMOylation pattern-associated genes (SPAGs) using the R package limma. These DEGs were also significantly associated with lung adenocarcinoma prognosis, indicating the critical role of SUMOylation patterns in dictating prognosis. KEGG functional annotation showed significant enrichment of SUMOylation-related genes to biological pathways such as cell cycle-related processes (cell cycle), cell adhesion (ECM-receptor interaction), immune activation signaling pathways (complement and coagulation cascades, IL 17 signaling pathway) (Figure 5A), while GO functional annotations revealed enrichment of biological processes such as regulation of the cell cycle phase transition, cell growth, humoral immune response, cell-cell junction and response to hypoxia (Supplementary Figures S4A–C).

Next, we constructed a protein interaction network (PPI, network type = physical subnetwork, minimum required interaction score = 0.7) of SPAGs using the STRING database and calculated the relationship between nodes using the Cytohubba plugin. We then classified genes with neighboring nodes ≥5 of 62 genes as key SPAGs and mapped their interaction (Figure 5B; Supplementary Figure S4D). Of the key genes, *CCNB1* had the most neighboring nodes, followed by *PLK1*, *BUB1B*, *CDK1*, and *BRCA1*. It has been recently reported that the key genes in the PPI network can represent the functional characteristics of all genes accurately. Differential expression analysis by the TCGA dataset showed that key genes were significantly differentially expressed in lung adenocarcinoma from normal tissue samples (Figure 5C). Meanwhile, key genes extensively indicated significant prognostic significance, and CBX7, ITGA8, ADRB2, and CAV1 were significant favorable factors for lung adenocarcinoma (Figure 5D).

Unsupervised consensus clustering was carried out based on the identified key SPAGs. Three genomic subtypes were obtained, namely, gene-T1, gene-T2, and gene-T3 by CDF curve (Figures 5E, F; Supplementary Figure S4E). PCA suggested significant differences among the three genomic subtypes identified based on key genes (Figure 5G). Kaplan–Meier survival analysis indicated that gene-T2 was associated with the worst, gene-T3 with the best, and gene-T1 with the intermediate prognostic outcome, indicating different clinical courses of the genomic subtypes (Figure 5H). Moreover, the heatmap revealed differential expression of the 62 key genes among the three genomic subtypes (Figure 5I). This is consistent with the results of the prognostic utility of key genes in lung adenocarcinoma, as the heatmap showed that *CAV1*, *ADRB2*, *CBX7*, and *ITGAB8* which were favorable factors for lung adenocarcinoma were remarkably upregulated in gene-T3, and the remaining risk factors for lung adenocarcinoma were significantly upregulated in the gene-T2 subtype. We then combined prognostic features such as age, sex, pathological stage, and genomic subtype with survival information for multicox analysis, which revealed that all the above prognostic features including genomic subtype are independent prognostic factors in lung adenocarcinoma; gene-T2 vs. gene-T1 (HR = 1.53, p = 0.003) and gene-T3 vs. gene- T1 (HR = 0.66, p = 0.008) (Figure 5J). Finally, we determined that SUMOylation regulatory genes were significantly differentially expressed among the three genomic subtypes and their expression in the gene-T1 subtype was intermediate to that in the gene-T2 and gene-T3 subtypes (Supplementary Figure S4F), indicating that different genomic subtypes further reflect the differences in the SUMOylation patterns.

3.5 Calculation and prognostic significance of the SUMOylation score

We aimed to determine the prognostic value and biological significance of SUMOylation modification patterns in lung adenocarcinoma and to develop a SUMOylation pattern-based prognostic model for the assessment of individual lung adenocarcinoma patients. For this, we first performed a univariate cox analysis of the 62 key SPAGs along with LASSO regression to screen the best variables. Eventually, eight SUMOylation key genes that were significantly differentially expressed between normal tissues and tissues in lung adenocarcinoma and significantly associated with prognosis in lung adenocarcinoma were screened, and a SUMOylation-based prognostic signature model was constructed based on the expression of these genes and their coefficients (Supplementary

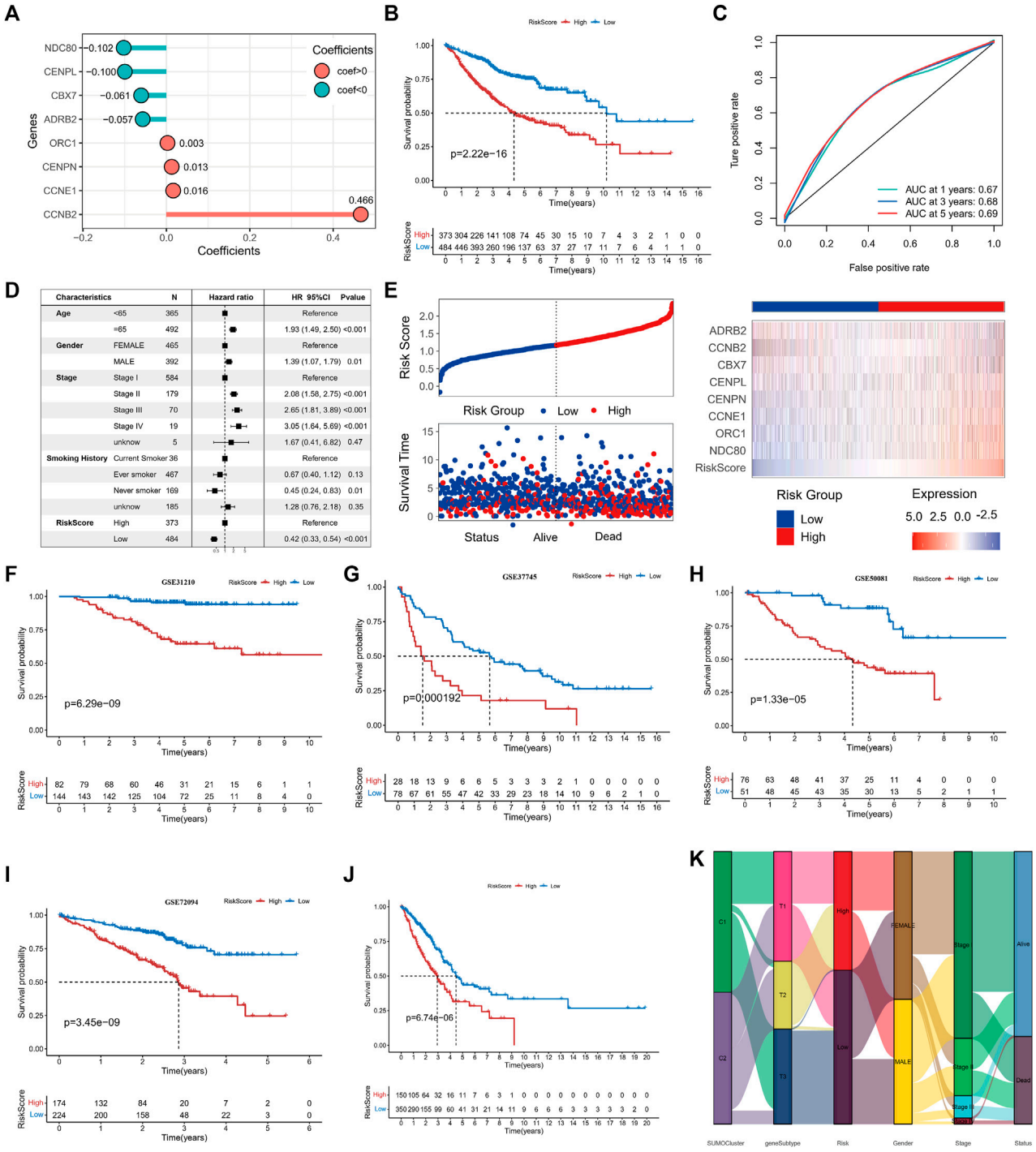


FIGURE 6
Construction of SUMOylation score and validation of prognostic value. **(A)** LASSO regression coefficients of model genes. **(B)** Kaplan-Meier curves for overall survival (OS) in lung adenocarcinoma patients between high- and low-SUMOylation score. **(C)** Operating characteristic curve for assessing the predictive performance of SUMOylation score for OS of lung adenocarcinoma, with AUCs of 0.67, 0.68, and 0.69 at 1, 3, and 5 years, respectively. **(D)** Multivariate cox analysis indicates that the SUMOylation score can work as an independent prognostic factor for lung adenocarcinoma. **(E)** The risk heatmap illustrates the variations in SUMOylation scores, patient deaths, and model gene expression levels. Red is high expression and blue is low expression. Validation of the prognostic significance of SUMOylation scores in independent datasets GSE31210 **(F)**, GSE37745 **(G)**, GSE50081 **(H)**, GSE72094 **(I)**. **(J)** Validation of significant differences in overall survival (OS) between patients with high- and low SUMOylation scores using the TCGA external dataset. **(K)** The alluvial map reveals the association between SUMOylation patterns, SUMOylation genomic subtypes, SUMOylation score groups, and other clinicopathological prognostic features.

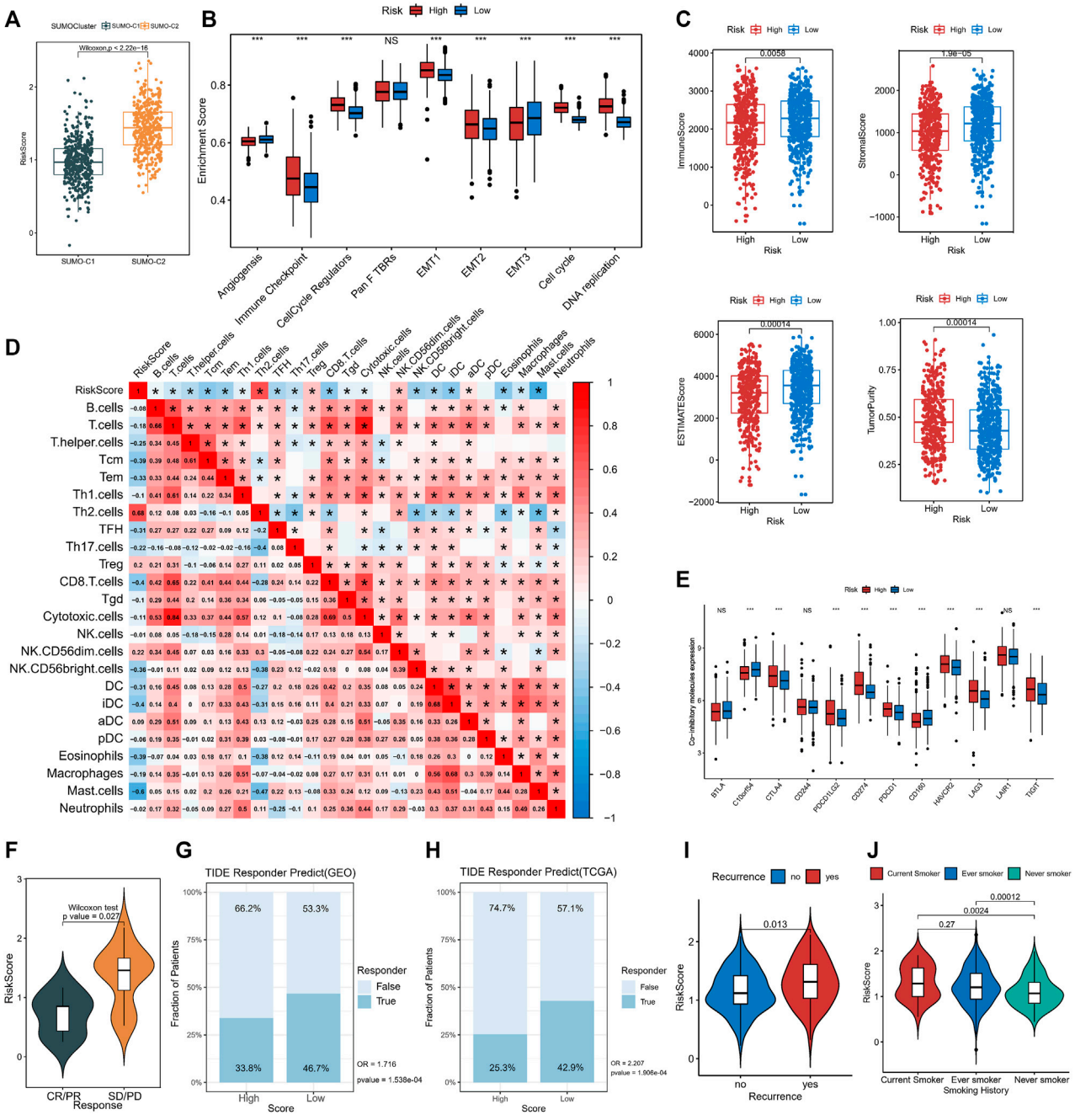


FIGURE 7
The association of SUMOylation score with tumor microenvironment and clinical characteristics. **(A)** The Wilcoxon test for significant differences in SUMOylation scores between the two SUMOylation patterns. **(B)** Oncogenic matrix microenvironment differences between high- and low-SUMOylation score. **(C)** Significant differences in ImmuneScore, StromalScore, ESTIMATEScore, and TumorPurity between different SUMOylation scores. **(D)** The correlation between SUMOylation score and immune cell infiltration. **(E)** Analysis of differential expression of co-inhibitory molecules between high- and low-SUMOylation score. **(F)** Comparison of SUMOylation scores between groups with and without immunotherapy response. Prediction of immunotherapy response in high and low-risk groups in GEO cohort **(G)** and TCGA cohort **(H)** based on TIDE algorithm. Association of lung adenocarcinoma recurrence **(I)** and smoking history **(J)** with SUMOylation score.

Figures S5A, B). The risk score was calculated formulas follows: risk score = Exp*NDC80**(−0.1024) + Exp*ORC1**0.0027+ Exp*CCNE1**0.0164+ Exp*CENPN**0.0127+ Exp*CENPL**(−0.0997) + Exp*CBX7**(−0.0606) + Exp*CCNB2**0.4661+ Exp*ADRB2**(−0.0568) (Figure 6A). All lung adenocarcinoma samples were divided into high-risk and low-risk groups using the Survival package to calculate the optimal cut-off values. Survival analysis revealed that patients with higher scores were significantly associated with worse prognosis (log-rank p = 2.22e-16) (Figure 6B); the 1-year ROC AUC was 0.67, 3-year AUC was 0.68, and 5-year AUC was 0.69, indicating that the risk score based on the SUMOylation model

had relatively superior prognostic accuracy compared to the vast array of existing prognostic models (Figure 6C). Multivariate Cox regression analysis incorporating age, gender, pathological stage, smoking history, and risk score confirmed that the risk score was an independent prognostic factor for assessing patient outcomes. (low vs. high, HR = 0.42 (0.33, 0.54), p < 0.001) (Figure 6D). ADRB2 and CBX7 expression were significantly negatively correlated with the risk score (Figure 6E). More deaths were observed in the high-risk group.

Four datasets, GSE31210 (Figure 6F), GSE37745 (Figure 6G), GSE50081 (Figure 6H), and GSE72094 (Figure 6I), and the external dataset TCGA-LUAD (Figure 6J) were used to further validate the stability and prognostic value of SUMOylation scores. The results of survival analysis gave *p*-values <0.001 for log-rank tests in all data sets, suggesting that the SUMOylation model risk score can consistently forecast patient prognostic outcomes; the ROC analysis also showed a high diagnostic value of the risk score in predicting lung adenocarcinoma prognosis (Supplementary Figures S5C–G). Multivariate cox regression analysis of TCGA lung adenocarcinoma cohort incorporating various prognostic factors such as age further confirms that the risk score is an independent risk factor for lung adenocarcinoma (low vs. high, HR = 0.52 (0.38, 0.71), $p < 0.001$) (Supplementary Figure S5H). To better understand the direct relationship between SUMOylation patterns and genomic subtypes, we constructed a Sankey diagram and observed that the vast majority of SUMO-C2 and a small fraction of SUMO-C1 comprise genomic subtype-T1 and then almost all of them comprise the Low Riskscore group (Figure 6K).

3.6 Experimental validation for genes of risk model

To verify the expression levels of model genes, we collected tumor tissues from 18 pairs of lung adenocarcinoma patients and the corresponding normal tissues adjacent to the cancer and performed quantitative real-time PCR. We also analyzed the expression levels and prognostic value of the model genes using the TCGA lung adenocarcinoma cohort. NDC80, CENPL, ORC1, CENPN, CCNE1, and CCNB2 were significantly upregulated in the TCGA cohort as well as in the 18 pairs of lung adenocarcinoma patient samples, and high expression was significantly associated with poor prognosis in lung adenocarcinoma. However, CBX7 and ADRB2 were significantly downregulated in lung adenocarcinoma tumor tissues and may serve as favorable prognostic factors for lung adenocarcinoma (Supplementary Figures S6A–H, $p < 0.05$).

SUMOylation score indicates immune microenvironment status and immunotherapy response in lung adenocarcinoma.

Comparison by the Wilcoxon rank-sum test showed that SUMO-C2 was associated with a higher risk score (Figure 7A), gene-T1 with an intermediate, gene-T2 with the highest, and gene-T3 with the lowest risk score (Supplementary Figure S7A), implying that differences in high and low-risk scores could reflect prognostic differences attributable to different SUMOylation patterns as well as different genomic subtypes. To further determine the relationship between SUMOylation scores and immune infiltration and other biological processes, the differences in the tumorigenic microenvironment between the groups with high- and low-risk scores were investigated. Immune checkpoint, cell cycle regulators, EMT1, EMT2, the cell cycle, and DNA replication were significantly active in the high-risk group (Figure 7B). Assessment of the level of immune infiltration in the Estimate package revealed significantly lower immune scores and higher tumor purity in the high-risk than in the low-risk group (Figure 7C). Moreover, Spearman correlation analysis exhibited a significant negative correlation between SUMOylation score and immune score, and a significant positive correlation between SUMOylation score and tumor purity (Supplementary Figure S7B), reinforcing the accuracy of the assessment of active biological processes in different SUMOylation patterns through risk scores.

Subsequently, Spearman analysis was then used to investigate the correlation between risk score and different immune cell subpopulations, and risk score was overall significantly negatively correlated with the levels of various immune cells. The risk score was significantly positively correlated with Th2 cells (r = 0.68) and Treg cells (r = 0.2)., and was negatively correlated with immune cells such as CD8+T cells (r = −0.4), Tem (r = −0.33), Tcm (r = −0.39), TFH (r = −0.31), CD56 (bright) natural killer cells (r = −0.36), DCs (r = −0.31), (immature Dendritic Cells)iDC (r = −0.4) and mast cells (r = −0.6). In addition, there was a close correlation between the degree of infiltration of Th1 cells, $CD8^+$ T cells, T cells, and cytotoxic cells, as well as that of macrophages and iDCs in lung adenocarcinoma. (Figure 7D). These findings suggest that SUMOylation regulatory genes are involved in immune cell infiltration and immune regulation of lung adenocarcinoma tumors. Of these, Th2, Treg, and $CD8^+$ T cells, and mast cells are involved in immune dynamic regulation (Supplementary Figure S7C). We then assessed immune checkpoint expression between the high- and low-risk groups to further investigate the effect of SUMOylation on immunotherapy response and found that the expression of clinically common immune checkpoint co-inhibitory molecules such as CD274, PDCD1, CTLA4, PDCD1LG2, HAVCR2, LAG3, and TIGIT was increased in the high-risk group (Figure 7E), while co-stimulatory molecules, such as CD2, CD28, and CD40LG, were decreased in the high-risk group (Supplementary Figure S7D), indicating that SUMOylation scores broadly modulate immune checkpoint expression. In addition, SUMOylation also suppresses the activation of the immune microenvironment by reducing the expression of MHC molecules (Supplementary Figure S7E).

A subsequent analysis of the role of SUMOylation patterns on immunotherapy response using the immunotherapy cohort reported in the lung cancer dataset GSE126044 indicated that the responder group (CR/PR) had low-risk scores and the non-responder group had high-risk scores (Figure 7F), explaining the worse immunotherapy response in high-risk patients. Meanwhile, we evaluated the immunotherapy response of lung adenocarcinoma patients in both high and low SUMOylation score groups in the GEO cohort (High-risk vs. Low-risk, OR is 1.716, *p*-value < 0.001, Figure 7G) and TCGA cohort (High-risk vs. Low-risk, OR is 2.207, p-value < 0.001, Figure 7H) based on

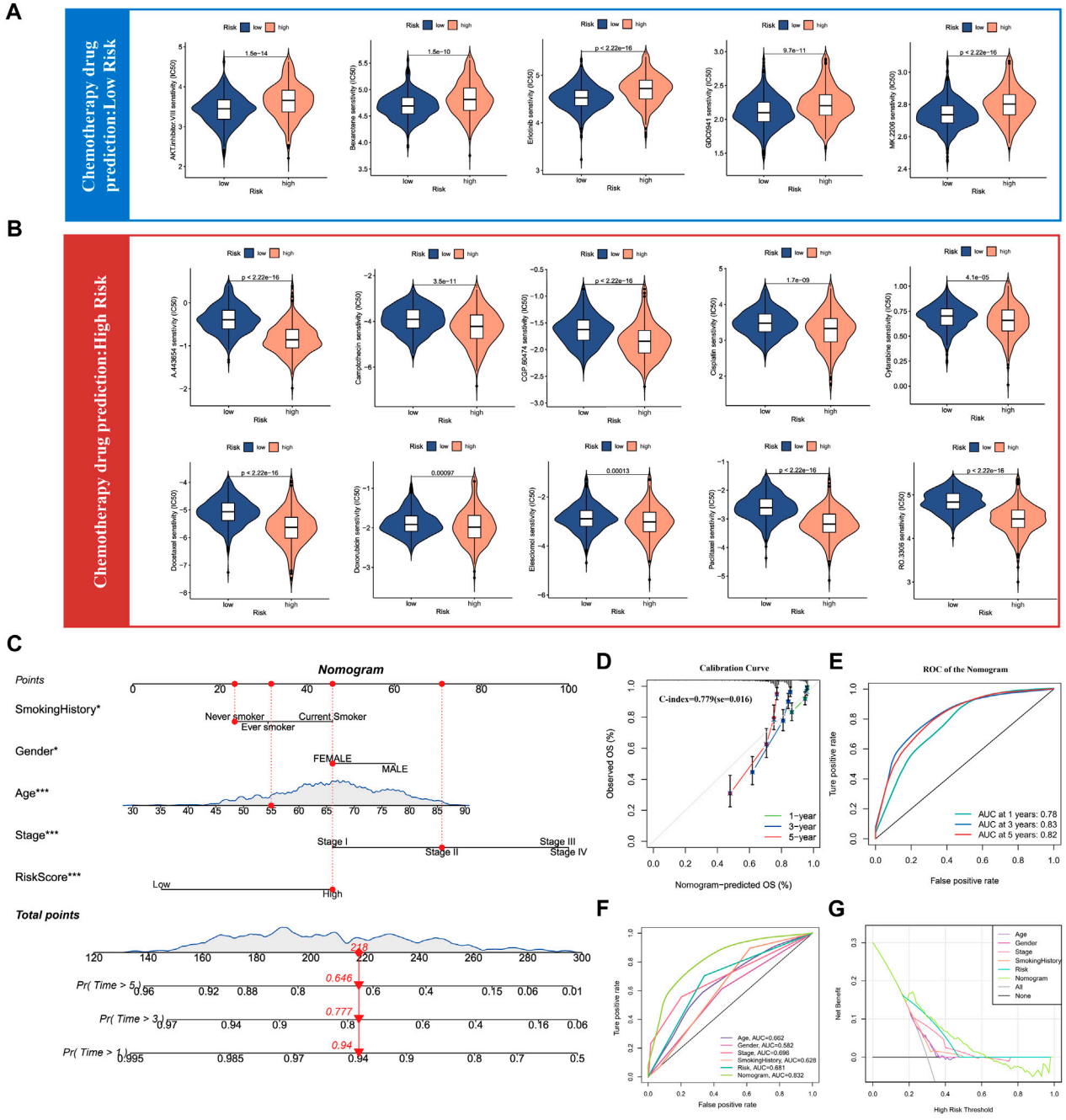


FIGURE 8
Drug sensitivity analysis of chemotherapy for lung adenocarcinoma and Nomogram construction. **(A)** Chemotherapy drug sensitivity analysis in patients with low SUMOylation score. **(B)** Chemotherapy drug sensitivity analysis in patients with high SUMOylation score. **(C)** Construction of Nomogram by combining independent prognostic features such as smoking history, gender, age, pathological stage, and SUMOylation score. **(D)** Calibration curves for evaluating the predictive performance of Nomogram for 1, 3, and 5-year OS in lung adenocarcinoma patients. **(E)** The AUC for assessing the accuracy of Nomogram to forecast 1-, 3-, and 5- years overall survival. **(F)** The AUC for predicting 3-year overall survival for comparing the accuracy of various prognostic characteristics. **(G)** Decision Curve Analysis (DCA) shows the efficacy of Nomogram and other prognostic features for clinical applications.

the TIDE algorithm and indicated that low-risk patients were more likely to benefit from immunotherapy. In addition to this, significantly high SUMOylation scores are observed in patients with recurrent lung adenocarcinoma (Figure 7I), and significantly high SUMOylation scores are observed in smokers (Figure 7J), suggesting that SUMOylation scores may be effective in comprehensively assessing clinical prognostic factors in patients with lung adenocarcinoma.

3.7 Prediction of chemotherapy drug sensitivity and construction of nomogram for lung adenocarcinoma

As the above results have confirmed that SUMOylation leads to remarkable changes in the tumor microenvironment in lung adenocarcinoma, therefore, we used the SUMOylation score to characterize the sensitivity of lung adenocarcinoma patients to various chemotherapeutic agents to guide the combination chemotherapy. Based on the mRNA levels in lung adenocarcinoma tissues, we calculated the IC50 of various chemotherapeutic agents for lung adenocarcinoma patients in different SUMOylation score groups using the pRRophetic package and found that the group with low SUMOylation score was more sensitive to AKT-inhibitor-VIII, bexarotene, erlotinib, GDC0941, MK-2206, among others than that with high SUMOylation score (Figure 8A). In contrast, the high SUMOylation score group was sensitive to more chemotherapeutic agents, including A-443654, camptothecin, CGP-60474, cisplatin, cytarabine, docetaxel, doxorubicin, elesclomol, paclitaxel, and RO-3306 (Figure 8B). These findings suggest that SUMOylation scores can accurately reflect significant biological differences between SUMOylation patterns and also correlate with clinical prognostic features.

Meanwhile, as we previously illustrated by multivariate Cox analysis that age, gender, pathological stage, smoking history, and SUMOylation score all can serve as independent prognostic factors for lung adenocarcinoma, we combined the above significant prognostic factors to construct column plots to more accurately and efficiently predict the overall survival (OS) of individual patients with lung adenocarcinoma (Figure 8C). Calibration curves showed that the OS at 1, 3, and 5 years predicted by the nomogram closely corresponded to the actual OS of lung adenocarcinoma patients (Figure 8D) and the time-dependent ROC showed high AUC of 0.78, 0.83, and 0.82 for 1-, 3-, and 5- years, respectively. Yang et al. reported that 0.9≥ AUC>0.8 indicates excellent discrimination (Shengping and Gilbert, 2017), indicating that the developed nomogram has a remarkably prognostic performance (Figure 8E). Furthermore, a comparison of the predictive performance of multiple clinical prognostic features revealed that the nomogram offered the best predictive utility relative to any other clinical feature as determined by the ROC curve (Figure 8F) as well as the DCA decision curve (Figure 8G). Finally, to further validate the accuracy of the multifactorial regression model Nomogram, we evaluated the AUC of the ROC curves of Nomogram using three independent lung adenocarcinoma cohorts and the 5-year AUC values were 0.79, 0.81, and 0.89, respectively (Supplementary Figures S7F–H7). Validation in three independent lung adenocarcinoma cohorts supports that the accuracy of our Nomogram is significantly better than a large number of reported prognostic models for lung adenocarcinoma (Mo et al., 2020; Song et al., 2020; Chen et al., 2021).

4 Discussion

SUMOylation performs a critical role in various biological processes encompassing immune regulation (K et al., 2021). Due to the extremely complex TME and immune background of lung adenocarcinoma, the modulatory role of SUMOylation on the TME of lung adenocarcinoma, especially on immune infiltration, is still poorly understood. Current studies are limited to individual SUMOylation regulatory genes, and the regulation of the TME mediated by integrated SUMOylation regulatory genes has not been investigated. Therefore, the identification of different SUMOylation patterns in the TME is valuable in further understanding the effect of SUMOylation on tumor immune responses in lung adenocarcinoma.

Previous studies have shown that regulation of the TME has a critical role in tumor progression and immunotherapeutic efficacy (Goliwas et al., 2021). In this study, we confirmed the tight correlation of SUMOylation regulatory genes and identified two distinct SUMOylation patterns associated with significant differences in the TME in terms of differential activation of oncogenic pathways and immune infiltration. Park S et al. classified Non-small cell lung cancer (NSCLC) into three different immunophenotypes based on tumor-infiltrating lymphocytes and immune checkpoint treatment response. Assessment of immune phenotypes is useful as a guide for determining prognosis as well as immunotherapy (Park et al., 2022). The SUMOylation patterns we identified also corresponded to distinct immunophenotypes. SUMO-C1 corresponded to a tumor immunoinflammatory phenotype with better overall survival, significant activation of immune pathways such as the IL2, inflammatory, complement, and allograft rejection pathway, and increased infiltration levels of anti-tumor immune cells. IL2 promotes not only the proliferation of cytotoxic T lymphocytes (CTLs) and natural killer cells but also the differentiation of CTLs to effector T cells (Borrelli et al., 2018). Meanwhile, CD8+T cells, Tem, cytotoxic cells, and Th17 cells were reported to promote anti-tumor immune processes (Deng et al., 2018). SUMO-C2 was consistent with reduced levels of immune infiltration and a microenvironment that promotes immune escape, corresponding to an immune-desert phenotype. Furthermore, a remarkable activation of the PI3K-AKT-mTOR signaling pathway, the cell cycle, EMT substrates, and oncogenic pathways suggested that SUMO-C2 also exhibited features of an immune exclusion phenotype. Notably, the levels of infiltration of Th2 cells, Treg cells, $CD8^+$ T cells, and Th17 cells varied with different SUMOylation patterns and SUMOylation scores. SUMOylation regulatory genes are diverse in their regulation of immune cells, and the regulatory mechanism of SUMOylation regulatory genes on lung adenocarcinoma immune cells needs to be further explored. Therefore, the identification of SUMOylation patterns could help determine immunotherapy response and patient prognoses.

Notably, SUMO-C2 showed significant activation of PI3K-AKT-mTOR pathway in TME, and excessive activation of the PI3K-AKT-mTOR pathway would lead to a combined phenotype of immunodeficiency and immune dysregulation (Nunes-Santos et al., 2019), which is critical for maintaining the immunosuppressive function of Tregs and (Myeloid-derived suppressor cells)MDSCs, while inhibition of the PI3K-AKT-mTOR pathway can reduce the expression of immunosuppressive factors as well as immune checkpoint ligands (O'Donnell et al., 2018). Furthermore, we found that SUMO-C2 was associated with the increased activation of the EMT pathway and increased expression of multiple immune checkpoint molecules such as

PD1. Jiang Y et al. showed that the overactivation of EMT was associated with the activation of different immune checkpoint molecules and induced tumor immune escape (Jiang and Zhan, 2020). Importantly, targeted inhibition of the PI3K-AKT-mTOR pathway maintains the antitumor immune function of $CD8^+$ T cells (Patton et al., 2006), while targeted inhibition of EMT pathways has been shown to remodel the TME and restore the antitumor microenvironment (Erin et al., 2020; Kumagai et al., 2020), which suggests that targeted inhibition of the above pathways combined with immunotherapy may be a more reasonable treatment option for patients with the SUMO-C2 SUMOylation pattern.

Further, we predicted potential target drugs for high SUMOylation scores. AKT inhibitors such as AKT-inhibitor-VIII, MK-2206 (O'Donnell et al., 2018), A-443654 (Han et al., 2007), and PI3K inhibitors such as GDC-0941 (Haagensen et al., 2012) can inhibit the PI3K-AKT-mTOR signaling pathway, thereby inhibiting tumor progression. Meanwhile, CGP-60474 (Han et al., 2018), RO-3306 (Kojima et al., 2009), and doxorubicin (Jin et al., 2020) inhibit the cell cycle and DNA replication to suppress tumor progression. However, the relationship between these drugs and SUMOylation and their role in lung adenocarcinoma progression is not yet known. In addition, camptothecin (Sanchez-Alcazar et al., 2003), bexarotene (Lowe and Plosker, 2000), cisplatin (Ghosh, 2019), and paclitaxel (Wang et al., 2000) can also exert apoptosis-inducing antitumor effects, and docetaxel (Borghaei et al., 2021) has been widely used in combination with nivolumab in the treatment of NSCLC. This suggests that the SUMOylation model and score can better guide chemotherapy as well as immunotherapy for lung adenocarcinoma.

Based on two SUMOylation patterns that were associated with different TMEs, due to the complexity of SUMOylation pattern SUMO-C2 immunophenotype, we mined SUMOylation pattern-related genes (SPAGs) and further determined the genomic subtype gene-T1 corresponding to immune exclusion phenotype, which would further define SUMOylation patterns as well as TME immune phenotypes. A m6A methylation modification score can accurately determine colorectal cancer TME and immune transcripts. High m6A scores are tightly associated with worse prognosis, inferior levels of antitumor immune infiltration, and poorer response to immunotherapy (Chong et al., 2021). To determine the TME status and more precisely guide the treatment of individual patients, we present a highly sensitive prognostic model based on SUMOylation scoring, which can serve as a prognostic biomarker for lung adenocarcinoma. Noteworthy, a large number of scoring systems have been previously reported for prognostic assessment of lung adenocarcinoma, Unlike the scoring systems constructed with the DNA repair-related prognostic signature (Chang and Lai, 2019), the cell cycle-related prognostic signature (Chen et al., 2021), immune-related signature (Song et al., 2020), and the angiogenesis-related signature (Qing et al., 2022). Our model has relatively superior predictive performance compared to other signature models and can be further used as a complement to clinical factors. Interestingly, SUMOylation scores were significantly negatively correlated with immune infiltration. The immunoinflammatory phenotype consistent with SUMO-C1 exhibited a lower SUMOylation score, while the immune-desert phenotype consistent with SUMO-C2 or gene-T3 showed the highest SUMOylation score. This implies that SUMOylation can modulate the TME as well as the immune background of lung adenocarcinoma, thus regulating the progression and prognosis of lung adenocarcinoma patients.

In summary, our study identified two different SUMOylation patterns mapping different immune phenotypes and TMEs, establishing that SUMOylation can regulate the lung adenocarcinoma TME and the infiltration of immune cells. The SUMOylation score we constructed can facilitate a more accurate assessment of the TME in lung adenocarcinoma, and the nomogram based on it has good clinical utility. As samples with high SUMOylation scores showed stronger immunosuppression and immune escape, as well as promoted the activation of pro-cancer pathways, compared with low scores, we further explored two SUMOylation patterns of potential targeting agents to guide chemical combination therapy. However, our study is limited in that we relied on a few immunotherapy response cohorts with NSCLC to confirm our findings, hence, our findings must be validated in larger cohorts. Furthermore, the clinical utility of the SUMOylation score and the constructed nomogram needs to be validated in clinical settings. In addition, the characterization of the underlying mechanisms by which SUMOylation regulates the tumor microenvironment and immune background in lung adenocarcinoma and other tumors warrants further research.

5 Conclusion

The SUMOylation patterns can well dictate the tumor microenvironment features, particularly immune cell infiltration status, in lung adenocarcinoma. The SUMOylaiton score is indicative of the relationship between SUMOylation and immune cell crosstalk and has remarkable prognostic value, and can be used to predict immunotherapy and chemotherapy response in lung adenocarcinoma. In conclusion, the SUMOylation model and score have a high value for determining the tumor microenvironment status and prognosis prediction.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE31210 https://www.ncbi.nlm.nih.gov/geo/, GSE37745 https://www.ncbi.nlm.nih.gov/geo/, GSE50081 https://www.ncbi.nlm.nih.gov/geo/, GSE72094.

Author contributions

Conception and data analysis of the paper: ZC, LT, and JY. Drafting of the manuscript: ZC and JY. Data collection and assistants in data analysis: XS and YL. Data visualization: ZS and HD. Article Revision: CX and XT. Funding and Supervising: JZ All authors participated in the work and agreed to take responsibility for all aspects of the work, thus ensuring that the research was accurate and that relevant issues were properly investigated and resolved.

Funding

This work was supported by the grants from Natural Science Foundation of Jiangsu province: BK20220250; Suzhou science and Technology Bureau (LCZX2019002).

Acknowledgments

We thank Bullet Edits Limited for the linguistic editing and the public databases for their data support. We are very grateful to Professor Jun Zhao for his valuable guidance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2023.1094588/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Mutual association and functional annotation of 33 SUMOylation regulatory genes. GO functional annotation **(A-B)** and KEGG functional annotation **(C)** of SUMOylarion regulatory genes. **(D)** Analysis of the correlation between SUMOylation regulatory genes in lung adenocarcinoma. **(E)** Principal component analysis of SUMOylation regulatory gene expression profiles based on 57 paired lung adenocarcinoma and normal samples. **(F)** Co-mutation between regulatory genes of SUMOylation. **(G)** Multivariate cox analysis for prognostic significance among SUMOylation regulatory genes.

SUPPLEMENTARY FIGURE S2
Unsupervised clustering based on SUMOylation regulatory genes and differences in biological pathways between clusters. **(A)** Consensus clustering graph for K = 3 to 6. **(B)** The heatmap shows the mRNA expression levels of SUMOylation regulatory genes in two SUMOylation patterns. **(C)** Comparison of HALLMARK biological pathways between two SUMOylation patterns.

SUPPLEMENTARY FIGURE S3
Relationship between SUMOylation regulatory genes with their patterns and immune microenvironment. Immune cell infiltration levels between SUMOylation patterns were assessed by **(A)** EPIC algorithm and **(B)**TIMER algorithm. **(C)** Correlation between SUMOylation regulatory genes and immune infiltrating cells.

SUPPLEMENTARY FIGURE S4
GO functional annotation and key genes analysis. Functional annotation of 906 SUMOylation-associated differential genes GO, **(A)** BP, **(B)** CC, **(C)** MF. SUMOylation-related differential genes are regarded as SUMOylation Pattern-Related Genes (SPAGs). **(D)** Computation of key genes in SPAGs by Cytohubba (Degree≥5). **(E)** Consensus clustering graph with K = 2,4–9. **(F)** Comparison of significant differences in SUMOylation regulatory gene expression among three SUMOylation genomic subtypes.

SUPPLEMENTARY FIGURE S5
Construction and validation of SUMOylation score. LASSO regression analysis of 61 significantly prognostically relevant key genes , **(A,B)** partial Likelihood Deviance = 13.77, SE = 0.1393, Lambda = 0.01725. Validation of the accuracy of SUMOylation scores to predict OS at 1-, 3-, and 5-year in the independent datasets GSE31210 **(C)**, GSE37745 **(D)**, GSE50081 **(E)**, GSE72094 **(F)**, and the external validation set TCGA-LUAD **(G)**, respectively. **(H)** Multivariate cox analysis based on prognostic characteristics such as SUMOylation score in the TCGA-LUAD cohort.

SUPPLEMENTARY FIGURE S6
Validation for model genes. mRNA expression differential analysis, survival analysis, and quantitative Real-time PCR for ADRB2, **(A)** CBX7 **(B)**, CCNB2 **(C)**, CCNE1 **(D)**, CENPL **(E)**, ORC1 **(F)**, NDC80 **(G)**, and CENPN **(H)**. (*$p < 0.05$, **$p < 0.01$).

SUPPLEMENTARY FIGURE S7
Relationship between SUMOylation score and immune microenvironment with other prognostic signatures. **(A)** Comparison of risk scores among genomic subtypes of SUMOylation. **(B)** Correlation of SUMOylation score with immune score and tumor purity. **(C)** Comparison of 24 immune cell GSVA scores between high- and low-SUMOylation scores. Comparison of co-stimulatory molecules **(D)** and MHC molecules **(E)** expression levels between high- and low-SUMOylation scores. Survival analysis and ROC of multivariate regression model Nomogram in independent lung adenocarcinoma cohorts with GSE31210**(F)**, GSE50081 **(G)**, GSE72094 **(H)**, the 5-year AUCs were 0.79, 0.77, and 0.89, respectively.

References

Anderson, N. M., and Simon, M. C. (2020). The tumor microenvironment. *Curr. Biol.* 30, R921–R925. doi:10.1016/j.cub.2020.06.081

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112. doi:10.1038/nature08460

Bawa-Khalfe, T., and Yeh, E. T. (2010). SUMO losing balance: SUMO proteases disrupt SUMO homeostasis to facilitate cancer development and progression. *Genes Cancer* 1, 748–752. doi:10.1177/1947601910382555

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Borghaei, H., Gettinger, S., Vokes, E. E., Chow, L. Q. M., Burgio, M. A., de Castro Carpeno, J., et al. (2021). Five-year outcomes from the randomized, phase III trials CheckMate 017 and 057: Nivolumab versus docetaxel in previously treated non-small-cell lung cancer. *J. Clin. Oncol.* 39, 723–733. doi:10.1200/JCO.20.01605

Borrelli, C., Ricci, B., Vulpis, E., Fionda, C., Ricciardi, M. R., Petrucci, M. T., et al. (2018). Drug-induced senescent multiple myeloma cells elicit NK cell proliferation by direct or exosome-mediated IL15 trans-presentation. *Cancer Immunol. Res.* 6, 860–869. doi:10.1158/2326-6066.CIR-17-0604

Botling, J., Edlund, K., Lohr, M., Hellwig, B., Holmberg, L., Lambe, M., et al. (2013). Biomarker discovery in non-small cell lung cancer: Integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin. Cancer Res.* 19, 194–204. doi:10.1158/1078-0432.CCR-12-1139

Chang, H. M., and Yeh, E. T. H. (2020). Sumo: From bench to bedside. *Physiol. Rev.* 100, 1599–1619. doi:10.1152/physrev.00025.2019

Chang, S. C., and Ding, J. L. (2018). Ubiquitination and SUMOylation in the chronic inflammatory tumor microenvironment. *Biochim. Biophys. Acta Rev. Cancer* 1870, 165–175. doi:10.1016/j.bbcan.2018.08.002

Chang, W. H., and Lai, A. G. (2019). Transcriptional landscape of DNA repair genes underpins a pan-cancer prognostic signature associated with cell cycle dysregulation and tumor hypoxia. *DNA Repair (Amst)* 78, 142–153. doi:10.1016/j.dnarep.2019.04.008

Chen, F., Song, J., Ye, Z., Xu, B., Cheng, H., Zhang, S., et al. (2021). Integrated analysis of cell cycle-related and immunity-related biomarker signatures to improve the prognosis prediction of lung adenocarcinoma. *Front. Oncol.* 11, 666826. doi:10.3389/fonc.2021.666826

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8 (4), S11. doi:10.1186/1752-0509-8-S4-S11Suppl

Cho, J. W., Hong, M. H., Ha, S. J., Kim, Y. J., Cho, B. C., Lee, I., et al. (2020). Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. *Exp. Mol. Med.* 52, 1550–1563. doi:10.1038/s12276-020-00493-8

Chong, W., Shang, L., Liu, J., Fang, Z., Du, F., Wu, H., et al. (2021). m(6)A regulator-based methylation modification patterns characterized by distinct tumor microenvironment immune profiles in colon cancer. *Theranostics* 11, 2201–2217. doi:10.7150/thno.52717

Deng, W., Lira, V., Hudson, T. E., Lemmens, E. E., Hanson, W. G., Flores, R., et al. (2018). Recombinant Listeria promotes tumor rejection by CD8(+) T cell-dependent remodeling of the tumor microenvironment. *Proc. Natl. Acad. Sci. U. S. A.* 115, 8179–8184. doi:10.1073/pnas.1801910115

Der, S. D., Sykes, J., Pintilie, M., Zhu, C. Q., Strumpf, D., Liu, N., et al. (2014). Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J. Thorac. Oncol.* 9, 59–64. doi:10.1097/JTO.0000000000000042

Eifler, K., and Vertegaal, A. C. O. (2015). SUMOylation-mediated regulation of cell cycle progression and cancer. *Trends Biochem. Sci.* 40, 779–793. doi:10.1016/j.tibs.2015.09.006

Erin, N., Grahovac, J., Brozovic, A., and Efferth, T. (2020). Tumor microenvironment and epithelial mesenchymal transition as targets to overcome tumor multidrug resistance. *Drug Resist Updat* 53, 100715. doi:10.1016/j.drup.2020.100715

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9, e107468. doi:10.1371/journal.pone.0107468

Ghosh, S. (2019). Cisplatin: The first metal based anticancer drug. *Bioorg Chem.* 88, 102925. doi:10.1016/j.bioorg.2019.102925

Goliwas, K. F., Deshane, J. S., Elmets, C. A., and Athar, M. (2021). Moving immune therapy forward targeting TME. *Physiol. Rev.* 101, 417–425. doi:10.1152/physrev.00008.2020

Haagensen, E. J., Kyle, S., Beale, G. S., Maxwell, R. J., and Newell, D. R. (2012). The synergistic interaction of MEK and PI3K inhibitors is modulated by mTOR inhibition. *Br. J. Cancer* 106, 1386–1394. doi:10.1038/bjc.2012.70

Han, E. K., Leverson, J. D., McGonigal, T., Shah, O. J., Woods, K. W., Hunter, T., et al. (2007). Akt inhibitor A-443654 induces rapid Akt Ser-473 phosphorylation independent of mTORC1 inhibition. *Oncogene* 26, 5655–5661. doi:10.1038/sj.onc.1210343

Han, H. W., Hahn, S., Jeong, H. Y., Jee, J. H., Nam, M. O., Kim, H. K., et al. (2018). LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent. *Sci. Rep.* 8, 14969. doi:10.1038/s41598-018-33039-0

Han, Y., Huang, C., Sun, X., Xiang, B., Wang, M., Yeh, E. T., et al. (2010). SENP3-mediated de-conjugation of SUMO2/3 from promyelocytic leukemia is correlated with accelerated cell proliferation under mild oxidative stress. *J. Biol. Chem.* 285, 12906–12915. doi:10.1074/jbc.M109.071431

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Horn, L., Spigel, D. R., Vokes, E. E., Holgado, E., Ready, N., Steins, M., et al. (2017). Nivolumab versus docetaxel in previously treated patients with advanced non-small-cell lung cancer: Two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). *J. Clin. Oncol.* 35, 3924–3933. doi:10.1200/JCO.2017.74.3062

Jiang, Y., and Zhan, H. (2020). Communication between EMT and PD-L1 signaling: New insights into tumor immune evasion. *Cancer Lett.* 468, 72–81. doi:10.1016/j.canlet.2019.10.013

Jin, Y., Wang, Y., Liu, X., Zhou, J., Wang, X., Feng, H., et al. (2020). Synergistic combination chemotherapy of lung cancer: Cisplatin and doxorubicin conjugated prodrug loaded, glutathione and pH sensitive nanocarriers. *Drug Des. Devel Ther.* 14, 5205–5215. doi:10.2147/DDDT.S260253

K, S. T., Joshi, G., Arya, P., Mahajan, V., Chaturvedi, A., and Mishra, R. K. (2021). SUMO and SUMOylation pathway at the forefront of host immune response. *Front. Cell Dev. Biol.* 9, 681057. doi:10.3389/fcell.2021.681057

Kessler, J. D., Kahle, K. T., Sun, T., Meerbrey, K. L., Schlabach, M. R., Schmitt, E. M., et al. (2012). A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* 335, 348–353. doi:10.1126/science.1212728

Kojima, K., Shimanuki, M., Shikami, M., Andreeff, M., and Nakakuma, H. (2009). Cyclin-dependent kinase 1 inhibitor RO-3306 enhances p53-mediated Bax activation and mitochondrial apoptosis in AML. *Cancer Sci.* 100, 1128–1136. doi:10.1111/j.1349-7006.2009.01150.x

Kumagai, S., Togashi, Y., Sakai, C., Kawazoe, A., Kawazu, M., Ueno, T., et al. (2020). An oncogenic alteration creates a microenvironment that promotes tumor progression by conferring a metabolic advantage to regulatory T cells. *Immunity* 53, 187–203.e8. doi:10.1016/j.immuni.2020.06.016

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). Timer: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77, e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Lightcap, E. S., Yu, P., Grossman, S., Song, K., Khattar, M., Xega, K., et al. (2021). A small-molecule SUMOylation inhibitor activates antitumor immune responses and potentiates immune therapies in preclinical models. *Sci. Transl. Med.* 13, eaba7791. doi:10.1126/scitranslmed.aba7791

Lowe, M. N., and Plosker, G. L. (2000). Bexarotene. *Am. J. Clin. Dermatol* 1, 245–250. doi:10.2165/00128071-200001040-00006; discussion 251-242

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548. doi:10.1038/nature25501III

Minguet, J., Smith, K. H., and Bramlage, P. (2016). Targeted therapies for treatment of non-small cell lung cancer--Recent advances and future perspectives. *Int. J. Cancer* 138, 2549–2561. doi:10.1002/ijc.29915

Mo, Z., Yu, L., Cao, Z., Hu, H., Luo, S., and Zhang, S. (2020). Identification of a hypoxia-associated signature for lung adenocarcinoma. *Front. Genet.* 11, 647. doi:10.3389/fgene.2020.00647

Nunes-Santos, C. J., Uzel, G., and Rosenzweig, S. D. (2019). PI3K pathway defects leading to immunodeficiency and immune dysregulation. *J. Allergy Clin. Immunol.* 143, 1676–1687. doi:10.1016/j.jaci.2019.03.017

O'Donnell, J. S., Massi, D., Teng, M. W. L., and Mandala, M. (2018). PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux. *Semin. Cancer Biol.* 48, 91–103. doi:10.1016/j.semcancer.2017.04.015

Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., et al. (2012). Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* 72, 100–111. doi:10.1158/0008-5472.CAN-11-1403

Park, S., Ock, C. Y., Kim, H., Pereira, S., Park, S., Ma, M., et al. (2022). Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. *J. Clin. Oncol.* 40, 1916–1928. doi:10.1200/JCO.21.02010

Patton, D. T., Garden, O. A., Pearce, W. P., Clough, L. E., Monk, C. R., Leung, E., et al. (2006). Cutting edge: The phosphoinositide 3-kinase p110 delta is critical for the function of CD4+CD25+Foxp3+ regulatory T cells. *J. Immunol.* 177, 6598–6602. doi:10.4049/jimmunol.177.10.6598

Qing, X., Xu, W., Liu, S., Chen, Z., Ye, C., and Zhang, Y. (2022). Molecular characteristics, clinical significance, and cancer immune interactions of angiogenesis-associated genes in gastric cancer. *Front. Immunol.* 13, 843077. doi:10.3389/fimmu.2022.843077

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 6, e26476. doi:10.7554/eLife.26476

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. doi:10.1016/j.cell.2014.12.033

Sanchez-Alcazar, J. A., Bradbury, D. A., Brea-Calvo, G., Navas, P., and Knox, A. J. (2003). Camptothecin-induced apoptosis in non-small cell lung cancer is independent of cyclooxygenase expression. *Apoptosis* 8, 639–647. doi:10.1023/A:1026147812000

Schabath, M. B., Welsh, E. A., Fulp, W. J., Chen, L., Teer, J. K., Thompson, Z. J., et al. (2016). Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* 35, 3209–3216. doi:10.1038/onc.2015.375

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Shengping, Y., and Gilbert, B. (2017). The receiver operating characteristic (ROC) curve. *Southwest Respir. Crit. Care Chronicles* 5.

Shevach, E. M. (2009). Mechanisms of foxp3+ T regulatory cell-mediated suppression. *Immunity* 30, 636–645. doi:10.1016/j.immuni.2009.04.010

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics. *CA Cancer J. Clin.* 72, 7–33.2022

Song, C., Guo, Z., Yu, D., Wang, Y., Wang, Q., Dong, Z., et al. (2020). A prognostic nomogram combining immune-related gene signature and clinical factors predicts survival in patients with lung adenocarcinoma. *Front. Oncol.* 10, 1300. doi:10.3389/fonc.2020.01300

Wang, R. T., Zhi, X. Y., Zhang, Y., and Zhang, J. (2013). Inhibition of SENP1 induces radiosensitization in lung cancer cells. *Exp. Ther. Med.* 6, 1054–1058. doi:10.3892/etm.2013.1259

Wang, T. H., Wang, H. S., and Soong, Y. K. (2000). Paclitaxel-induced cell death: Where the cell cycle and apoptosis come together. *Cancer* 88, 2619–2628. doi:10.1002/1097-0142(20000601)88:11<2619::aid-cncr26>3.0.co;2-j

Won, T. J., Lee, Y. J., Hyung, K. E., Yang, E., Sohn, U. D., Min, H. Y., et al. (2015). SUMO2 overexpression enhances the generation and function of interleukin-17-producing CD8$^+$ T cells in mice. *Cell Signal* 27, 1246–1252. doi:10.1016/j.cellsig.2015.03.001

Wu, Z., Huang, H., Han, Q., Hu, Z., Teng, X. L., Ding, R., et al. (2022). SENP7 senses oxidative stress to sustain metabolic fitness and antitumor functions of CD8+ T cells. *J. Clin. Invest.* 132, e155224. doi:10.1172/JCI155224

Zou, S., Tong, Q., Liu, B., Huang, W., Tian, Y., and Fu, X. (2020). Targeting STAT3 in cancer immunotherapy. *Mol. Cancer* 19, 145. doi:10.1186/s12943-020-01258-7

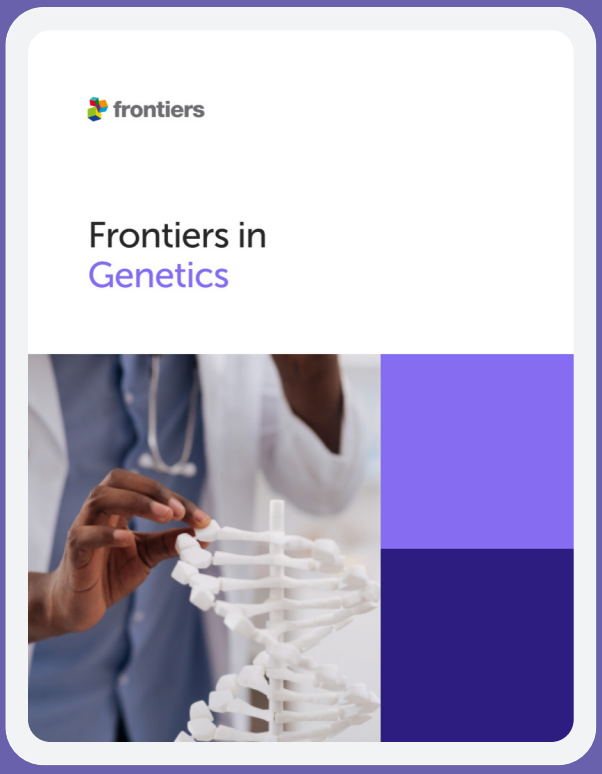
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact