frontiers | Research Topics

Characterizing the uncharacterized human proteins

Edited by
Andy T. Y. Lau, Mee-Hyun Lee, Chi-Ming Wong and Yan-Ming Xu

Published in
Frontiers in Genetics
Frontiers in Molecular Biosciences

ISSN 1664-8714
ISBN 978-2-8325-2452-7
DOI 10.3389/978-2-8325-2452-7

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

Characterizing the uncharacterized human proteins

Topic editors

Andy T. Y. Lau — Shantou University, China
Mee-Hyun Lee — Dongshin University, Republic of Korea
Chi-Ming Wong — Hong Kong Polytechnic University, SAR China
Yan-Ming Xu — Shantou University, China

Citation

Lau, A. T. Y., Lee, M.-H., Wong, C.-M., Xu, Y.-M., eds. (2023). *Characterizing the uncharacterized human proteins*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2452-7

Table of contents

04 Editorial: Characterizing the uncharacterized human proteins
Yan-Ming Xu, Mee-Hyun Lee, Chi-Ming Wong and Andy T. Y. Lau

07 Mxi1-0 Promotes Hypoxic Pulmonary Hypertension *Via* ERK/c-Myc-dependent Proliferation of Arterial Smooth Muscle Cells
Liang Dong, Xinning Liu, Bo Wu, Chengwei Li, Xiaomin Wei, Gulinuer Wumaier, Xiujuan Zhang, Jing Wang, Jingwen Xia, Yuanyuan Zhang, Ruzetuoheti Yiminniyaze, Ning Zhu, Jing Li, Daibing Zhou, Youzhi Zhang, Shuanghui Li, Junzhu Lv and Shengqing Li

20 A Potential Indicator ARRDC2 Has Feasibility to Evaluate Prognosis and Immune Microenvironment in Ovarian Cancer
Mengjun Zhang, Yunduo Liu, Yuan Liu, Siyu Hou, Hao Li, Ying Ma, Can Wang and Xiuwei Chen

37 Efficient Detection of the Alternative Spliced Human Proteome Using Translatome Sequencing
Chun Wu, Xiaolong Lu, Shaohua Lu, Hongwei Wang, Dehua Li, Jing Zhao, Jingjie Jin, Zhenghua Sun, Qing-Yu He, Yang Chen and Gong Zhang

46 The Epigenetic Regulation of Nonhistone Proteins by SETD7: New Targets in Cancer
Chengyao Chiang, Heng Yang, Lizhi Zhu, Chunlan Chen, Cheng Chen, You Zuo and Duo Zheng

57 YT521-B homology domain family proteins as N6-methyladenosine readers in tumors
Heng Yang, Chengyao Chiang, Qinhong Luo, Chunlan Chen, Junrong Huang, Lizhi Zhu and Duo Zheng

65 Advances in molecular characterization of myeloid proliferations associated with Down syndrome
Jixia Li and Maggie L. Kalev-Zylinska

85 Synergistic effects of rare variants of ARHGAP31 and FBLN1 *in vitro* in terminal transverse limb defects
Hong Tian, Fan Chu, Yingjie Li, Mengmeng Xu, Wenjiao Li and Chuanzhou Li

101 Single-gene knockout-coupled omics analysis identifies C9orf85 and CXorf38 as two uncharacterized human proteins associated with ZIP8 malfunction
Heng Wee Tan, Yan-Ming Xu, Zhan-Ling Liang, Na-Li Cai, Yu-Yao Wu and Andy T. Y. Lau

115 Screening and functional analysis of the differential peptides from the placenta of patients with healthy pregnancy and preeclampsia using placental peptidome
Tingting Chen, Zhongxiao Zhang, Qin Lu and Jun Ma

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 May 2023
DOI 10.3389/fgene.2023.1203691

OPEN ACCESS

EDITED AND REVIEWED BY
Maxim B. Freidin,
Queen Mary University of London,
United Kingdom

*CORRESPONDENCE
Andy T. Y. Lau,
andytylau@stu.edu.cn

RECEIVED 11 April 2023
ACCEPTED 14 April 2023
PUBLISHED 04 May 2023

CITATION
Xu Y-M, Lee M-H, Wong C-M and Lau ATY (2023), Editorial: Characterizing the uncharacterized human proteins. *Front. Genet.* 14:1203691. doi: 10.3389/fgene.2023.1203691

Editorial: Characterizing the uncharacterized human proteins

Yan-Ming Xu [1], Mee-Hyun Lee [2], Chi-Ming Wong [3] and Andy T. Y. Lau [1]*

[1]Laboratory of Cancer Biology and Epigenetics, Department of Cell Biology and Genetics, Shantou University Medical College, Shantou, Guangdong, China, [2]College of Korean Medicine, Dongshin University, Naju, Republic of Korea, [3]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China

KEYWORDS

uncharacterized human proteins, proteoforms, open reading frame (orf), alternative splicing, genomics, proteomics, systems biology

Editorial on the Research Topic
Characterizing the uncharacterized human proteins

As of the first week of April 2023, according to the data from the UniProt Knowledgebase (https://www.uniprot.org) (UniProt Consortium, 2023), the human proteome has a record of 20,422 canonical as well as 21,998 non-canonical protein isoforms. These figures are noteworthy because the number of non-canonical isoforms is almost equal to the number of canonical ones, suggesting that "non-canonical" isoforms may be equally important as their canonical counterparts. Despite the fact that scientists have studied canonical isoforms in greater detail over the years, there are still hundreds to thousands of uncharacterized canonical and non-canonical isoforms as their biological functions are yet to be revealed. Besides, the discovery of protein encoding ncRNAs (Jackson et al., 2018) has added another layer of complexity to the situation, and as such, the current reported total number of human proteoforms (Smith and Kelleher, 2018) is likely to be underestimated. The Research Topic titled "*Characterizing the uncharacterized human proteins*" showcases the latest discoveries in the field of unstudied or under-studied proteins/isoforms. It consists of 9 published manuscripts, including 6 research and 3 review articles, with 66 different contributors. This Research Topic covers a wide range of studies, from individual uncharacterized proteins to high-throughput analyses of novel peptides/protein variants in both normal and pathological conditions.

Among the research works covered in this Research Topic, Tan et al. unveiled the potential relationships between essential micronutrients and uncharacterized human proteins C9orf85 and CXorf38, where their expressions could be selectively induced by manganese and selenium. Dong et al. demonstrated that Max interacting protein 1–0 (Mxi1-0), a functional isoform of Mxi1, potentiates hypoxic pulmonary hypertension through MEK/ERK/c-Myc-mediated proliferation of pulmonary arterial smooth muscle cells. Tian et al. reported on a human case carrying two rare variants of ARHGAP31 [c.2623G>A (*p*.Glu875Lys)] and FBLN1 [c.1649G>A (*p*.Arg550His)], where the synergistic effects of these two protein mutants may potentiate the terminal transverse limb defects (TTLD), expanding the clinical complexity of mutant gene product interactions in genetic disorders. Zhang et al. studied arrestin domain containing 2 (ARRDC2), a protein in the α-arrestin family, in ovarian cancer

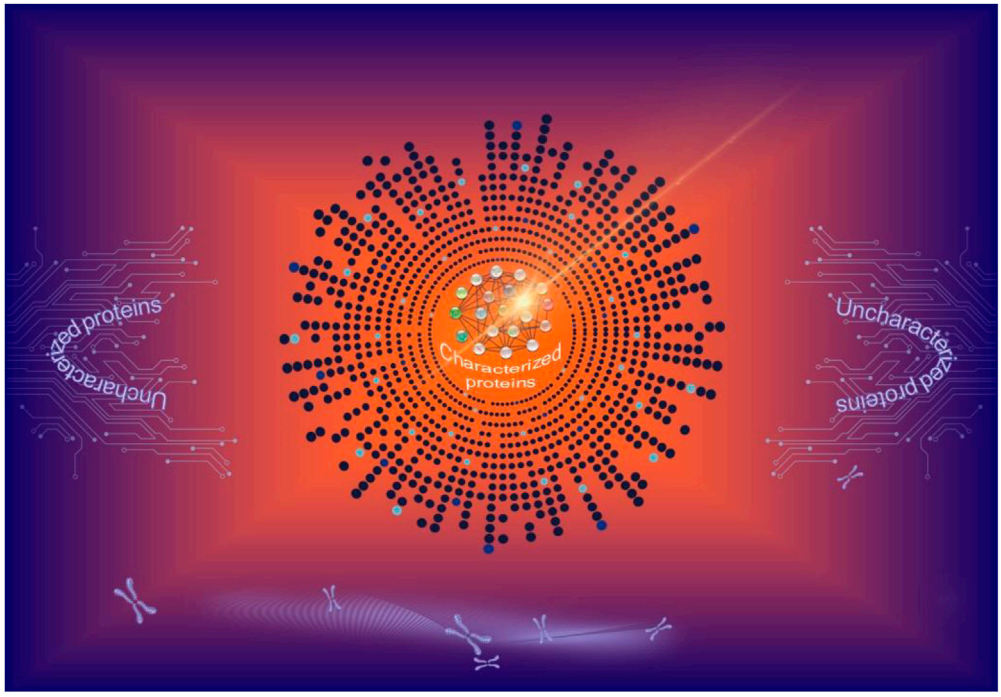


FIGURE 1
The exploration of the uncharted territories of the human proteome. As time goes on, all uncharacterized human proteins will gradually be characterized and merged to the existing protein network.

(OC). They found that high ARRDC2 expression level is associated with malignant biological behavior and poor overall survival of OC, suggesting that ARRDC2 can be used as a potential indicator to evaluate the prognosis of OC. Chen et al. used an untargeted proteomic approach with LC-MS/MS to screen and functionally analyze peptides from the placenta of healthy subjects versus patients with preeclampsia (PE). They identified a differentially-expressed peptide named placenta-derived peptide (PDP, with the sequence AASAKKKNKKGKTISL), derived from the precursor protein eukaryotic translation initiation factor 4B, which could bind to TGF-β1 and impact the Smad signaling pathway, demonstrating that placental bioactive peptides may regulate placental function during the progression of PE. Wu et al. used translatome sequencing to investigate alternative splicing (AS) isoforms in human hepatocellular carcinoma MHCC97H cell line. They identified 50 novel protein isoforms in mass spectrometry datasets, demonstrating the potential of translatome sequencing in investigating the proteome of AS isoforms.

Furthermore, this Research Topic includes excellent literature reviews. Li and Kalev-Zylinska provided a comprehensive and up-to-date summary of molecular alterations in myeloid leukemia associated with Down syndrome, including the aberrant expression of proteins on chromosome 21 (such as C21orf66) and GATA1 mutations, which drive expression of a truncated GATA1 protein. Chiang et al. focused on reviewing a lysine methytransferase named SETD7 (also known as KIAA1717). Besides its well-known epigenetic regulatory role as a histone lysine methyltransferase, SETD7 can also methylate other nonhistone proteins (over 30 substrates, including transcriptional-related proteins and enzymes). Thus, manipulating SETD7 and subsequently its substrate methylation levels might be possible strategies for cancer intervention. Lastly, Yang et al. discussed the YT521-B homology domain family proteins [YTHDFs, including YTHDF1 (also known as C20orf21), YTHDF2 and YTHDF3], in which more and more studies have supported these N6-methyladenosine readers could play a key role in tumor transcription, translation, protein synthesis, tumor stemness, epithelial–mesenchymal transition, immune escape, and chemotherapy resistance.

Although much more collaborative efforts are required to unveil the mysteries of the cellular functions of uncharacterized human proteins, the works in this Research Topic would certainly arouse the attention of scientists in the urgent need to fill in the knowledge gaps in this frontier in the years to come. With this goal being set, as time goes on, sooner or later all uncharacterized human proteins would ideally be characterized and have their names rewritten in history (Figure 1).

Author contributions

ATYL took the lead in conceiving this Research Topic. As guest editors, ATYL, Y-MX, and M-HL invited authors and oversaw manuscript review. ATYL and Y-MX wrote the editorial with input from M-HL and C-MW. All authors listed herein have made their intellectual contributions to and approved the work for publication.

Acknowledgments

We would like to thank Miss Xiao-Yun Zhao for her assistance on the design of this Research Topic image.

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

Jackson, R., Kroehling, L., Khitun, A., Bailis, W., Jarret, A., York, A. G., et al. (2018). The translation of non-canonical open reading frames controls mucosal immunity. *Nature* 564 (7736), 434–438. doi:10.1038/s41586-018-0794-7

Smith, L. M., and Kelleher, N. L. (2018). Proteoforms as the next proteomics currency. *Science* 359 (6380), 1106–1107. doi:10.1126/science.aat1884

UniProt Consortium (2023). UniProt: The universal protein Knowledgebase in 2023. *Nucleic Acids Res.* 51 (D1), D523–D531. doi:10.1093/nar/gkac1052

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 March 2022
doi: 10.3389/fgene.2022.810157

Mxi1-0 Promotes Hypoxic Pulmonary Hypertension *Via* ERK/c-Myc-dependent Proliferation of Arterial Smooth Muscle Cells

*Liang Dong[1†], Xinning Liu[1†], Bo Wu[2†], Chengwei Li[1], Xiaomin Wei[1], Gulinuer Wumaier[1], Xiujuan Zhang[1], Jing Wang[1], Jingwen Xia[1], Yuanyuan Zhang[1], Ruzetuoheti Yiminniyaze[1], Ning Zhu[1], Jing Li[1], Daibing Zhou[1], Youzhi Zhang[1], Shuanghui Li[1], Junzhu Lv[1] and Shengqing Li[1]**

[1]Department of Pulmonary and Critical Care Medicine, Huashan Hospital, Fudan University, Shanghai, China, [2]Department of Lung Transplantation, Wuxi People's Hospital, Wuxi, China

Background: Hypoxic pulmonary hypertension (HPH) is a challenging lung arterial disorder with remarkably high incidence and mortality, and so far patients have failed to benefit from therapeutics clinically available. Max interacting protein 1–0 (Mxi1-0) is one of the functional isoforms of Mxi1. Although it also binds to Max, Mxi1-0, unlike other Mxi1 isoforms, cannot antagonize the oncoprotein c-Myc because of its unique proline rich domain (PRD). While Mxi1-0 was reported to promote cell proliferation *via* largely uncharacterized mechanisms, it is unknown whether and how it plays a role in the pathogenesis of HPH.

Methods: GEO database was used to screen for genes involved in HPH development, and the candidate players were validated through examination of gene expression in clinical HPH specimens. The effect of candidate gene knockdown or overexpression on cultured pulmonary arterial cells, e.g., pulmonary arterial smooth muscle cells (PASMCs), was then investigated. The signal pathway(s) underlying the regulatory role of the candidate gene in HPH pathogenesis was probed, and the outcome of targeting the aforementioned signaling was evaluated using an HPH rat model.

Results: Mxi1 was significantly upregulated in the PASMCs of HPH patients. As the main effector isoform responding to hypoxia, Mxi1-0 functions in HPH to promote PASMCs proliferation. Mechanistically, Mxi1-0 improved the expression of the proto-oncogene c-Myc *via* activation of the MEK/ERK pathway. Consistently, both a MEK inhibitor, PD98059, and a c-Myc inhibitor, 10058F4, could counteract Mxi1-0-induced PASMCs proliferation. In addition, targeting the MEK/ERK signaling significantly suppressed the development of HPH in rats.

Conclusion: Mxi1-0 potentiates HPH pathogenesis through MEK/ERK/c-Myc-mediated proliferation of PASMCs, suggesting its applicability in targeted treatment and prognostic assessment of clinical HPH.

Keywords: hypoxic pulmonary hypertension (HPH), max interacting protein 1–0 (Mxi1-0), pulmonary arterial smooth muscle cells (PASMCs), cell proliferation, MEK/ERK, c-Myc

OPEN ACCESS

Edited by:
Andy T Y Lau,
Shantou University, China

Reviewed by:
Tan Heng Wee,
Shantou University, China
Z XY,
Shantou University, China

****Correspondence:***
Shengqing Li
shengqingli@hotmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *06 November 2021*
Accepted: *08 February 2022*
Published: *23 March 2022*

Citation:
Dong L, Liu X, Wu B, Li C, Wei X, Wumaier G, Zhang X, Wang J, Xia J, Zhang Y, Yiminniyaze R, Zhu N, Li J, Zhou D, Zhang Y, Li S, Lv J and Li S (2022) Mxi1-0 Promotes Hypoxic Pulmonary Hypertension Via ERK/c-Myc-dependent Proliferation of Arterial Smooth Muscle Cells. Front. Genet. 13:810157. doi: 10.3389/fgene.2022.810157

INTRODUCTION

Hypoxic pulmonary hypertension (HPH) due to lung diseases and/or chronic hypoxia is among the most common groups of pulmonary hypertension with high mortality (Badesch et al., 2010; Galiè et al., 2015). Although the two key pathological processes, vasoconstriction and pulmonary vascular remodeling, have been adequately investigated (Dunham-Snary et al., 2017; Woo et al., 2019; Young et al., 2019), the 3 years survival rate of HPH is significantly worse than of other categories of HPH (Hurdman et al., 2012). In advanced lung disease, regional hypoxic vasoconstriction improves the matching of perfusion and alveolar ventilation, but subsequently leads to the increase in pulmonary circulation pressure (Naeije and Dedobbeleer, 2013; Li et al., 2018), underlining the difficulty of HPH treatment despite the availability of novel drugs (Prins et al., 2017; Cassady and Reed, 2019). Therefore, there is an urgent need to develop novel therapeutics based on the identification of key players in the pathogenesis of HPH. In particular, although hypoxia-induced cell proliferation of vascular endothelial and smooth muscle cells is a hallmark of HPH, the underlying mechanisms remain largely elusive.

Max interacting protein 1–0 (Mxi1-0) is an alternative transcript of Mxi1 involved in the Myc-Max-Mad network (Zervos et al., 1993; Dugast-Darzacq et al., 2004). Mxi1-0 shares exons 2 to 6 with the originally cloned Mxi1 (here-in-after referred to as Mxi1-1). The common basic region helix-loop-helix/leucine zipper (bHLH/LZ) enables Mxi1-0 to bind to Max and regulate cell behavior (Hurlin and Huang, 2006). Compared with Mxi1-1, Mxi1-0 has an alternative first exon (exon 0), encoding a different SIN3 interaction domain (SID) (Engstrom et al., 2004), and is reported to promote proliferation of various types of cells, such as endothelial cells and neuroblastoma cells (Armstrong et al., 2013; Wu et al., 2017). However, it is uncharacterized whether and how Mxi1-0 plays a regulatory role in HPH-related over-proliferation of pulmonary arterial endothelial cells (PAECs) or pulmonary arterial smooth muscle cells (PASMCs).

In this study, we found in clinical specimens that Mxi1-0 was crucially involved in the development of HPH. Mxi1-0 but not Mxi1-1 improved the mitosis of PASMCs. Mechanistically, Mxi1-0 was transcriptionally activated by hypoxia and upregulates the proto-oncogene c-Myc *via* MEK/ERK signaling.

MATERIALS AND METHODS

Sample Collection and Ethics Statement

6 chronic lung disease (CLD) samples with or without hypertension were obtained during lung transplantation, and three samples of donor lung tissues were taken from the lungs that were not transplanted. All experiments related to human samples were conducted in accordance with the Declaration of Helsinki and were approved by the ethics committee. All subjects provided written informed consent prior to participation in the study.

Hematoxylin and Eosin Staining

The fixed lungs were sliced in the mid-sagittal plane, embedded in paraffin, and cut into 5 μm thick sections with a microtome. Then, the sections were placed on glass slides, stained with hematoxylin and eosin (HE) staining for morphological analysis, and visualized under an Olympus BX41 microscope (Tokyo, Japan).

Immunohistochemistry and Immunofluorescence Staining

Paraffin-embedded lung tissue sections were deparaffinized in xylene and rehydrated in a graded ethanol series to PBS. Antigen retrieval was performed by pressure cooking in citrate buffer for 10 min. The sections were permeabilized by incubation with 0.3% Triton X-100 and blocked with 5% donkey serum albumin in a humidified chamber for 1h, and were immunostained with primary antibodies to Mxi1 (1:200, ab28740, Abcam Ltd., Cambridge, United Kingdom) or α-smooth muscle actin (α-SMA) (1:200, 48938, CST, Boston, United States). After overnight incubation, sections were washed and incubated with the respective secondary antibodies, donkey anti-rabbit IgG (1:1,000, Jackson Immuno, PA, United States), alexa 594 donkey anti-rabbit IgG (1:1,000, Jackson Immuno, PA, United States), or oralexa 488 donkey anti-mouse IgG (1:1,000, Jackson Immuno, PA, United States) for 1 h. For immunohistochemistry, sections were counterstained with hematoxylin and detected by incubation with the DAB substrate. For immunofluorescence staining, sections were counterstained with nuclear DAPI (1:1,000) and mounted with fluorescent mounting medium.

Cell Culture and Transfection

Primary human PASMCs were purchased from ScienCell, Inc (#3110) and cultured in smooth muscle cell medium (SMCM, ScienCell). Primary human PAECs were purchased from ScienCell, Inc (#3100) and cultured in endothelial cell medium (ECM, ScienCell). Cells were used at passages 4–7. In the cell-growth assay, PASMCs were exposed to normoxia, hypoxia, and hypoxia plus various antagonists as indicated. Cells in the normoxia group were maintained at 37°C in 21% O_2, 74% N_2, and 5% CO_2 (Forma 370, Thermo, United States). Cells in the hypoxia groups were separately cultured at 1% O_2, 94% N_2, and 5% CO_2 (Forma 3131, Thermo, United States). When cells reached 80% confluence, they were transfected with different siRNAs using Lipofectamine 3000 Transfection Reagent (Invitrogen). 6 h after transfection, cells were cultured in serum containing medium for a resting period of 24 h, followed by hypoxia exposure for different times. Mxi1-1 siRNA (target sequence: 5′-CGUCGCACAUGUUCC GGAACG-3′) and Mxi1-0 siRNA (target sequence: 5′-CAGCGAGAACUCGAUGGAGAATT-3′) were synthesized by GenePharma (Shanghai Gene Pharma Co.,). As a control,

commercially available non-targeting siRNA (si-Control) was used.

Immunocytofluorescence Assay

Human PASMCs grown on chamber slides were fixed with 4% paraformaldehyde, permeabilized by incubation with 0.3% Triton X-100 and blocked with 5% donkey serum albumin for 2 h. Then, slides were immunostained overnight with primary antibodies to Mxi1 (1:100, ab28740, Abcam Ltd., Cambridge, United Kingdom), followed by 1 h incubation with a secondary antibody, alexa 594 donkey anti-rabbit IgG (1:200, Jackson Immuno, PA, United States). After incubation, slides were counterstained with nuclear DAPI (1:1,000) and mounted with fluorescent mounting medium. Fluorescent images were taken with fluorescence microscope (Olympus, Japan).

Lentiviral Infection

The recombinant lentiviruses expressing Flag-tagged Mxi1-1 and HA-tagged Mxi1-0 were purchased from Genscript (Jiangsu, China). To prepare Mxi1-1 or Mxi1-0 overexpression viral particles, HEK293T cells were co-transfected with each viral vector and the packaging vectors (pMD2.0G and psPAX) using JetPEI purchased from Qbiogene (Montreal, Canada) following the manufacturer's protocol. The medium was replaced 4 h after transfection, and cells were cultured for a further 36 h. Viral particles were harvested, filtered using a 0.45 μm syringe filter, and combined with 10 μg/ml polybrene (Sigma, MO, United States). PASMCs at 60% confluency were treated with these particles overnight. The culture medium was then replaced with fresh complete growth medium, and cells were cultured for a further 24 h and selected with puromycin (1 μg/ml). The selected cells were used in further experiments.

Quantitative Real-Time PCR

The mRNA level was determined by qRT-PCR. Total RNA was extracted from cells or tissues using Trizol reagent (Invitrogen, CA, United States) according to the manufacturer's protocol. RNA was subsequently reverse transcribed into cDNA using ReverTra Ace qPCR RT Kit (TOYOBO, Japan). cDNA was amplified and detected using Hieff qPCR SYBR Green Master Mix (Yeasen, Shanghai, China). The PCR primers were as follows: 5′-GGACCT GACTGACTACCTCAT-3′ and 5′-CGTAGCACAGCTTCTCCT TAAT-3′ for β-actin, 5′-GAGGCTGCCGAGTTTTTGG-3′ and 5′-TCGGCATGGACGGGAAT-3′ for Mxi1-1, and 5′-GAGACCGAC ACACACTCCCATA-3′ and 5′-CGAAAAGCCGGCCTGACT-3′ for Mxi1-0. Fold change of RNA species was analyzed with the formula ($2^{-\Delta\Delta Ct}$), and was normalized to β-actin expression.

Western Blotting

Total lysates of cells or tissues were homogenized in RIPA lysis buffer (Beyotime, Shanghai, China) supplemented with a protease inhibitor, phenylmethyl sulfonyl fluoride (1 mM). Equivalent amounts of protein were separated by SDS-polyacrylamide gels and transferred to 0.22 μM nitrocellulose membranes (Millipore, MA, United States). After blocking, the membranes were probed with one of the following primary antibodies overnight at 4°C: anti-Mxi1 (1:1,000, sc-130627, Santa Cruz, TX, United States), anti-MEK (1:1,000, 4,694, CST, Boston, United States), anti-phospho-MEK (1:1,000, 9,154, CST, Boston, United States), anti-ERK (1:1,000, 4,695, CST, Boston, United States), anti-phospho-ERK (1:1,000, 4,370, CST, Boston, United States) and anti-c-Myc (1:1,000, 13987, CST, Boston, United States). Then, membranes were incubated with secondary antibodies for 1 h at room temperature. Bound antibodies were detected using a Super ECL Detection Reagent (Yeasen, Shanghai, China) and imaged on a Tanon Western blotting detection system (Tanon, Shanghai, China).

Cell Viability Assay

Cells were seeded into 96-well plates at a density of 2000/well and incubated with either vehicles or inhibitors. After exposure to nomorxia or hypoxia, a total of 110 μL of DMEM containing the CCK-8 solution (Beyotime, Shanghai, China) [CCK8:DMEM (v/v) = 1:10] was added to each well, and cells were incubated for 4 h. Finally, cell viability was determined by measuring the absorbance at 450 nm using a multi-well spectrophotometer (Bio-RAD, CA, United States).

Cell Proliferation Assay

Proliferation of cells was detected with a BeyoClickEdU Cell Proliferation Kit with Alexa Fluor 647 (Beyotime, Shanghai, China) following the manufacturer's protocol. Briefly, cells were seeded into 12-well plates at a density of 10000/well and incubated with siRNA or vehicle. After exposure to nomorxia or hypoxia for 48h, cells were treated with EdU (20 μm) for 3 h and subjected to fixing and permeabilization. Then, cells were exposed in click additive solution for 30 min followed by Hoechst staining for 10min, and finally observed with fluorescence microscope (Olympus, Japan).

Animal Models

Adult male Sprague-Dawley rats weighing 150–200 g were purchased from the Slake Company (Shanghai, China). All protocols and surgical procedures were approved by the Fudan University Veterinary Medicine Animal Care and Use Committee. Animals were randomly divided into five groups (n = 5/group): 1) normoxia (Nor), 2) chronic hypoxia (Hyp), 3) chronic hypoxia and treatment with DMSO (DMSO), 4) chronic hypoxia and treatment with 0.15 mg/kg/tiw PD98059 (PD0.15), and 5) chronic hypoxia and treatment with 0.3 mg/kg/tiw PD98059 (PD0.3). PD98059 was administered *via* intraperitoneal injection. Rats in the normoxia group were housed at ambient barometric pressure for 28 days (-718 mmHg, PO_2 maintained in 150.6 mmHg). Rats in the hypoxia groups were housed in a hypobaric hypoxia chamber depressurized to 380 mmHg (PO_2 in 79.6 mmHg) for 8 h/day for 28 days as previously described (Li et al., 2019). All animals were raised under a 12 h:12 h light-dark cycle and were freely supplied food and water. The room temperature was maintained at 25°C, and the bedding was changed once per week.

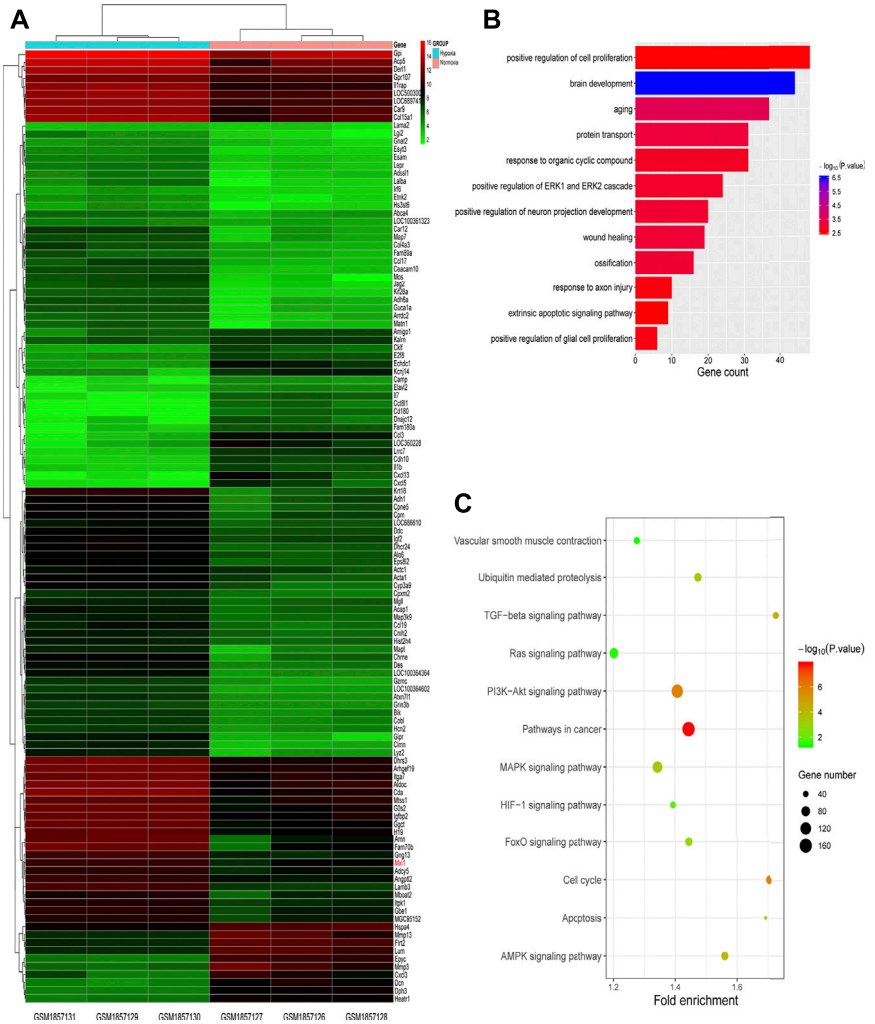


FIGURE 1 | Screening for hypoxic pulmonary hypertension (HPH)-related genes. **(A)** Heatmap of genes differentially expressed in normal and hypoxic pulmonary arterial smooth muscle cells (PASMCs) in GEO database (dataset: GSM1857126~1857131). **(B,C)** GO **(B)** and KEGG **(C)** enrichment analyses for differentially expressed genes described in **(A)**.

Echocardiography and Hemodynamic Analysis

After 28 days, rats were fasted overnight and initially anesthetized with isoflurane inhalation. Echocardiography was performed with Visual Sonics Vevo 2100 ultrasoundmachine and 12S rodent probe (GE Healthcare, CT, United States) to determine pulmonary artery acceleration time (PAAT) and tricuspid annulusplain systolic excursion (TAPSE). Data were analyzed with EchoPAC software (GE Healthcare, CT, United States). Then, haemodynamic analysis was performed as previously described (Xia et al., 2018). Briefly, anesthesia was given to rats with 20% ethylurethanm *via* injecting intraperitoneally (4 ml/kg). After intubation, right ventricular systolic pressure (RVSP) was recorded. To investigate right ventricular hypertrophy (RVH), the right ventricle (RV), the left ventricle plus septum (LV + S) and the body weight (BW) were weighed, and the RV/(LV + S) ratio and RV/BW ratio were determined.

Statistical Analysis

ImageJ software was used to scan the gray level of Western blotting images, and spss24.0 software was used for data statistical analysis. Statistical significance was assessed by comparing mean (±SD) values with Student's t-test for independent groups. $p < 0.05$ was considered as statistically significant. The data were plotted with graphpad prism 8.0 software.

RESULTS

Mxi1 Is Involved in the Clinical Development of Hypoxic Pulmonary Hypertension

To identify candidate players in HPH pathogenesis, we first screened for genes differentially expressed between normal and hypoxic PASMCs using the GEO database (**Figure 1A**, dataset:

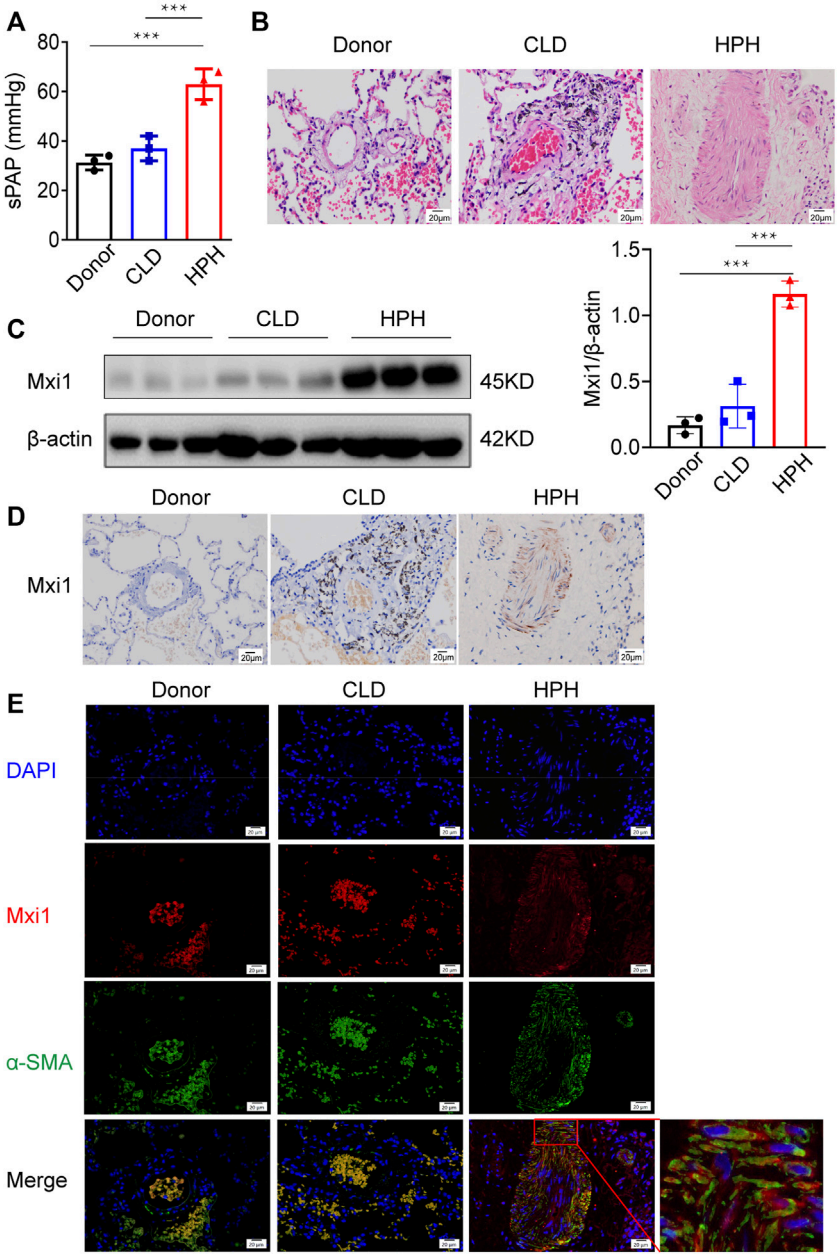


FIGURE 2 | Mxi1 is upregulated in PASMCs of HPH patients. Lung samples of three donors, three chronic lung disease (CLD) patients and three hypoxic pulmonary hypertension (HPH) patients were collected and sectioned. **(A)** Hemodynamic analysis of patients before collection of samples. **(B)** Hematoxylin and eosin staining of paraffin-fixed lung sections was used for morphological analysis of pulmonary arteries. **(C)** The expression of Mxi1 was examined *via* Western blotting analysis, followed by quantification of the blots through densitometry and normalization to β-actin. **(D)** Representative paraffin lung tissue sections from donors, CLD, and HPH patients were subjected to immunohistochemical staining of Mxi1. **(E)** Representative sections in all groups were also subjected to immunofluorescence staining of Mxi1 and α-SMA with nuclei counterstained by DAPI (blue), and the image of HPH merge was zoomed. Scale bar, 20 μm. Data are shown as means ± SDs. For statistical significance, ***represents $p < 0.001$ compared with HPH patients.

GSM1857126-1857131). As a result, we obtained 602 genes, of which 200 were up-regulated and 402 were down-regulated in hypoxia samples. GO and KEGG enrichment analyses were performed using DAVID 6.8 database. The GO results indicated that genes were significantly enriched in cell proliferation, aging and response to organic cyclic compound, mainly concentrated in biological processes related to proliferation (**Figure 1B**), whereas KEGG analysis demonstrated that genes were mainly enriched in growth and differentiation related pathways, such as Ras signaling pathway, PI3K-Akt signaling pathway, MAPK signaling pathway, FOXO signaling pathway, cell cycle and AMPK signaling pathway (**Figure 1C**).

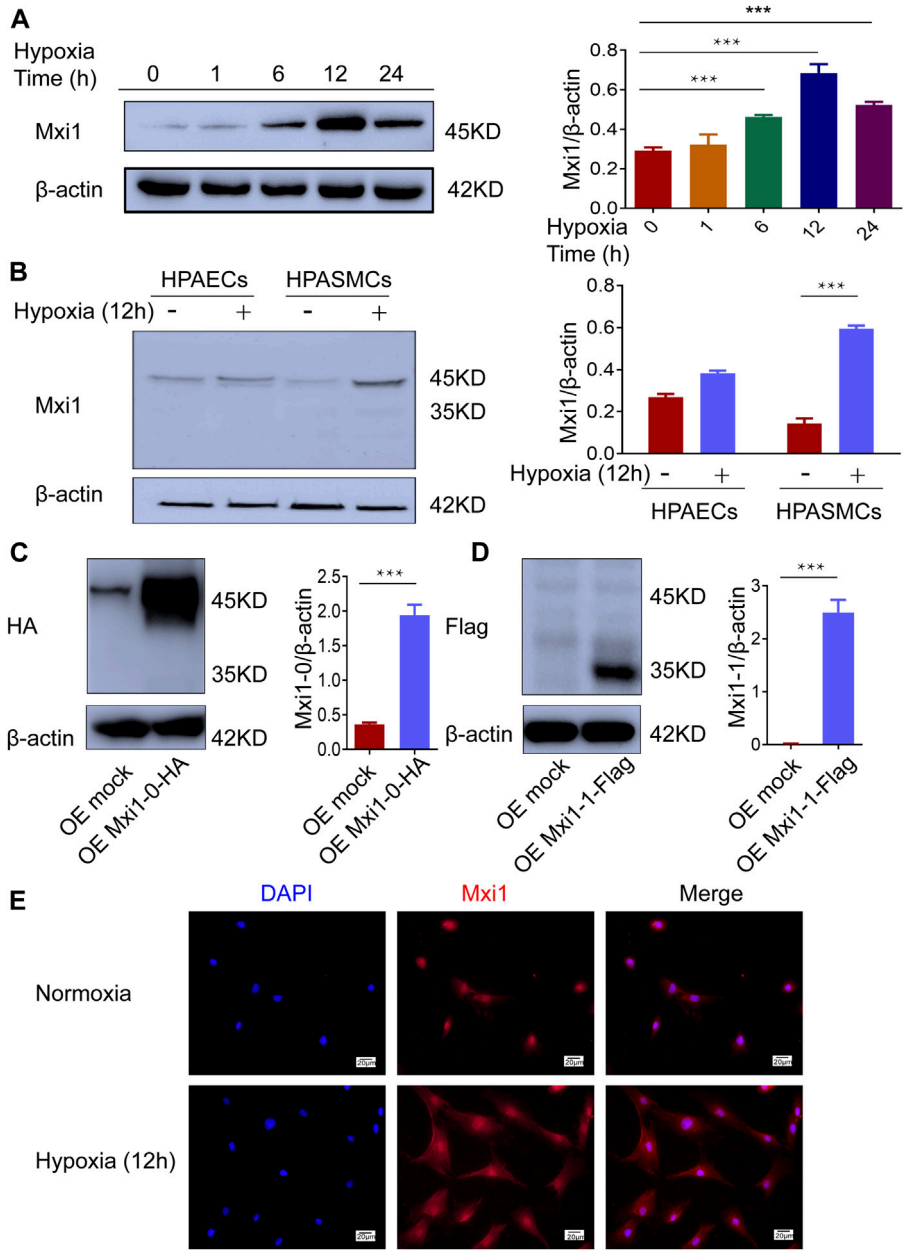


FIGURE 3 | Hypoxia induces expression of Mxi1-0 but not Mxi1-1 in PASMCs. **(A)** PASMCs were exposed to hypoxia (1% O_2) for indicated periods of time. Cell lysates were prepared and subjected to Western blotting assay. **(B)** Pulmonary arterial endothelial cells (PAECs) and PASMCs were exposed to normoxia (21% O_2) or hypoxia for 12 h, and were subjected to Western blottinganalyses. **(C,D)** PASMCs were transfected with constructs for HA-tagged Mxi1-0 **(C)** or Flag-tagged Mxi1-1 **(D)**, and were subjected to Western blotting analyses. **(E)** PASMCs were exposed to normoxia or hypoxia for 12 h, and were subjected to immunostaining with a Mxi1 antibody and counterstaining of the nuclei with DAPI. Scale bar, 20 μm. Data from three independent experiments are shown as means ± SDs. For statistical significance, ***represents $p < 0.001$ compared to normoxia or the mock-transfected group.

Among the most significantly changed genes in HPH is Mxi1, which is remarkably upregulated in HPH and established to participate in the regulation of cell division (Armstrong et al., 2013; Wu et al., 2017). We next investigated the expression of Mxi1 in clinical HPH samples. Specimens of nine patients, including 3 with CLD but not hypertension, 3 with HPH and three donors of normal lungs, were collected and examined (**Figures 2A,B**). Western blotting analysis indicated that Mxi1 was significantly higher in pulmonary of HPH tissues than in those of CLD patients or donors, suggesting a correlation between Mxi1 expression and HPH occurrence (**Figure 2C**). Immunohistochemical staining of the lung tissues of HPH patients demonstrated specific upregulation of Mxi1 in the medial layer, consisting mainly of PASMCs (**Figure 2D**).

Consistently, immunofluorescent staining showed that Mxi1 co-localized with α-SMA (**Figure 2E**). Thus, Mxi1 expression is induced in PASMCs during the development of HPH.

Hypoxia Induces Expression of Mxi1-0 but Not Mxi1-1 in Pulmonary Arterial Smooth Muscle Cells

We next explored the expression of Mxi1 isoforms in pulmonary arterial cells when exposed to hypoxia. Western blotting analysis showed that Mxi1 was induced by hypoxia in a time-dependent manner and peaked when exposed to hypoxic conditions for 12 h (**Figure 3A**). Hypoxia significantly upregulated Mxi1 in PASMCs but not PAECs (**Figure 3B**). Consistent with the assays using clinical pulmonary specimens (**Figure 2C**), we detected a 45 kD protein using the pan-Mxi1 antibody (**Figure 3B**). In addition, to determine the major isoforms induced by hypoxia, we generated vectors for tagged Mxi1-0 and Mxi1-1, and ectopically overexpressed these proteins in PASMCs. Western blotting analyses using antibodies recognizing the tag sequence detected 2 proteins with distinct molecular weights (**Figures 3C,D**), suggesting that Mxi1-0 was the predominant isoform upregulated by hypoxia (**Figure 3A**). Unlike Mxi1-1 that was reported to exhibit a predominant nuclear localization (Engstrom et al., 2004; Erichsen et al., 2015), Mxi1-0 resided both in the cytoplasma and the nucleus (**Figure 3E**). These findings strongly suggest that Mxi1-0 is remarkably upregulated in PASMCs when exposed to hypoxic conditions.

Mxi1-0 but Not Mxi1-1 Mediates Hypoxia-Induced Pulmonary Arterial Smooth Muscle Cells Proliferation

Accumulating evidence has indicated that hypoxemia in the pulmonary vessels leads to increased PASMCs proliferation (Shah, 2012). To test whether Mxi1-0 plays a pathogenic role in the induction of this phenotype, we designed siRNAs that could specifically knockdown Mxi1-0 (siMxi1-0) or Mxi1-1 (siMxi1-1) (**Figure 4A**). While hypoxic treatment significantly promoted the growth of PASMCs, this was ablated by knockdown of Mxi1-0 but not that of Mxi1-1 (**Figure4B**). Conversely, overexpression of Mxi1-0 but not Mxi1-1 improved the growth rates of PASMCs exposed either to normoxia or hypoxia (**Figure 4C**). In line with these observations, we found that Mxi1-0 silencing in hypoxia-treated PASMCs reduced the ratios of EdU-positive cells, whereas Mxi1-0 overexpression enhanced these ratios in both normoxia- and hypoxia-exposed cells (**Figure 4D**), suggesting an essential role of Mxi1-0 in hypoxia-elicited mitosis of PASMCs. Collectively, these data indicate that Mxi1-0 but not Mxi1-1 drives the hyper-proliferative response in hypoxia-exposed PASMCs.

Mxi1-0 Potentiates Pulmonary Arterial Smooth Muscle Cells Proliferation *Via* Upregulation of c-Myc

C-Myc is a proto-oncogene, which plays a key role in the regulation of cell growth and proliferation (Nesbit et al., 1999; Baluapuri et al., 2020). Previous studies have demonstrated that hypoxia leads to the high expression of c-Myc in the occurrence of pulmonary vascular diseases, and that Mxi1 is involved in the regulation of the transcription factor, c-Myc (Zervos et al., 1993; Voelkel et al., 2013). To mechanistically understand how Mxi1-0 might promote proliferation, we assessed the expression of c-Myc in hypoxic PASMCs further subjected to Mxi1-0 silencing or overexpression. As a result, hypoxic exposure increased c-Myc levels, which was further augmented by overexpression and counteracted by knockdown of Mxi1-0 (**Figures 5A,B**). By constrast, Mxi1-1 overexpression or knockdown failed to affect the expression of c-Myc in PASMCs cultured in normoxic or hypoxic conditions (**Figures 5C,D**), consistent with previous findings that Mxi1 forms a complex with Max to repress the transactivity rather than directly reduce the expression of c-Myc (Huang et al., 2018). We next probed whether Mxi1-0 promotes PASMCs growth *via* upregulation of c-Myc in hypoxic conditions. While hypoxia-induced proliferation of PASMCs was significantly impaired by a c-Myc inhibitor, 10058F4, this cannot be rescued by overexpression of Mxi1-0 (**Figures 5E,F**). These observations indicate that the pro-proliferative effect of Mxi1-0 is mediated by upregulation of c-Myc in hypoxic PASMCs.

Mxi1-0 Upregulates c-Myc Through MEK/ERK Signaling in Hypoxia-Exposed Pulmonary Arterial Smooth Muscle Cells

Previous studies suggested that the pathobiology of HPH was associated with the activation of MEK/ERK signaling, which was documented to promote c-Myc expression (Dang et al., 2006). In addition, bioinformatic analysis in the early stage of this study showed that the function of Mxi1-0 was enriched in Ras/MAPK signaling pathway. Hence, we examine whether Mxi1-0 upregulates c-Myc *via* MEK/ERK signaling in hypoxia-treated PASMCs. Indeed, hypoxia induced phosphorylation of MEK and ERK, which was inhibited by knockdown of Mxi1-0 but not Mxi1-1 (**Figures 6A,B**). Hypoxia-induced upregulation of c-Myc was counteracted by PD98059, a classical MEK antagonist, in a dose-dependent manner (**Figure 6C**). Similar to the inhibition of c-Myc, antagonizing MEK/ERK signaling abrogated hypoxia-induced hypoxia-elicited overgrowth of PASMCs, which was not rescured by overexpression of Mxi1-0 (**Figure 6D**). These data suggest that Mxi1-0 enhanced the proliferation of PASMCs through MEK/ERK-mediated upregulation of c-Myc.

We further explore the function of MEK/ERK signaling in HPH pathogenesis using rat models (**Figure 7A**). Sustained exposure of animals to hypoxia is sufficient to induce pulmonary hypertension as determined by hemodynamic (echocardiography and right heart catheter) and morphological analysis, e.g., decreased PAAT and TAPSE (**Figure 7B**). Treatment of rats with PD98059 significantly relieved hypertension as revealed by restored PAAT and TAPSE (**Figure 7B**). In addition, PD98059 treatment dramatically decreased RVSP and ameliorated RVH indices [including RV/BW ratio and RV/(LV + S) ratio], both of

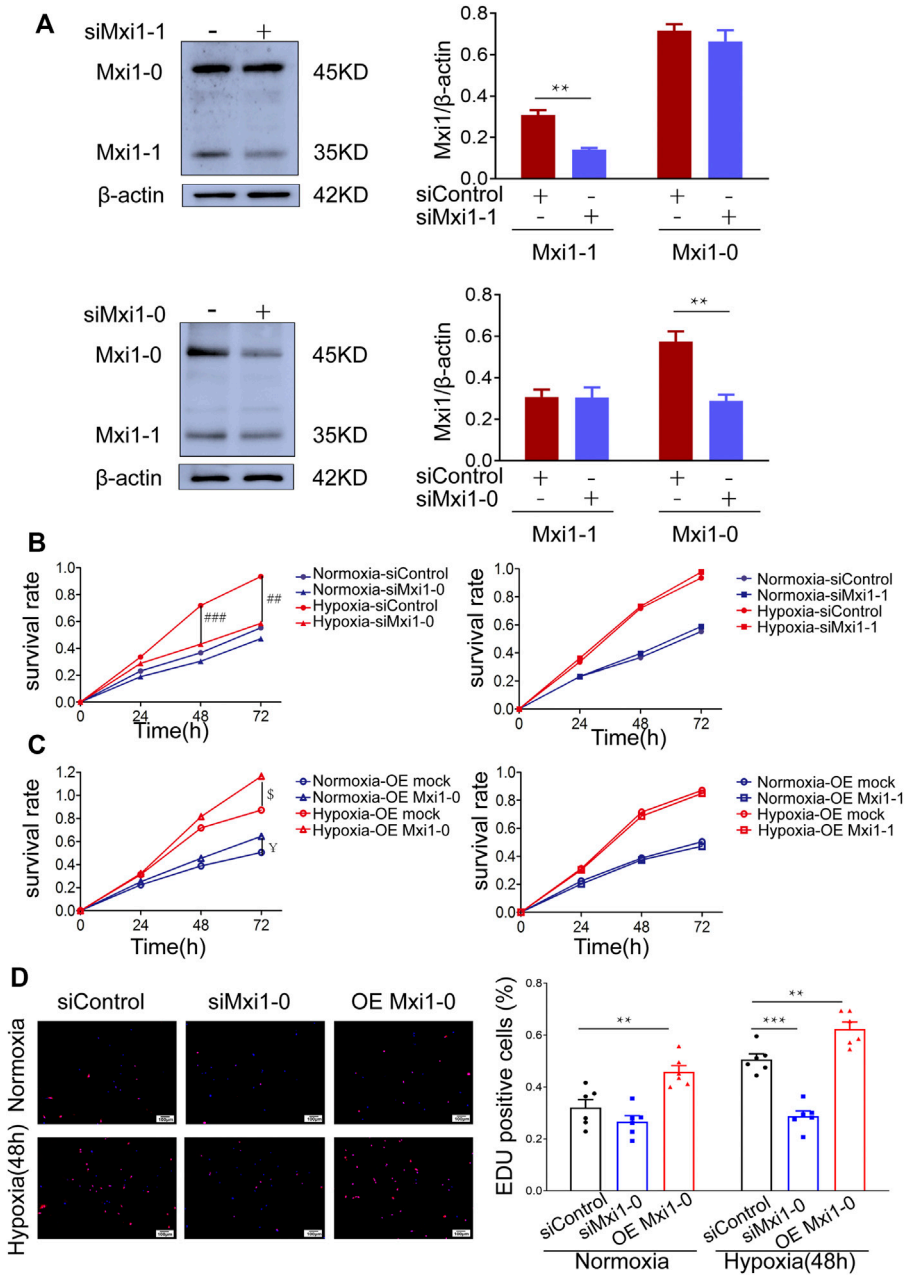


FIGURE 4 | Mxi1-0 but not Mxi1-1 promotes PASMCs growth under hypoxic conditions. **(A)** PASMCs were transfected with siRNAs targeting either Mxi1-0 or Mxi1-1, and were subjected to Western blotting analyses. **(B)** PASMCs cultured in normoxic or hypoxic conditions were transfected with siRNAs targeting either Mxi1-0 or Mxi1-1, and CCK-8 assays were performed on indicated times. **(C)** PASMCs cultured in normoxic or hypoxic conditions were infected with recombinant lentiviruses to overexpress Mxi1-0 or Mxi1-1, and CCK-8 assays were performed on indicated times. **(D)** PASMCs were transfected with Mxi1-0-targeted siRNAs or infected with recombinant lentiviruses to overexpress Mxi1-0, and were cultured in normoxic or hypoxic conditions for 48 h. Cells were then subjected to immunofluorescence staining for EdU, followed by microscopy and calculation of the ratios of EdU-positive cells in three random fields. Scale bar, 100 μm. Data from three independent experiments are shown as means ± SDs. For statistical significance, *represents $p < 0.05$ compared to siControl, **represents $p < 0.01$ compared to siControl, ***represents $p < 0.001$ compared to siControl, $^{\#\#}$ represents $p < 0.01$ compared to Hypoxia-siControl, $^{\#\#\#}$ represents $p < 0.001$ compared to Hypoxia-siControl, $^{\$}$represents $p < 0.05$ compared to Hypoxia-OEmock, $^{¥}$represents $p < 0.05$ compared to Normoxia-OEmock.

which are major characters of HPH (**Figure 7C**). HE staining revealed that PD98059 significantly improved hypoxia-induced pulmonary artery remodeling by reducing the population of PASMCs (**Figure 7D**). These data demonstrated that inhibition of MEK/ERK signaling protects rats against chronic hypoxia-induced HPH.

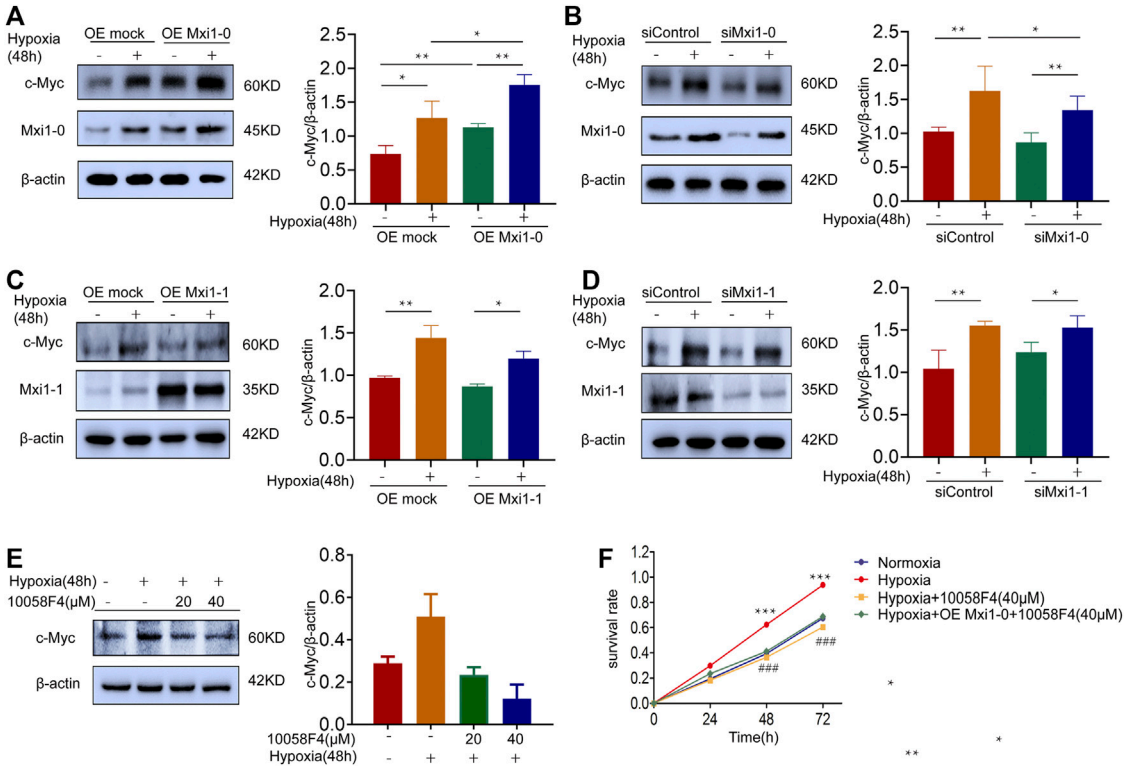


FIGURE 5 | Mxi1-0 upregulates c-Myc production in hypoxic PASMCs. **(A)** PASMCs were infected with recombinant lentiviruses to overexpress Mxi1-0, and were cultured under normoxic or hypoxic conditions for 48 h. Cells were then subjected to Western blotting analysis. **(B)** PASMCs were transfected with Mxi1-0-targeted siRNAs, and were cultured under normoxic or hypoxic conditions for 48 h. Cells were then subjected to Western blotting analysis. **(C)** PASMCs were infected with recombinant lentiviruses to overexpress Mxi1-1, and were cultured under normoxic or hypoxic conditions for 48 h. Cells were then subjected to Western blotting analysis. **(D)** PASMCs were transfected with Mxi1-1-targeted siRNAs, and were cultured under normoxic or hypoxic conditions for 48 h. Cells were then subjected to Western blotting analysis. **(E)** PASMCs were cultured under normoxic or hypoxic conditions for 48 h and treated with indicated doses of 10058F4. Cells were then subjected to Western blotting analysis. **(F)** PASMCs were cultured under normoxic or hypoxic conditions, treated with 10058F4, and infected with control or Mxi1-0-overexpressing lentiviruses. Cells were then subjected to CCK-8 assays on indicated times. Data from three independent experiments are shown as means ± SDs. For statistical significance, *represents $p < 0.05$ compared between two groups, **represents $p < 0.01$ between two groups, ***represents $p < 0.001$ compared between Hypoxia and Normoxia, ### represents $p < 0.001$ compared between Hypoxia and Hypoxia + 10058F4 (40 μM).

DISCUSSION

HPH is a serious pulmonary disorder with systemic complications that aggravate its clinical consequences. To date, no effective treatment for this life-threatening disease has advanced, which is at least partially ascribed to the lack of appropriate biomarkers (Maron et al., 2016; Thompson and Lawrie, 2017). In the present study, we found that Mxi1 was overexpressed in the pulmonary arteries of HPH patients. Immunohistochemical staining showed that Mxi1 was expressed abundantly in the medial layer mainly composed of PASMCs, which was further validated by immunofluorescence indicating that Mxi1 overlapped with α-SMA, a specific biomarker of smooth muscle cells. Although Mxi1 has been extensively studied in the context of carcinogenesis and well documented as a tumor suppressor (Zervos et al., 1993; Huang et al., 2018), this is probably the first study to determine that it is also an important player in HPH, reminiscent of multiple shared pathomechanisms between pulmonary hypertension and cancer (Negi et al., 2021).

Protein isoforms are generated from the same gene due to transcription from different promoter, alternative splicing or varied translation initiation sites (Yap and Makeyev, 2016). Although these isoforms are structurally and functionally similar, it is also common that they play distinct roles or are competitively involved in a physiological or pathological process (Li et al., 2016). The Mxi1 gene is located on chromosome 10q24-q25 (Wechsler et al., 1994), and encodes proteins of three different isoforms, Mxi1-0, Mxi1-1 and Mxi1 WR, of which Mxi1 WR is considered to have no biological function due to lack of SID. Although it is highly homologous with Mxi1-1, Mxi1-0 has an additional N-terminal sequence consisting of 92 amino acids (Engstrom et al., 2004). This novel sequence has a proline rich domain (PRD), which is responsible for the cytoplasmic localization of Mxi1-0 (Armstrong et al., 2013). In the study, we established that Mxi1-0, but not Mxi1-1, was highly expressed in PASMCs of HPH patients. Mxi1-0, but not Mxi1-1, was induced by hypoxia and plays an essential role in the proliferation of PASMCs. These observations were in contrast to Mxi1-1, which was reported to suppress cell growth through binding to Max and

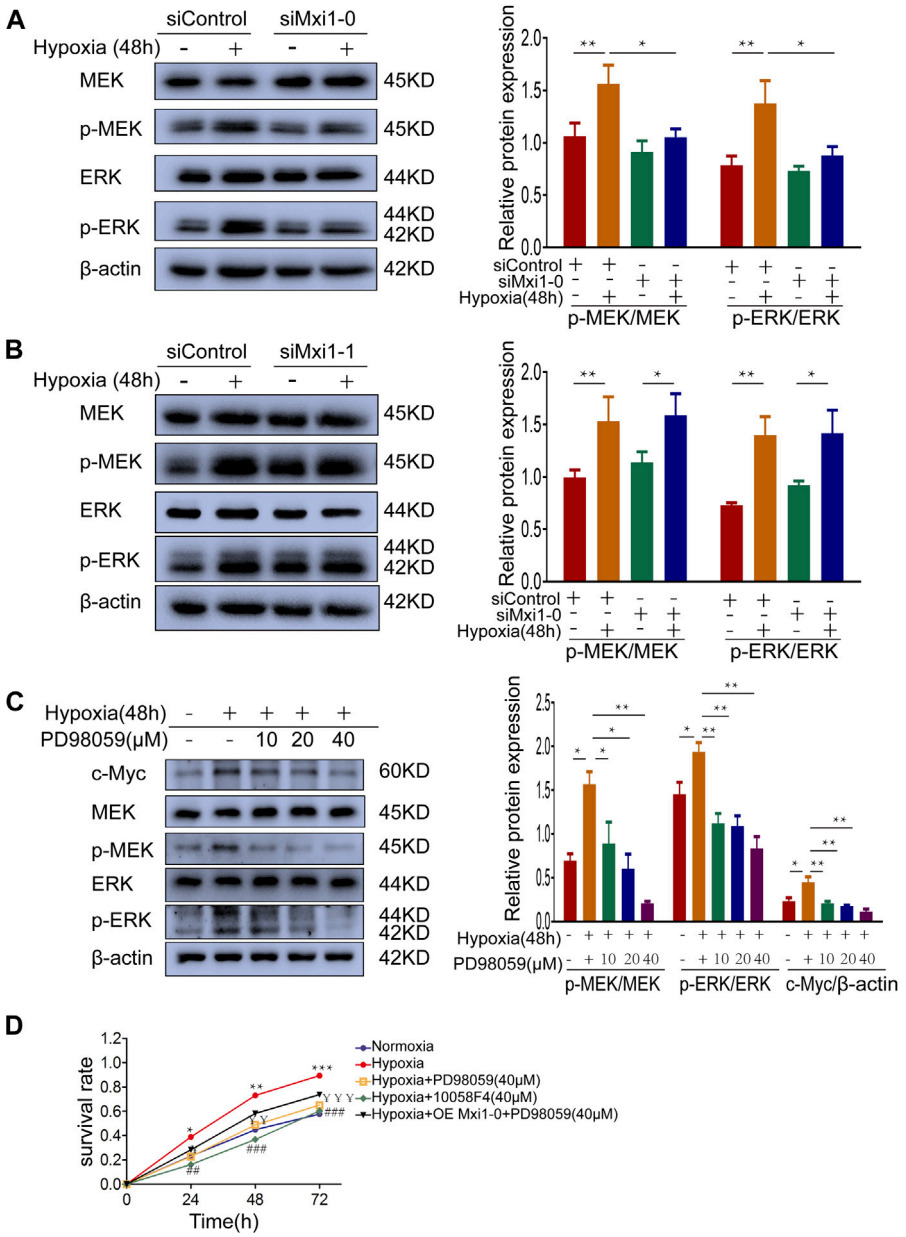


FIGURE 6 | Mxi1-0 upregulates c-Myc through MEK/ERK signaling in hypoxic PASMCs. **(A,B)** PASMCs tranfected with siRNAs targeting Mxi1-0 **(A)** or Mxi1-1 **(B)** were cultured under normoxic or hypoxic conditions for 48 h. Western blotting assay was performed with the indicated antibodies. All the phospho-protein levels were measured by densitometry and normalized to that of β-actin. **(C)** PASMCs were cultured under normoxic or hypoxic conditions for 48 h and treated with different doses of PD98059, and the levels of indicated proteins were determined by Western blotting analysis and quantified by densitometry. **(D)** PASMCs were cultured under normoxic or hypoxic conditions, treated with PD98059 or 10058F4, and infected with control or Mxi1-0-overexpressing lentiviruses. Cells were then subjected to CCK-8 assays on indicated times. Data from three independent experiments are shown as means ± SDs. For statistical significance, *represents $p < 0.05$ compared between two groups or to Normoxia **(D)**, **represents $p < 0.01$ compared between two groups or to Normoxia **(D)**, $^{¥}$represents $p < 0.05$ compared between Hypoxia and Hypoxia + PD98059 (40 μM), $^{¥¥}$represents $p < 0.01$ compared between Hypoxia and Hypoxia + PD98059 (40 μM),$^{¥¥¥}$represents $p < 0.001$ compared between Hypoxia and Hypoxia + PD98059 (40 μM), $^{\#\#}$ represents $p < 0.01$ compared between Hypoxia and Hypoxia+10058F4 (40 μM), $^{\#\#\#}$represents $p < 0.001$ compared between Hypoxia and Hypoxia + 10058F4 (40 μM).

impairing the transcriptional activity of c-Myc (Wechsler et al., 1997; Manni et al., 2002). The unique PRD in Mxi1-0 recruits specific protein chaperones may explain why Mxi1-0 has different cellular functions from Mxi1-1 (Dugast-Darzacq et al., 2004; Hurlin and Huang, 2006). A recent study showed that the deletion of PRD converts Mxi1-0 into a potent suppressor of c-Myc, which is considered to be a key mediator of HPASMCs proliferation (Dugast-Darzacq et al., 2007; Zhang et al., 2019). Other studies suggested that although Mxi1-0 could bind to Max protein, it failed to inhibit c-Myc-dependent transcription but

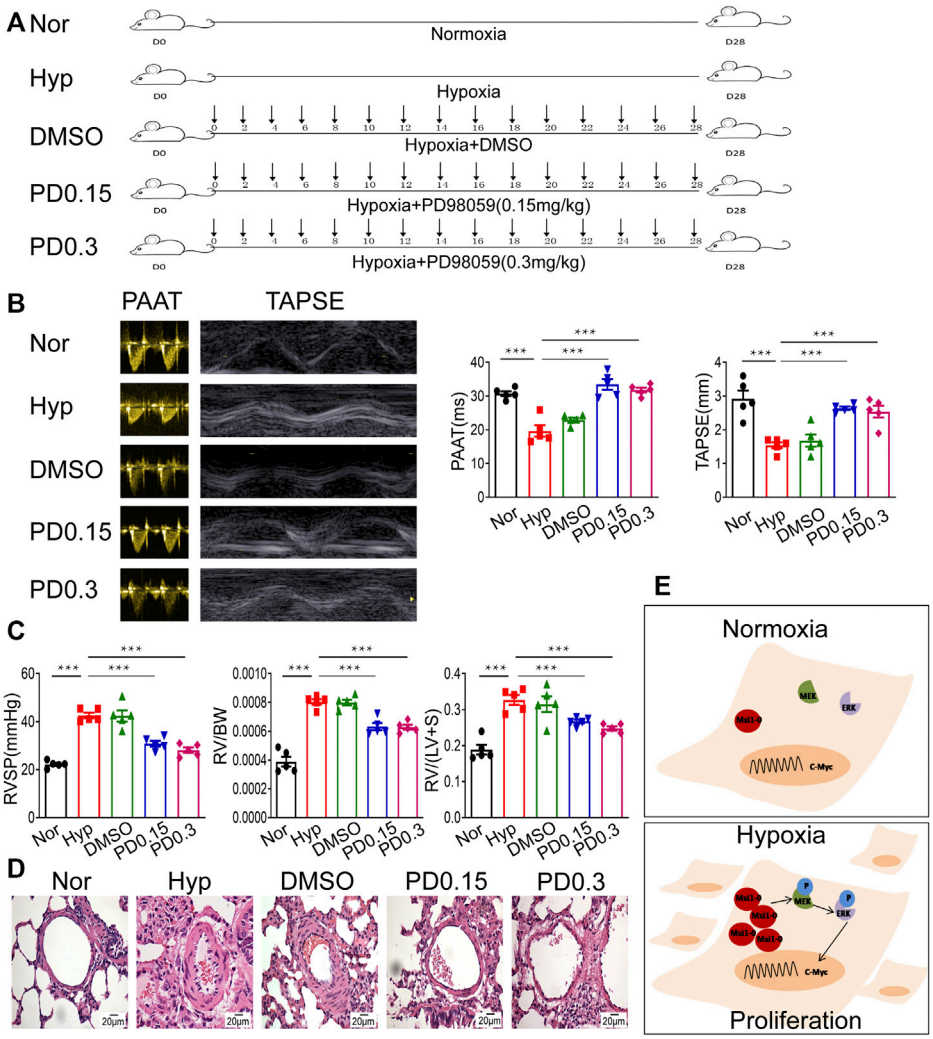


FIGURE 7 | Inhibition of MEK/ERK signaling protects rats against HPH. **(A)** A rat model of hypoxic pulmonary hypertension (HPH) was generated ($n = 5$ animals per group). Rats exposed to chronic hypoxia were treated with vehicle (DMSO) or 0.15 mg/kg or 0.3 mg/kg PD98059. **(B)** Rats in all groups were subjected to echocardiography and measurement of PAATand TAPSE. **(C)** After intubation for rats described in **(A)**, RVSP was recorded, and the right ventricular hypertrophy ratio of RV/BW and RV/LV + S were calculated. **(D)** Hematoxylin and eosin staining of paraffin-fixed lung sections prepared from rats described in **(A)** was performed for morphological analysis of the pulmonary arteries. **(E)** A diagram showing that hypoxia-induced Mxi1-0 promotes PASMCs proliferation *via* MEK/ERK/c-Myc signaling in the context of HPH. Scale bar, 20 μm. Data are shown as means ± SDs. For statistical significance, ***represents $p < 0.001$ compared to hypoxia. PAAT, pulmonary artery acceleration time; TAPSE, tricuspid annulusplain systolic excursion; RVSP, right ventricular systolic pressure; RV, right ventricle; BW, body weight; LV + S, left ventricle plus septum.

might promote the transcription of the proto-oncogene c-Myc (Engstrom et al., 2004; Boult et al., 2008). Nonetheless, the detailed mechanisms underlying the role of Mxi1-0 in the division of PASMCs remain to be further dissected, e.g., whether Mxi1-0 and Mxi1-1 play opposite roles in the regulation of Max-Myc interaction or whether Mxi1-0 participates in the cellular machineries responsible for degradation of c-Myc.

MAPK signaling pathway is involved in the regulation of various biological functions of cells (Uehling and Harris, 2015). The activation of this pathway licenses the expression of a large cohort of genes regulating cell proliferation, differentiation and vascular development (Lei et al., 2015). MEK/ERK signaling is a classical MAPK signal transduction pathway and serves an important regulator of pulmonary hypertension (Preston et al., 2006). Consistently, we found here that HPH-related genes were significantly enriched in MAPK signaling pathway and ERK related biological processes. Mxi1-0, as a novel regulator of HPH, potentiates PASMC proliferation through MEK/ERK-dependent upregulation of c-Myc (**Figure 7E**). These findings are in agreement with previous studies showing that Mxi1-0 activates MEK/ERK signaling and improved the proliferation of human umbilical vein endothelial cells (HUVECs), and that Mxi1-0 underlies hypoxia-induced vascular endothelial growth factor production by hepatic carcinoma cells (Hu et al., 2017; Wu et al., 2017). Although it is still unknown how Mxi1-0 activates MEK/ERK signaling in PASMCs, we found that inhibition of this

canonical pathway significantly repressed the development of pulmonary hypertension in rats exposed to chronic hypoxia.

Collectively, we report a hitherto unrecognized crucial role of Mxi1-0 in HPH, thereby providing rationale for the applicability of Mxi1-0 and downstream signaling as candidate targets for clinical treatment and potential biomarkers for prognostic assessment of HPH.

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 author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Huashan institutional review board. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Fudan University Veterinary Medicine Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

SQL designed the experiment and prepared the initial manuscript. LD, XL, and BW conducted experimental operation. CL, GW, and JW cultivated the cells. XW, JX, and XZ were responsible for experimental calculations. YYZ, RY, NZ, JL, DZ, YZZ, SHL, and JZL were involved in the experimental analysis. LD, XL, and BW contributed equally to this manuscript. SQL was a Correspondent.

FUNDING

This work was supported by the National Natural Science Foundation (Nos. 81670045 and 81970048) and by the General Project of Shanghai Municipal Health Commission (No.201840296).

ACKNOWLEDGMENTS

We are grateful for the financial support by the National Natural Science Foundation and by the General project of Shanghai Municipal Health Commission.

REFERENCES

Armstrong, M. B., Mody, R. J., Ellis, D. C., Hill, A. B., Erichsen, D. A., and Wechsler, D. S. (2013). N-myc Differentially Regulates Expression of MXI1 Isoforms in Neuroblastoma. *Neoplasia* 15 (12), 1363–1370. PubMed PMID: 24403858; PubMed Central PMCID: PMC3884527. doi:10.1593/neo.11606

Badesch, D. B., Raskob, G. E., Elliott, C. G., Krichman, A. M., Farber, H. W., Frost, A. E., et al. (2010). Pulmonary Arterial Hypertension. *Chest* 137 (2), 376–387. PubMed PMID: 19837821. doi:10.1378/chest.09-1140

Baluapuri, A., Wolf, E., and Eilers, M. (2020). Target Gene-independent Functions of MYC Oncoproteins. *Nat. Rev. Mol. Cel Biol* 21 (5), 255–267. PubMed PMID: 32071436; PubMed Central PMCID: PMC7611238. doi:10.1038/s41580-020-0215-2

Boult, J. K. R., Tanière, P., Hallissey, M. T., Campbell, M. J., and Tselepis, C. (2008). Oesophageal Adenocarcinoma Is Associated with a Deregulation in the MYC/MAX/MAD Network. *Br. J. Cancer* 98 (12), 1985–1992. PubMed PMID: 18493233; PubMed Central PMCID: PMC2441969. doi:10.1038/sj.bjc.6604398

Cassady, S. J., and Reed, R. M. (2019). Pulmonary Hypertension in COPD: A Case Study and Review of the Literature. *Medicina* 55 (8), 432. PubMed PMID: 31382489; PubMed Central PMCID: PMC6723523. doi:10.3390/medicina55080432

Dang, C. V., O'Donnell, K. A., Zeller, K. I., Nguyen, T., Osthus, R. C., and Li, F. (2006). The C-Myc Target Gene Network. *Semin. Cancer Biol.* 16 (4), 253–264. PubMed PMID: 16904903. doi:10.1016/j.semcancer.2006.07.014

Dugast-Darzacq, C., Grange, T., and Schreiber-Agus, N. B. (2007). Differential Effects of Mxi1-Srα and Mxi1-Srβ in Myc Antagonism. *FEBS J.* 274 (17), 4643–4653. PubMed PMID: 17697116. doi:10.1111/j.1742-4658.2007.05992.x

Dugast-Darzacq, C., Pirity, M., Blanck, J. K., Scherl, A., and Schreiber-Agus, N. (2004). Mxi1-SRα: a Novel Mxi1 Isoform with Enhanced Transcriptional Repression Potential. *Oncogene* 23 (55), 8887–8899. PubMed PMID: 15467743. doi:10.1038/sj.onc.1208107

Dunham-Snary, K. J., Wu, D., Sykes, E. A., Thakrar, A., Parlow, L. R. G., Mewburn, J. D., et al. (2017). Hypoxic Pulmonary Vasoconstriction. *Chest* 151 (1), 181–192. PubMed PMID: 27645688; PubMed Central PMCID: PMC5310129. doi:10.1016/j.chest.2016.09.001

Engstrom, L. D., Youkilis, A. S., Gorelick, J. L., Zheng, D., Ackley, V., Petroff, C. A., et al. (2004). Mxi1-0, an Alternatively Transcribed Mxi1 Isoform, Is Overexpressed in Glioblastomas. *Neoplasia* 6 (5), 660–673. PubMed PMID: 15548375; PubMed Central PMCID: PMC1531670. doi:10.1593/neo.04244

Erichsen, D. A., Armstrong, M. B., and Wechsler, D. S. (2015). Mxi1 and Mxi1-0 Antagonize N-Myc Function and Independently Mediate Apoptosis in Neuroblastoma. *Translational Oncol.* 8 (1), 65–74. PubMed PMID: 25749179; PubMed Central PMCID: PMC4350643. doi:10.1016/j.tranon.2015.01.002

Galiè, N., Humbert, M., Vachiery, J.-L., Gibbs, S., Lang, I., Torbicki, A., et al. (2015. 2015 ESC/ERS Guidelines for the Diagnosis and Treatment of Pulmonary Hypertension. *Eur. Respir. J.* 46 (4), 903–975. PubMed PMID: 26318161. doi:10.1183/13993003.01032-2015

Hu, Z., Dong, N., Lu, D., Jiang, X., Xu, J., Wu, Z., et al. (2017). A Positive Feedback Loop between ROS and Mxi1-0 Promotes Hypoxia-Induced VEGF Expression in Human Hepatocellular Carcinoma Cells. *Cell Signal.* 31, 79–86. PubMed PMID: 28065785. doi:10.1016/j.cellsig.2017.01.007

Huang, Y., Hu, K., Zhang, S., Dong, X., Yin, Z., Meng, R., et al. (2018). S6K1 Phosphorylation-dependent Degradation of Mxi1 by β-Trcp Ubiquitin Ligase Promotes Myc Activation and Radioresistance in Lung Cancer. *Theranostics* 8 (5), 1286–1300. PubMed PMID: 29507620; PubMed Central PMCID: PMC5835936. doi:10.7150/thno.22552

Hurdman, J., Condliffe, R., Elliot, C. A., Davies, C., Hill, C., Wild, J. M., et al. (2012). ASPIRE Registry: Assessing the Spectrum of Pulmonary Hypertension Identified at a REferral centre. *Eur. Respir. J.* 39 (4), 945–955. PubMed PMID: 21885399. doi:10.1183/09031936.00078411

Hurlin, P. J., and Huang, J. (2006). The MAX-interacting Transcription Factor Network. *Semin. Cancer Biol.* 16 (4), 265–274. PubMed PMID: 16908182. doi:10.1016/j.semcancer.2006.07.009

Lei, Z., van Mil, A., Brandt, M. M., Grundmann, S., Hoefer, I., Smits, M., et al. (2015). MicroRNA-132/212 Family Enhances Arteriogenesis after Hindlimb Ischaemia through Modulation of the Ras-MAPK Pathway. *J. Cel. Mol. Med.* 19 (8), 1994–2005. PubMed PMID: 25945589; PubMed Central PMCID: PMC4549050. doi:10.1111/jcmm.12586

Li, D., Wang, B., Wang, H., and Liu, Q. (2018). Prognostic Significance of Pulmonary Hypertension in Patients with Cystic Fibrosis. *Medicine* 97 (7),

e9708. PubMed PMID: 29443734; PubMed Central PMCID: PMC5839836. doi:10.1097/MD.0000000000009708

Li, W., Liu, C.-C., Kang, S., Li, J.-R., Tseng, Y.-T., and Zhou, X. J. (2016). Pushing the Annotation of Cellular Activities to a Higher Resolution: Predicting Functions at the Isoform Level. *Methods* 93, 110–118. PubMed PMID: 26238263. doi:10.1016/j.ymeth.2015.07.016

Li, Y., Yang, L., Dong, L., Yang, Z.-w., Zhang, J., Zhang, S.-l., et al. (2019). Crosstalk between the Akt/mTORC1 and NF-Kb Signaling Pathways Promotes Hypoxia-Induced Pulmonary Hypertension by Increasing DPP4 Expression in PASMCs. *Acta Pharmacol. Sin* 40 (10), 1322–1333. PubMed PMID: 31316183; PubMed Central PMCID: PMC6786428. doi:10.1038/s41401-019-0272-2

Manni, I., Tunici, P., Cirenei, N., Albarosa, R., Colombo, B. M., Roz, L., et al. (2002). Mxi1 Inhibits the Proliferation of U87 Glioma Cells through Down-Regulation of Cyclin B1 Gene Expression. *Br. J. Cancer* 86 (3), 477–484. PubMed PMID: 11875718; PubMed Central PMCID: PMC2375210. doi:10.1038/sj.bjc.6600065

Maron, B. A., Machado, R. F., and Shimoda, L. (2016). Pulmonary Vascular and Ventricular Dysfunction in the Susceptible Patient (2015 Grover Conference Series). *Pulm. Circ.* 6 (4), 426–438. PubMed PMID: 28090285; PubMed Central PMCID: PMC5210067. doi:10.1086/688315

Naeije, R., and Dedobbeleer, C. (2013). Pulmonary Hypertension and the Right Ventricle in Hypoxia. *Exp. Physiol.* 98 (8), 1247–1256. PubMed PMID: 23625956. doi:10.1113/expphysiol.2012.069112

Negi, V., Yang, J., Speyer, G., Pulgarin, A., Handen, A., Zhao, J., et al. (2021). Computational Repurposing of Therapeutic Small Molecules from Cancer to Pulmonary Hypertension. *Sci. Adv.* 7 (43), eabh3794. PubMed PMID: 34669463. doi:10.1126/sciadv.abh3794

Nesbit, C. E., Tersak, J. M., and Prochownik, E. V. (1999). MYC Oncogenes and Human Neoplastic Disease. *Oncogene* 18 (19), 3004–3016. PubMed PMID: 10378696. doi:10.1038/sj.onc.1202746

Preston, I. R., Hill, N. S., Warburton, R. R., and Fanburg, B. L. (2006). Role of 12-lipoxygenase in Hypoxia-Induced Rat Pulmonary Artery Smooth Muscle Cell Proliferation. *Am. J. Physiology-Lung Cell Mol. Physiol.* 290 (2), L367–L374. PubMed PMID: 16199435. doi:10.1152/ajplung.00114.2005

Prins, K. W., Duval, S., Markowitz, J., Pritzker, M., and Thenappan, T. (2017). Chronic Use of PAH-specific Therapy in World Health Organization Group III Pulmonary Hypertension: a Systematic Review and Meta-analysis. *Pulm. Circ.* 7 (1), 145–155. PubMed PMID: 28680574; PubMed Central PMCID: PMC5448533. doi:10.1086/690017

Shah, S. J. (2012). Pulmonary Hypertension. *Jama* 308 (13), 1366–1374. PubMed PMID: 23032553. doi:10.1001/jama.2012.12347

Thompson, A. A. R., and Lawrie, A. (2017). Targeting Vascular Remodeling to Treat Pulmonary Arterial Hypertension. *Trends Molecular Medicine* 23 (1), 31–45. PubMed PMID: 27989641. doi:10.1016/j.molmed.2016.11.005

Uehling, D. E., and Harris, P. A. (2015). Recent Progress on MAP Kinase Pathway Inhibitors. *Bioorg. Med. Chem. Lett.* 25 (19), 4047–4056. PubMed PMID: 26298497. doi:10.1016/j.bmcl.2015.07.093

Voelkel, N. F., Mizuno, S., and Bogaard, H. J. (2013). The Role of Hypoxia in Pulmonary Vascular Diseases: a Perspective. *Am. J. Physiology-Lung Cell Mol. Physiol.* 304 (7), L457–L465. PubMed PMID: 23377344. doi:10.1152/ajplung.00335.2012

Wechsler, D. S., Shelly, C. A., Petroff, C. A., and Dang, C. V. (1997). MXI1, a Putative Tumor Suppressor Gene, Suppresses Growth of Human Glioblastoma Cells. *Cancer Res.* 57 (21), 4905–4912. PubMed PMID: 9354456.

Wechsler, D. S., Hawkins, A. L., Li, X., Jabs, E. W., Griffin, C. A., and Dang, C. V. (1994). Localization of the Human Mxi1 Transcription Factor Gene (MXI1) to Chromosome 10q24-Q25. *Genomics* 21 (3), 669–672. PubMed PMID: 7959753. doi:10.1006/geno.1994.1336

Woo, K. V., Ornitz, D. M., and Singh, G. K. (2019). Diagnosis and Pathophysiological Mechanisms of Group 3 Hypoxia-Induced Pulmonary Hypertension. *Curr. Treat. Options. Cardio Med.* 21 (3), 16. PubMed PMID: 30903302. doi:10.1007/s11936-019-0718-3

Wu, W., Hu, Z., Wang, F., Gu, H., Jiang, X., Xu, J., et al. (2017). Mxi1-0 Regulates the Growth of Human Umbilical Vein Endothelial Cells through Extracellular Signal-Regulated Kinase 1/2 (ERK1/2) and Interleukin-8 (IL-8)-dependent Pathways. *PloS one* 12 (6), e0178831. PubMed PMID: 28575053; PubMed Central PMCID: PMC5456372. doi:10.1371/journal.pone.0178831

Xia, J., Yang, L., Dong, L., Niu, M., Zhang, S., Yang, Z., et al. (2018). Cefminox, a Dual Agonist of Prostacyclin Receptor and Peroxisome Proliferator-Activated Receptor-Gamma Identified by Virtual Screening, Has Therapeutic Efficacy against Hypoxia-Induced Pulmonary Hypertension in Rats. *Front. Pharmacol.* 9, 134. PubMed PMID: 29527168; PubMed Central PMCID: PMC5829529. doi:10.3389/fphar.2018.00134

Yap, K., and Makeyev, E. V. (2016). Functional Impact of Splice Isoform Diversity in Individual Cells. *Biochem. Soc. Trans.* 44 (4), 1079–1085. PubMed PMID: 27528755; PubMed Central PMCID: PMC4984447. doi:10.1042/BST20160103

Young, J. M., Williams, D. R., and Thompson, A. A. R. (2019). Thin Air, Thick Vessels: Historical and Current Perspectives on Hypoxic Pulmonary Hypertension. *Front. Med.* 6, 93. PubMed PMID: 31119132; PubMed Central PMCID: PMC6504829. doi:10.3389/fmed.2019.00093

Zervos, A. S., Gyuris, J., and Brent, R. (1993). Mxi1, a Protein that Specifically Interacts with Max to Bind Myc-Max Recognition Sites. *Cell* 72 (2), 223–232. PubMed PMID: 8425219. doi:10.1016/0092-8674(93)90662-a

Zhang, C., Ma, C., Zhang, L., Zhang, L., Zhang, F., Ma, M., et al. (2019). MiR-449a-5p Mediates Mitochondrial Dysfunction and Phenotypic Transition by Targeting Myc in Pulmonary Arterial Smooth Muscle Cells. *J. Mol. Med.* 97 (3), 409–422. PubMed PMID: 30715622. doi:10.1007/s00109-019-01751-7

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: 18 May 2022
doi: 10.3389/fgene.2022.815082

A Potential Indicator ARRDC2 Has Feasibility to Evaluate Prognosis and Immune Microenvironment in Ovarian Cancer

Mengjun Zhang[1], *Yunduo Liu*[1], *Yuan Liu*[1], *Siyu Hou*[2], *Hao Li*[1], *Ying Ma*[1], *Can Wang*[1] *and Xiuwei Chen*[1]*

[1]*Department of Gynecology, Harbin Medical University Cancer Hospital, Harbin, China,* [2]*Department of Gynecology, Beijing Shijitan Hospital, Capital Medical University, Beijing, China*

Background: The abnormal expression of α-arrestin protein family plays a key regulatory role in the occurrence and development of many cancers, including colorectal cancer and cervical cancer, and is inseparable from changes in the tumor immune microenvironment. However, the role of ARRDC2, an important member of this family, in the malignant biological process of ovarian cancer (OC) has not been reported, and its role in the change of the immune microenvironment is also unknown.

Methods: In this study, HPA, TCGA, GEO and other databases were used to explore the role of ARRDC2 in the prognosis assessment of ovarian cancer. Then, GO, KEGG analysis and GSEA analysis of the biological processes and cell signaling pathways that ARRDC2 may be involved in activated or inhibited. In addition, the TIMER and TISIDB database were used to conduct in-depth research on the role of ARRDC2 in the change of the immune microenvironment of ovarian cancer. The CMap database explored and screened drugs that may be used for treatment. Through cell transfection, CCK-8, Ki-67 immunofluorescence, wound healing, transwell and clone formation assay, the effect of ARRDC2 knockdown on the malignant biological behavior of OC cells were explored.

Results: There were significant differences between OC and ARRDC2 mRNA and protein levels. High ARRDC2 expression level is associated with poor overall survival and can be used as an independent prognostic factor. Interestingly, ARRDC2 expression is positively correlated with B cells, Neutrophils, Dendritic cells and CD8+ T cells, signifying that ARRDC2 may be related to infiltration of immune cells. ARRDC2 and its co-expressed genes are enriched in cell signaling pathways related to the immune system. We explored two possible drugs for the treatment of ovarian cancer. Finally, the results of *in vitro* experiments indicated that knockdown of ARRDC2 may inhibit malignant phenotypes such as proliferation and migration of OC cells.

Conclusion: The differentially expressed ARRDC2 may be a potential prognostic indicator and can be used as a novel biomarker for exploring the immune microenvironment of ovarian cancer.

Keywords: arrestin domain containing 2, biomarker, immunity, prognosis, ovarian cancer

OPEN ACCESS

Edited by:
Mee-Hyun Lee,
Dongshin University, South Korea

Reviewed by:
Chel Hun Choi,
Samsung, South Korea
Xiaoying Xue,
Second Hospital of Hebei Medical University, China

***Correspondence:**
Xiuwei Chen
1427@hrbmu.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *15 November 2021*
Accepted: *02 May 2022*
Published: *18 May 2022*

Citation:
Zhang M, Liu Y, Liu Y, Hou S, Li H, Ma Y, Wang C and Chen X (2022) A Potential Indicator ARRDC2 Has Feasibility to Evaluate Prognosis and Immune Microenvironment in Ovarian Cancer. Front. Genet. 13:815082. doi: 10.3389/fgene.2022.815082

IMPACT STATEMENT

Ovarian cancer is one of the most common malignant tumors in female reproductive system. Recently, ARRDC gene family is shown in many studies to play a critical role in tumor growth and invasion, which may shed light on study on OV. ARRDC2 is a member of the ARRDC gene family and has been suggested to be involved in various cellular processes of tumorigenesis and progression, such as invasion, migration, proliferation, transformation, and survival. However, little is known about the significance of ARRDC2 in ovarian cancer. Our research may provide new ideas and directions for the diagnosis and treatment of ovarian cancer. ARRDC2 may provide a new development direction for gene targeted therapy and immunotherapy of gynecological tumors. The discovery of a new and effective specific biomarker is of great significance not only for the medical research field of gynecological tumors, but also for the entire medical research field.

INTRODUCTION

In 2018, there were approximately 295,000 new cases of OV (ovarian cancer) and 185,000 deaths in the world. As this tumor is asymptomatic during initial progression and there are no clear early screening methods, it is usually diagnosed at the advanced stage, resulting in an overall 5-years survival rate of less than 40% (Whiteman and Wilson, 2016). The rapid development of high-throughput sequencing technology and transcription research is expected to increase the early diagnosis rate. Although a large number of new proto-oncogenes and tumor suppressor genes that can be used for diagnosis have been discovered, the survival results of ovarian cancer have not been greatly improved. And in terms of treatment, immunotherapy has evolved rapidly over the past 20 years, giving patients with ovarian cancer, known as "immunogenic tumors," more access to treatment (Santoiemma et al., 2016). However, the response rate of ovarian cancer patients to existing immunotherapy is not satisfactory. Obviously, the understanding of the tumor immune microenvironment of ovarian cancer is still insufficient, and more in-depth research on it and finding specific genes that potentially affect the tumor immune microenvironment and can be used as immunotherapy targets can help improve this situation. In short, it is of great significance to find biomarkers that may be used in early diagnosis and immunotherapy and try to explore their mechanisms.

The mammalian α-arrestin family consists of five structural domain-containing arrestin proteins (ARRDC1-5) and TXNIP. Arrestin domain containing 2 (ARRDC2) is an enigmatic member of the arrestin protein family that plays an important role in the regulation of G protein-coupled receptors (GPCRs) (Dores et al., 2015; Tian et al., 2016). Numerous recent studies have established a link between α-arrestin family and cancer. ARRDC3 and TXNIP were considered to be tumor suppressor genes that regulate a variety of cellular processes. For instance, ARRDC3 was decreased in prostate cancer, breast cancer and colorectal cancer (Huang et al., 2012; Rafiq et al., 2013). However, tumor relevance studies of ARRDC2 have not been reported. Given that members of the arrestin protein family play a momentous role in the biology of tumors, the function of ARRDC2 in tumors, especially in ovarian cancer, has attracted great interest to us.

We are the first study to investigate the important role of the ARRDC2 of the arrestin protein family in the occurrence, development and poor prognosis of ovarian cancer. Here, we performed a deep dive into the TCGA database and the GEO database to determine the impact of ARRDC2 on the progression and poor prognosis of ovarian cancer. The interrelationships between ARRDC2 and immune cell infiltration, immune checkpoints and chemokines were also explored by the Timer database and TISIDBD database. Subsequently, these data were correlated with the clinical regression and prognosis of OV patients. The results showed that high expression in OV and affected the clinical prognosis of patients. Excitingly, a close correlation was found between ARRDC2 expression and the tumor immune microenvironment of the tumor including infiltration of immune cells, immune checkpoint and chemokines. We sought to explore the cellular signaling pathways associated with ARRDC2 and potential small molecule drugs. Finally, the effect of ARRDC2 on the malignant phenotype of OC was confirmed by knockdown of ARRDC2 in OC cell lines. Overall, the results of multiple high-throughput data and a series of rigorous *in vitro* experiments confirmed that ARRDC2 might drive the malignant biological behavior of OC cells. In conclusion, this study attempts to explore potential as a new immune-related prognostic biomarker for OV patients, which may open up a new approach for the combination of immunotherapy and gene therapy for OV patients.

MATERIALS AND METHODS

Data Collection

The Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/) is a world-recognized data-rich public platform, and the public sequencing data in this database have contributed significantly to oncology research. After searching and screening, three data sets (GSE29450, GSE10971, and GSE19829) containing gene expression data were selected. Among them, GSE19829 additionally includes prognostic information such as overall survival time. Microarray data from the GSE29450 (OV = 10, Normal = 10) and GSE10971 (OV = 13, Normal = 24) datasets from GEO were used to study ARRDC2 gene in OV and in normal control. The GSE19829 (OV = 28) data set containing survival information was used in a survival meta-analysis related to the expression level of ARRDC2 by combining with the survival information of TCGA (OV = 372). The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database has a large amount of transcriptomic data such as gene expression data and DNA methylation data. Such a powerful database of massive information has largely improved molecular research of tumor. Therefore, transcriptome data, methylation data and corresponding clinical data of 372 OV patients were collected from the TCGA database. These data were

then used to explore the expression level of ARRDC2 and its relationship with specific clinical features and prognosis. At the same time, the relationship between ARRDC2 and changes in methylation sites was also explored.

Cell Culture and Cell Transfection

Ovarian cancer cell lines (SKOV3 and A2780) and corresponding normal cell lines (IOSE80) were provided by Shanghai Sun Ran HAKATA Cell Bank (http://www.xrshbio.com/). The cells were cultured at 37°C in a 5% CO_2 incubator using DMEM medium containing 10% fetal bovine serum (FBS. Gibco), 100 U/mL penicillin and 0.1 g/L streptomycin in DMEM medium. When the cells proliferated to about 80%–90% of the bottom of the culture vessel, the cells were passaged and isolated by digestion with 0.25% trypsin. Cells were seeded in 6-well medium plates and 100 pmol siRNA was transfected into each plate using siRNA-Mate (GenePharma, Shanghai,China) following the product instructions standard procedure, and detected the transfection efficiency by real-time quantitative polymerase chain reaction (RT-qPCR) at 24 h after transfection and Western blot at 48 h after transfection. GAPDH was set as an internal reference and its primer sequences were as follows: (GAPDH-F:5′-CAAGGTCATCCATGACAACTTTG-3′, GAPDH-R:5′-GTCCACCACCCTGTTGCTGTAG-3′). The primer sequences of ARRDC2 were as follows: (ARRDC2-516-F:5′-GUGUCCGCUACUGUAUCAATT-3′, ARRDC2-516-R:3′-UUGAUACAGUAGCGGACACTT-5′, ARRDC2-149-F:5′-GACAAGGGUGAAAGCGUUCUTT-3′, ARRDC2-149-R:3 ′-AGAACGCUUUCACCUUGUCTT-5′). An empty sequence was constructed as a control (si-NC). 24 h after transfection, the complete medium was replaced according to the instructions, and the knockdown efficiency experiment verified that the ARRDC2-516 sequence was effective and was used for subsequent experiments.

RT-qPCR

Expression of ARRDC2 in human ovarian cancer cells was detected using RT-qPCR. Total RNA was extracted from the cells using Total RNA Kit I kit (Omega Biotek). RNA reverse transcription was performed under the guidance of NovoScript Plus All-in-one 1st Strand cDNA Synthesis SuperMix (Novoprotein) was performed. The relative expression levels of ARRDC2 were determined by RT-qPCR using NovoStart SYBR qPCR SuperMix Plus (Novoprotein) kit, and GAPDH was used as an internal reference control for ARRDC2 using the $2^{-\Delta\Delta Ct}$ method. The primer sequences of GAPDH and ARRDC2 were as follows: (GAPDH-F: 5′-CAAGGTCATCCATGACAACTTTG-3′, GAPDH-R: 5′-GTCCACCACCCTGTTGCTGTAG-3′, ARRDC2-F: 5′- CCCGATCCTGGTACTGTAACC-3′, ARRDC2-R: 5′- CGTTGTCGATCTCGGCAAAGA-3′). The thermal cycling conditions were as follows: Initial denaturation at 95°C for 10 min, denaturation at 95°C for 10 s, annealing and extension at 60°C for 30 s, for a total of 40 cycles.

Survival Meta-Analysis

A systematic search in large authoritative databases (such as PubMed and Web of Science) did not reveal any previous studies on the carcinogenicity and poor prognosis of ARRDC2. Therefore, this study combined data from two datasets (GSE19829 and TCGA RNA sequences) in a survival meta-analysis to reveal the prognostic significance of ARRDC2 on OV for the first time. The heterogeneity between studies was assessed by Q test (I^2 statistics). The fixed effects model is applicable when there is no heterogeneity or $I^2 < 50\%$. Otherwise, a random-effects model was applied. The random effects model was applied to this study according to the specific situation.

ARRDC2 Related Gene Enrichment Analysis

Go (gene ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis is carried out using David's online tool. In short, the list of genes that are positively and negatively related to ARRDC2 obtained through Pearson analysis based on TCGA data was uploaded to the DAVID database for Go and KEGG pathway enrichment analysis. In addition, we applied the "cluster Profiler" R package for GO enrichment analysis. We used Spearman correlation analysis to describe the correlation between quantitative variables without normal distribution. *p*-values less than 0.05 were considered statistically significant. Gene Set Enrichment Analysis (GSEA) is an analytical tool for analyzing cellular signaling pathways developed jointly by MIT and Harvard University. The RNA sequencing data from TCGA is batch-corrected and normalized, and then divided into "H group" (ARRDC2 high expression group) or "L group" (ARRDC2 low expression group). Enrichment analysis was performed using GSEA software (version 4.0.3). The number of changes was set to 1,000 and the genomic database was set to Kyoto Encyclopedia of Genes and Genomes (KEGG) cell signaling pathway ($p < 0.05$ was considered as significantly enriched).

Immune Databases (TIMER and TISIDB)

The Tumor Immunology Estimation Resource (TIMER; https://cistrome.shinyapps.io/timer) is a rich tumor immunology and genetics database available for automated analysis and visualization of data from the TCGA database of 10,897 pan-cancer samples. Firstly, we analyzed the correlation of ARRDC2 expression with the abundance of six immune cell types (neutrophils, CD4+ T cells, B cells, dendritic cells, CD8+ T cells and macrophages) in OV using the TIMER algorithm. Secondly, we also explored the prognostic value of ARRDC2 in OV patients with different immune cell abundance. In addition, we examined the impact of ARRDC2 gene copy number alterations on immune infiltration. Finally, we explored the co-expression relationship of ARRDC2 with common immune checkpoint-encoding genes including PD-1 (PDCD1), PDL1 (CD274) and PDL2 (PDCD1LG2) in tumor purity-corrected state. Then, TISIDB (http://cis.hku.hk/TISIDB/index.php) were used for verification and analysis the relationship between ARRDC2 and immune infiltrating cells and immune checkpoint. To further explore the relationship between ARRDC2 and immune checkpoints, TISIDB database was mined and validated.

Immune-Related Kaplan-Meier Survival Analysis

The prognostic value of different expression levels of ARRDC2 is analyzed through this database. Using this database, overall survival (OS) of OV patients were analyzed at different immune cell infiltrations. Patient samples were divided into

high and low expression groups according to ARRDC2 gene expression levels and evaluated using Kaplan-Meier survival plots (p value < 0.05, false positive rate < 0.05). Hazard ratio (HR) had 95% confidence intervals and log-rank p values.

Co-Expression Analysis and Cmap Analysis

Co-expression analysis was performed by Pearson method, and 10 genes positively and negatively correlated with ARRDC2 were obtained based on correlation coefficients and p-values for constructing the correlation between genes and genes. Subsequently, the obtained co-expressed genes were used for the screening and prediction of small molecule drugs by the Connectivity Map (CMap, https://portals.broadinstitute.org/cmap/) database constructed by Prof. Lamb et al. Finally, 2D and 3D structural maps and chemical formulas of the drugs were obtained in the PubChem database.

Western Blotting

Western blot was used to determine the expression level of ARRDC2 and the efficiency of knockdown. Cell lysis was performed using a lysis solution containing 1% protease inhibitor. After complete lysis on ice for 5 min, the cells were centrifuged for 15 min using a 4°C centrifuge at 12,000 g and the supernatant was immediately collected. The corresponding protein concentrations were quantified using the BCA method (Thermo Fisher Scientific, Waltham, United States). Protein electrophoresis separation was performed using 10% SDS-PAGE gels. The proteins in the gels were then transferred to methanol-treated PVDF membranes. 5% skim milk was used to seal the gels for 1 h. ARRDC2 and GAPDH primary antibodies were mixed with PVDF membranes and incubated overnight in a refrigerator at 4°C. PVDF membranes were washed three times and mixed with horseradish peroxidase-conjugated rabbit secondary antibodies and incubated at room temperature for 1 h. After that, the proteins were washed again with TBST. After that, the membranes were washed again 3 times with TBST. Finally, the membrane was developed with ECL developer and imaged using a charge-coupled camera LAS4000 (Fujifilm, Tokyo, Japan). The grayscale values of the strips were measured by ImageJ software (version v.1.52).

CCK8 Assay

The cell proliferation ability after transfection was studied by CCK8 (Cell Counting Kit-8). The cell suspension was evenly plated in a 96-well plate in a volume of 100 μl per well. After the cells adhered, 10 μl of CCK-8 reagent (Yeasen, Shanghai, China) was added to the culture medium, and the microplate reader was performed after incubation for 2 h. Set the detection time as 0, 12, 24, and 48 h respectively. OD (optical density) was measured at a wavelength of 450 nm using a microplate reader (Thermo Fisher Scientific, Waltham, United States).

Clone Formation Assay

Plate the transfected cell suspension evenly in a 6-well plate at 200 cells/well. Cells in 6-well plates were then grown in complete medium. On the 4th day, the medium could be replaced with fresh medium, and cell colony formation was observed after 10 days. Fix cell colonies with 4% paraformaldehyde for 30 min. Cell colonies were stained with 1% crystal violet for 5 min. Formed cell colonies were observed and imaged under a microscope. The number of colonies formed was measured and counted by ImageJ (version 1.52).

Ki-67 Cell Immunofluorescence

Cells were added evenly in 24-well plates, and when the cell density reached 50–60%, 4% paraformaldehyde was used to fix the cells. Cell permeabilization was performed with 0.5% Triton X-100 for 20 min. After washing with PBS, block with 10% goat serum for 1.5 h at 37°C. Primary antibody (Ki-67, 1:200) was added to each well and incubated overnight at 4°C in the dark. Add secondary antibody and incubate at 37°C for 1 h in the dark. Finally, nuclei were stained with DAPI and incubated in the dark for 10 min. The images were observed under a fluorescence microscope and randomly collected from five fields of view. The numbers of viable and proliferating cells were measured and recorded using ImageJ software (version 1.52).

Wound-Healing Assay

Transfected ovarian cancer cells were seeded in 6-well plates at the appropriate density. A 200 μl sterile pipette tip was used to make a straight scratch on each well, followed by three washes with 1 × PBS to remove detached cells. Change complete medium to fresh serum-free medium. Scratches at the same site were photographed with a microscope at 0 and 24 h, respectively. Statistical analysis of wound healing rates was performed using ImageJ (version 1.52).

Transwell Assay

100 μl of the transfected ovarian cancer cell suspension was seeded into the upper chamber of a transwell plate (Corning Costar, Shanghai, China). Add 600 μl of minimal essential medium and 20% FBS to the lower chamber of the transwell plate. Transwell plates were incubated for 24 h at 37°C and 5% CO_2. Cells that crossed the membrane into the lower compartment were then fixed in 95% ethanol for 15 min. Stain with 0.1% crystal violet for 15 min. Then use a cotton swab to wipe the upper chamber without passing through the cells. Use a microscope to observe the cells and take pictures. Measure and record the number of migrating cells using ImageJ (version 1.52).

Statistical Analysis

Statistical data analysis was performed using R software (version 3.6.1). Survival and clinicopathological characteristics data were obtained from the TCGA database and the GEO database. Then, the overall survival of ARRDC2 was determined by Kaplan-Meier method. Univariate COX and multivariate COX analyses were used to analyze the factors affecting the prognosis of patients with OV. Student's t-test, Kruskal Wallis test, and Wilcoxon signed-rank test was widely used to compare statistical indicators (p < 0.05 was considered to be statistically significant).

RESULTS

Correlation Between ARRDC2 Expression and Clinical Characteristics of OV

ARRDC2 has abnormally high expression in a variety of human malignant tumors (**Figure 1A**). The analysis of a total of 34 normal ovarian tissues and 23 ovarian cancer tissues from two GEO datasets (GSE29450 and GSE10971), revealed that the

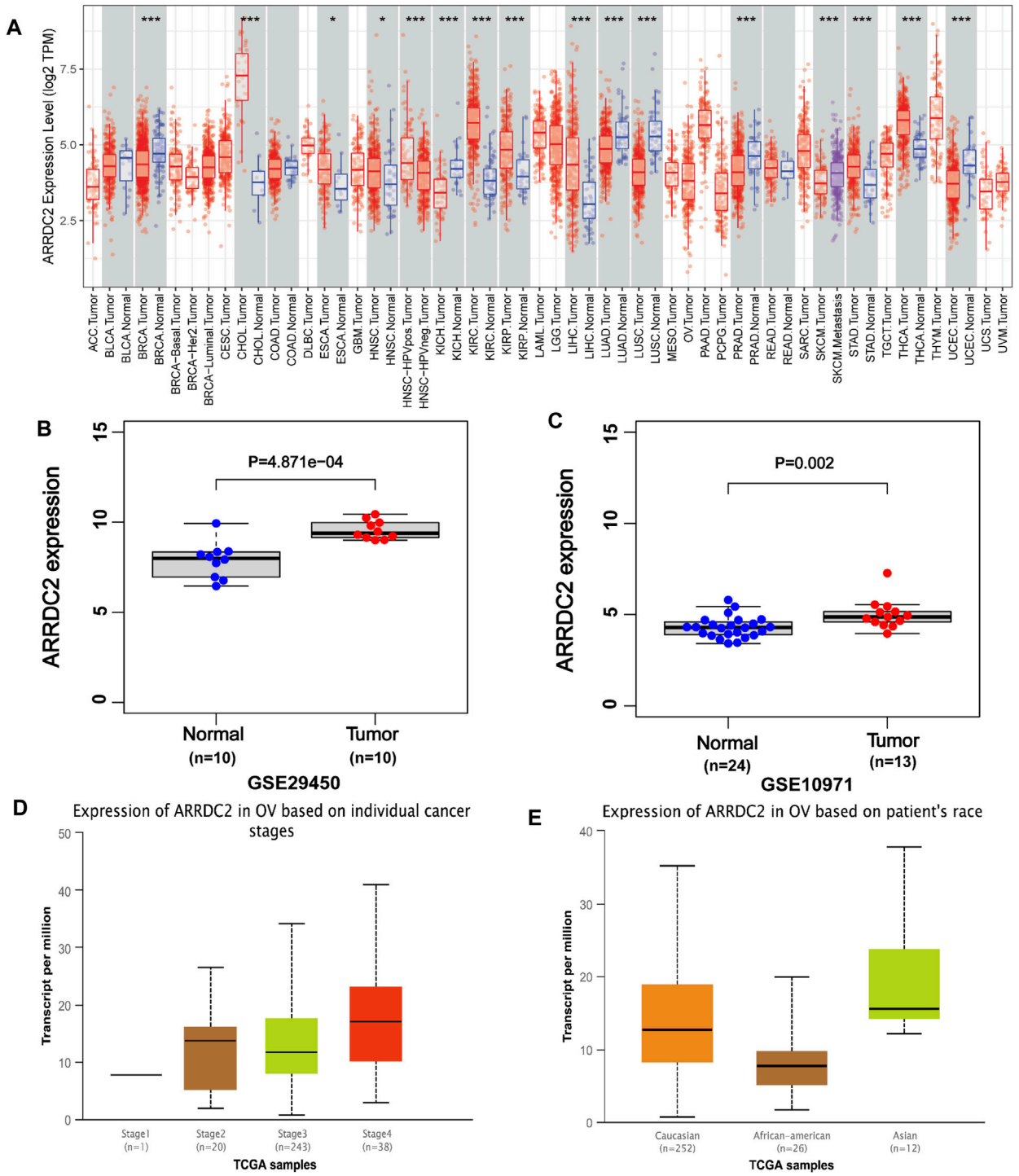


FIGURE 1 | The expression of ARRDC2 (mRNA, gene microarray and gene sequencing) in OVs. **(A)** Expression of ARRDC2 in pan-cancer. **(B)** Box plot based on the expression level of ARRDC2 in the GSE29450 (OV = 10, Normal = 10). **(C)** Box plot based on the expression level of ARRDC2 in the GSE10971 (OV = 13, Normal = 24). **(D)** The expression level of ARRDC2 in OV based on individual FIGO stage. **(E)** The expression level of ARRDC2 in OV based on the race of patient.

ARRDC2 expression was high and statistically significant in tumor tissues (**Figures 1B,C**). RT-qPCR was performed to verify the results of the above analysis. In addition, the relationship between ARRDC2 expression and clinical characteristics in 361 tumor samples from TCGA in the UALCAN database was explored. Correlation analysis showed that the expression of ARRDC2 was mainly positively correlated with FIGO stage and race (**Figures 1D,E**). It can be seen that

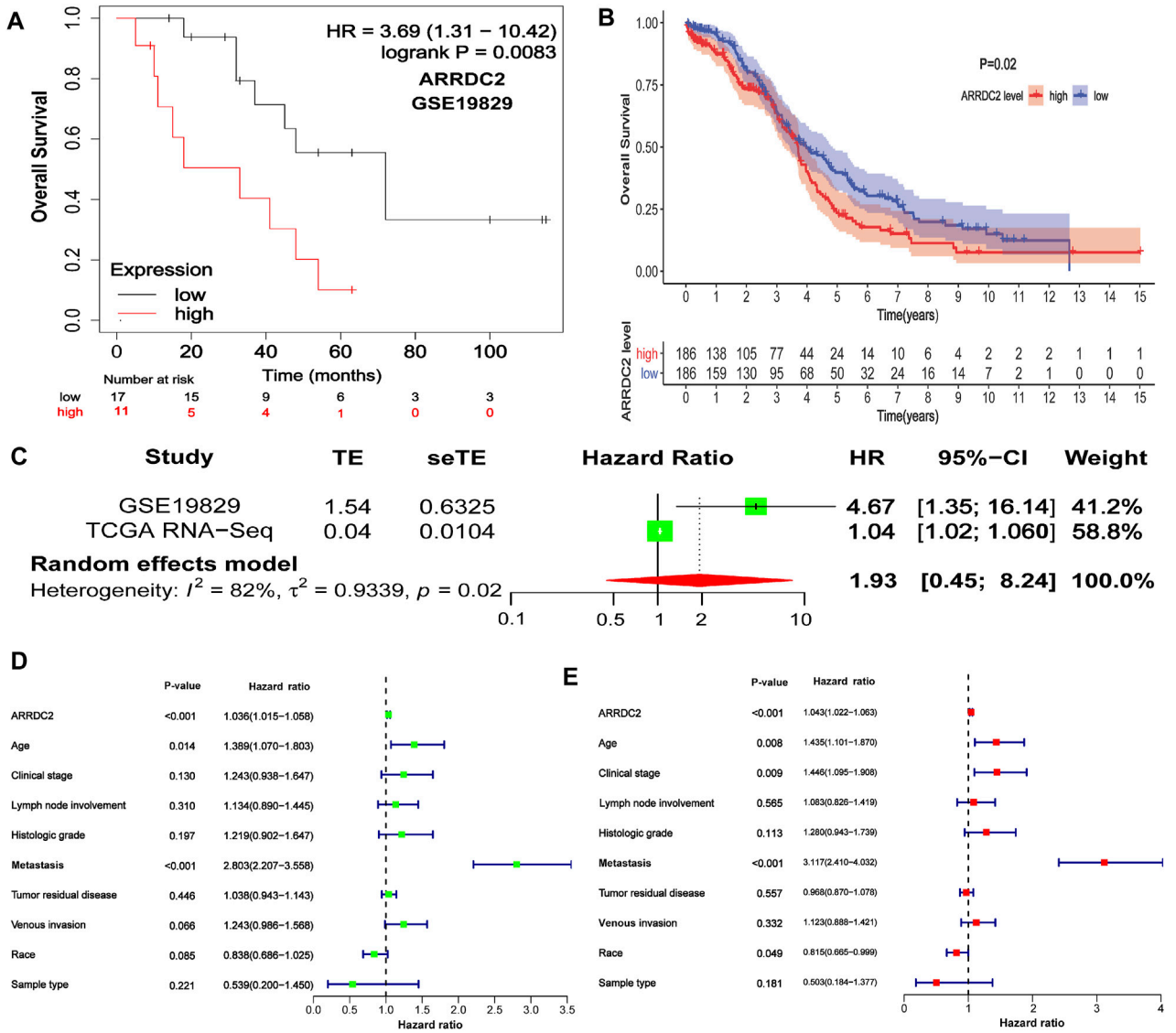


FIGURE 2 | The correlation between ARRDC2 and the poor prognosis of patients with OV. **(A)** The Kaplan-Meier survival curve revealed that the high expression of ARRDC2 lead to a poor prognosis in OVs (GSE19829, N = 28, p = 0.0083). **(B)** The Kaplan-Meier survival curve revealed that the high expression of ARRDC2 lead to a poor prognosis in OVs (TCGA, N = 372, p = 0.02). **(C)** Forest plot of high ARRDC2 expression with poor OS in OV patients based on survival meta-analysis of two datasets (GSE19829 and TCGA RNA-Seq, p = 0.02). **(D,E)** Analysis of univariate and multivariate factors affecting the prognosis of patients with OV. **(D)** Univariate cox analysis. **(E)** Multivariate cox analysis.

ARRDC2 expression was higher in patients with advanced FIGO stages (III and IV) than early FIGO stages (I and II), detailed clinical features are shown in **Supplementary Table S1**. In addition, we also explored the differences in ARRDC2 mRNA levels expression between different race groups, and ARRDC2 expression was significantly higher in Asian race than in African American race. In conclusion, our study explored up that ARRDC2 was extremely high expressed in ovarian cancer tumor in TCGA and GEO databases and was closely related to important clinical factors such as FIGO stage. Therefore, further studies on ARRDC2 are needed to explore its value in OV.

The Prognostic Value of ARRDC2 in OV Patients

To further explore the prognostic value of ARRDC2 for patients, Kaplan-Meier survival analysis was used to assess the relationship between ARRDC2 expression levels and OS. The KM survival analysis of 372 samples from TCGA showed that high ARRDC2 expression was associated with shorter overall survival in OV patients, p = 0.02 (**Figure 2B**). The GEO data also obtained the same results (**Figure 2A**). Univariate Cox analysis showed that the high expression of ARRDC2 (HR = 1.036; 95% CI = 1.015–1.058; p < 0.001), age (HR = 1.389;

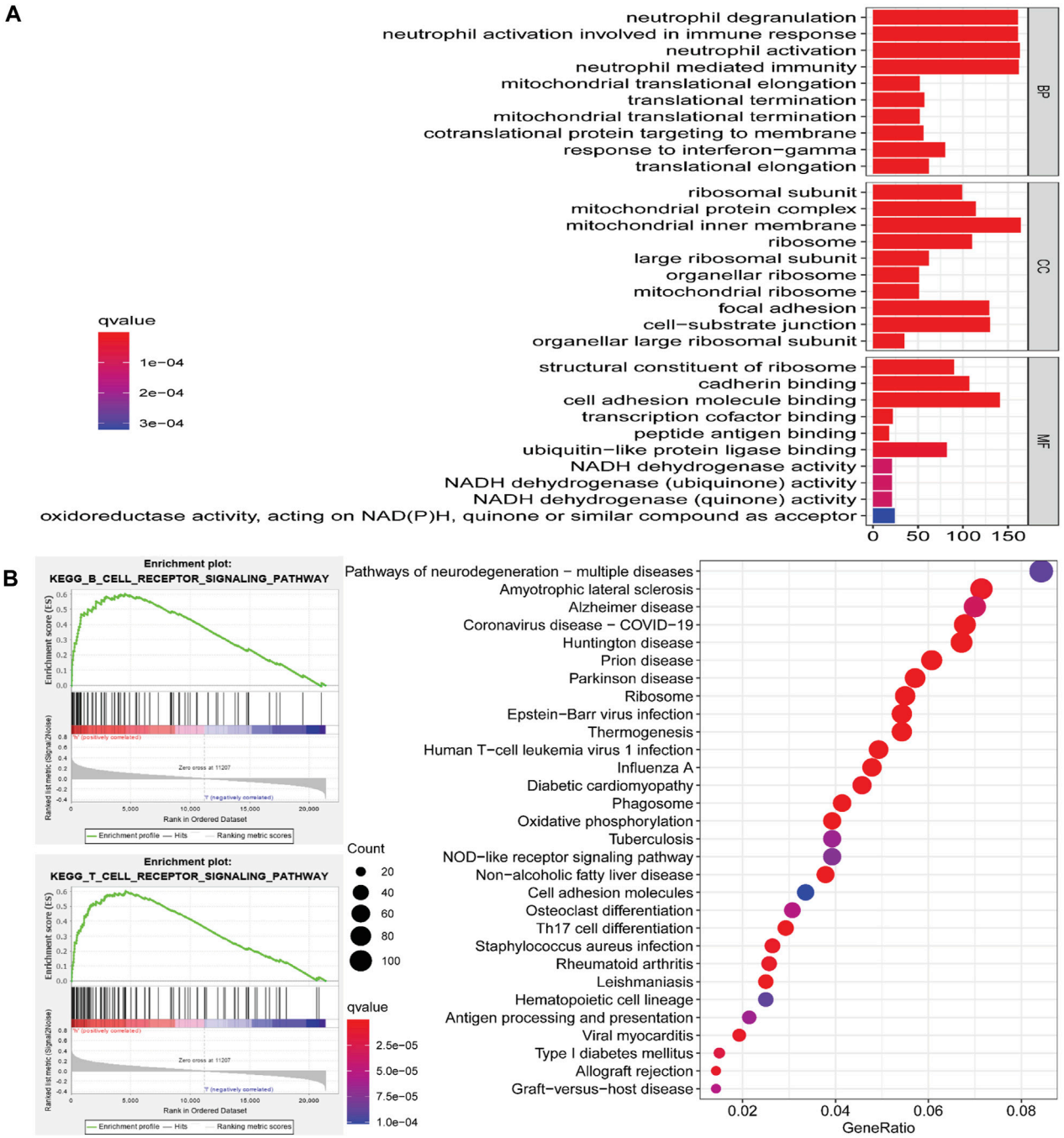


FIGURE 3 | Go (Gene ontology) functional annotation and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of ARRDC2 in OV. **(A)** Go functional annotation. Biological Process, BP (neutrophil degranulation, neutrophil mediated immunity, etc.). Cellular Component, CC (ribosomal subunit, mitochondrial protein complex, ribosome, large ribosomal subunit, etc.). Molecular Function, MF (structural constituent of ribosome, cadherin binding, transcription cofactor binding, etc.). **(B)** KEGG pathway enrichment analysis (B cell receptor signaling pathway, T cell signaling pathway, Human T-cell leukemia virus 1 infection signaling pathway, NOD-Like receptor signaling pathway, Th17 cell differentiation, etc.). GSEA enrichment analysis results of ARRDC2 (B cell receptor signaling pathway and T cell signaling pathway).

95% CI = 1.070–1.803; $p = 0.014$) and metastasis (HR = 2.803; 95% CI = 2.207–3.558; $p < 0.001$) were high risk factors (**Figure 2D**). Multivariate Cox analysis showed that the high expression of ARRDC2 (HR = 1.043; 95% CI = 1.022–1.063; $p < 0.001$), age (HR = 1.435; 95% CI = 1.101–1.870; $p = 0.008$) and person neoplasm cancer status (HR = 3.117; 95% CI = 2.410–4.032; $p < 0.001$) were high risk factors (**Figure 2E**). Overall, it is not difficult to see that ARRDC2 as an

independent prognostic risk factor is abnormally high in OV patients.

Survival Meta-Analysis

Although we have explored the impact of ARRDC2 on the survival outcome of OV patients, to increase the credibility and scientific validity of this study, we collected different mRNA expression data (GSE19829 and TCGA RNA-Seq) for meta-analysis from two datasets, which contained a total of 400 samples. The results showed that high expression of ARRDC2 was a risk factor in patients with OV (HR = 1.93; 95% CI = 0.45–8.24, $p = 0.02$) (**Figure 2C**). In summary, it can be seen that ARRDC2 can be used as a good biomarker for predicting the overall survival of OV patients.

Functional Annotation and Signaling Pathway Enrichment Analysis of ARRDC2

To further explore the potential molecular mechanisms of ARRDC2 in tumorigenesis of OV, we attempted to screen a series of pathways and biological functions by co-expressed genes of ARRDC2. The results of Go function annotation analysis are shown in **Figure 3A**. Among them, those enriched in biological processes included neutrophil degranulation and neutrophil-mediated immunity; those enriched in cellular components included ribosomal subunits, mitochondrial protein complexes, ribosomes and large ribose subunits; those enriched in cellular components included ribosomes, cadherin binding and transcription cofactor binding. The major enriched signaling pathways include B cell, T cell, Human T-cell leukemia virus 1 infection, NOD-Like receptor and Th17 cell differentiation. Importantly, we found that both ARRDC2 enriched functions and signaling pathways are closely related to immunity. In order to further verify the enrichment of ARRDC2 in the immune-related signaling pathways of ovarian cancer, GSEA was used to analyze the data of two groups of OV patients from the TCGA database (ARRDC2 high expression group and ARRDC2 low expression group). The result showed that B cell receptor signaling pathway and T cell signaling pathway were the most important enriched signaling pathways (FDR < 0.25, $p < 0.05$) (**Figure 3B**). In conclusion, ARRDC2 may affect the malignant progression and poor survival outcomes of patients with ovarian cancer through immunomodulatory effects.

Relationships of ARRDC2 With Tumor Immune Infiltration

In this study, eight types of infiltrating immune cells in the TIMER database were used to evaluate the relationship between ARRDC2 expression and immunity. The expression level of ARRDC2 was positively related to the infiltration of CD8+ T cells, neutrophils, B cells and dendritic cells (**Figures 4A,B**). However, there was no significant correlation between ARRDC2 expression and CD4+ T cells and Macrophage. In addition, the SCNA module was chosen to analyze the relationship between the somatic copy number alteration of ARRDC2 and different immune cell infiltrations. As shown in **Figure 4C**, the somatic copy number alteration of ARRDC2 correlated with infiltration of CD8+ T cells, neutrophils, B cells and dendritic cells. Meanwhile, numerous studies have confirmed that inhibition of the immune checkpoint pathway is an auspicious therapeutic pathway for the induction of effective anti-cancer immunity. Therefore, we analyzed the correlation between the expression levels of ARRDC2 and genes encoding immune checkpoints (**Figure 4D**), such as PD1 (PDCD1), PDL1 (CD274), PDL2 (PDCD1LG2) and CTLA4. The results showed that the expression levels of ARRDC2 were positively correlated with PD1, PDL1, PDL2, and CTLA4. As mentioned previously, the expression of ARRDC2 was closely correlated with the level of immune infiltration and positively correlated with immune checkpoint.

To further verify our speculation, the correlation between ARRDC2 expression and immune infiltration was explored by the TISIDB immune database, and the results were consistent with the TIMER database (**Figure 5A**). In addition to exploring the relationship between the ARRDC2 gene and immune cell infiltration as well as immune checkpoint using the TISIDB database, we also investigated MHC and chemokines. The results showed (**Figures 5B,C**) that the ARRDC2 was positively associated with MHC-related genes (B2M, HLA-DMA, HLA-DPA1, HLA-DRA, HLA-DRB1, and HLA-E) and chemokine-related genes (CCL17, CCL13, CCL5, CCL3, CCL4, and CX3CL1). Previous part of this study has shown that ARRDC2 was an independent influence on poor prognosis in OV patients. Thence, we hypothesized that ARRDC2 may affect the prognosis of OV patients partly due to immunological aspects. The Kaplan-Meier plotter database was used to verify the results of the study that the high expression of ARRDC2 during immune cell infiltration leads to a reduction in the overall survival of patients with ovarian cancer. As seen in **Figures 6A–I**, in the presence of different immune cell infiltration (B Cell, BASOPHILS, CD4+ T Cell, CD8+ T Cell, Eosinophils, Neutrophil, Macrophage, Mesenchymal stem cell, Natural killer T-cell and Th1cell infiltration), patients with high expression of ARRDC2 had shorter OS than those in the low expression group ($p < 0.05$).

Co-Expression Analysis and Drug Prediction of ARRDC2

To better understand the function of ARRDC2, we used co-expression analysis to determine the association of ARRDC2 with other genes. The top five positively and negatively associated genes with ARRDC2 were calculated as shown in the circular plot (**Figure 7A**). The results showed that ARRDC2 was negatively associated with ZNF22, HDGFL3, H2AFY2, SPINDOC, and MSI1, while positively correlated with TRPM2, FCGR2C, FCGR1CP, MIR3671, and RGS1 (**Figure 7B**). Subsequently, the co-expressed gene data was used to screen potential gene therapy drugs through the CMap and Pubchem databases. We predicted two possible gene therapy drugs for ARRDC2: Mercaptopurine and Apigenin (**Figures 7C,D**).

Expression of ARRDC2 Was Negatively Regulated by DNA Methylation

DNA methylation is one of the most intensively studied epigenetic modifications in mammals, and the importance of

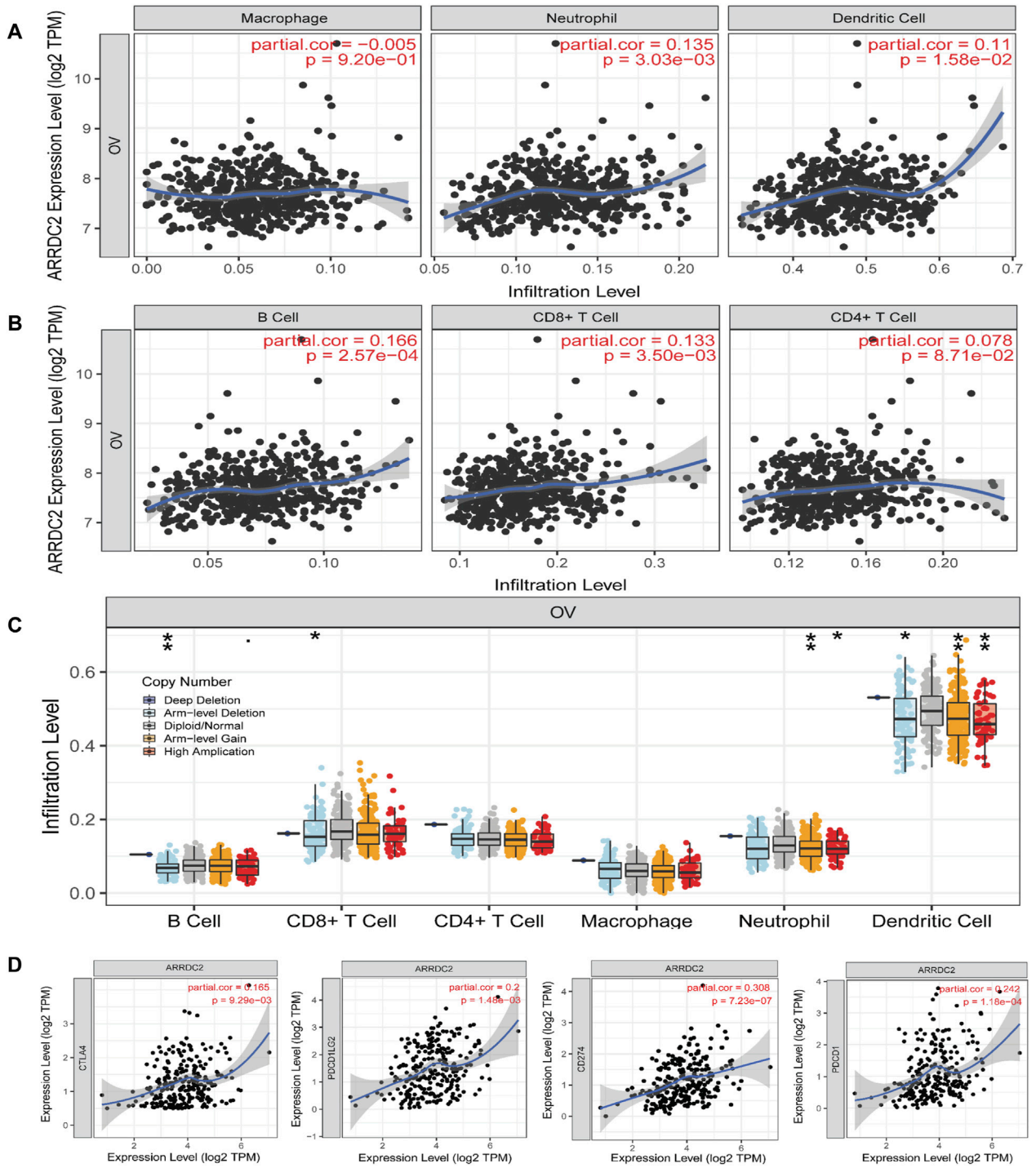


FIGURE 4 | The relationship between the expression of ARRDC2 and proportion of immune infiltrates in TIMER database. **(A,B)** Infiltration of various immune cells. **(A)** Macrophage, Neutrophil, Dendritic Cell. **(B)** B Cell, CD8+ T Cell, CD4+ T Cell. **(C)** Association between ARRDC2 gene copy number and immune cell infiltration levels in OV cohorts (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(D)** In TIMER database, the relationship between Immune checkpoint (CTLA4, PDCD1, CD274, PDCD1LG2) and ARRDC2 gene expression was analyzed by Spearman correlation analysis ($p < 0.05$). Correlation coefficients (rho values) and p values were shown.

altered DNA methylation in tumor formation continues to be revealed. Therefore, we utilized RNA-seq data and DNA methylation data to explore the relationship between DNA CpG site methylation levels and ARRDC2 mRNA expression based on the TCGA data. As shown in **Supplementary Figure S2A**, ovarian cancer tissues showed low levels of ARRDC2 gene

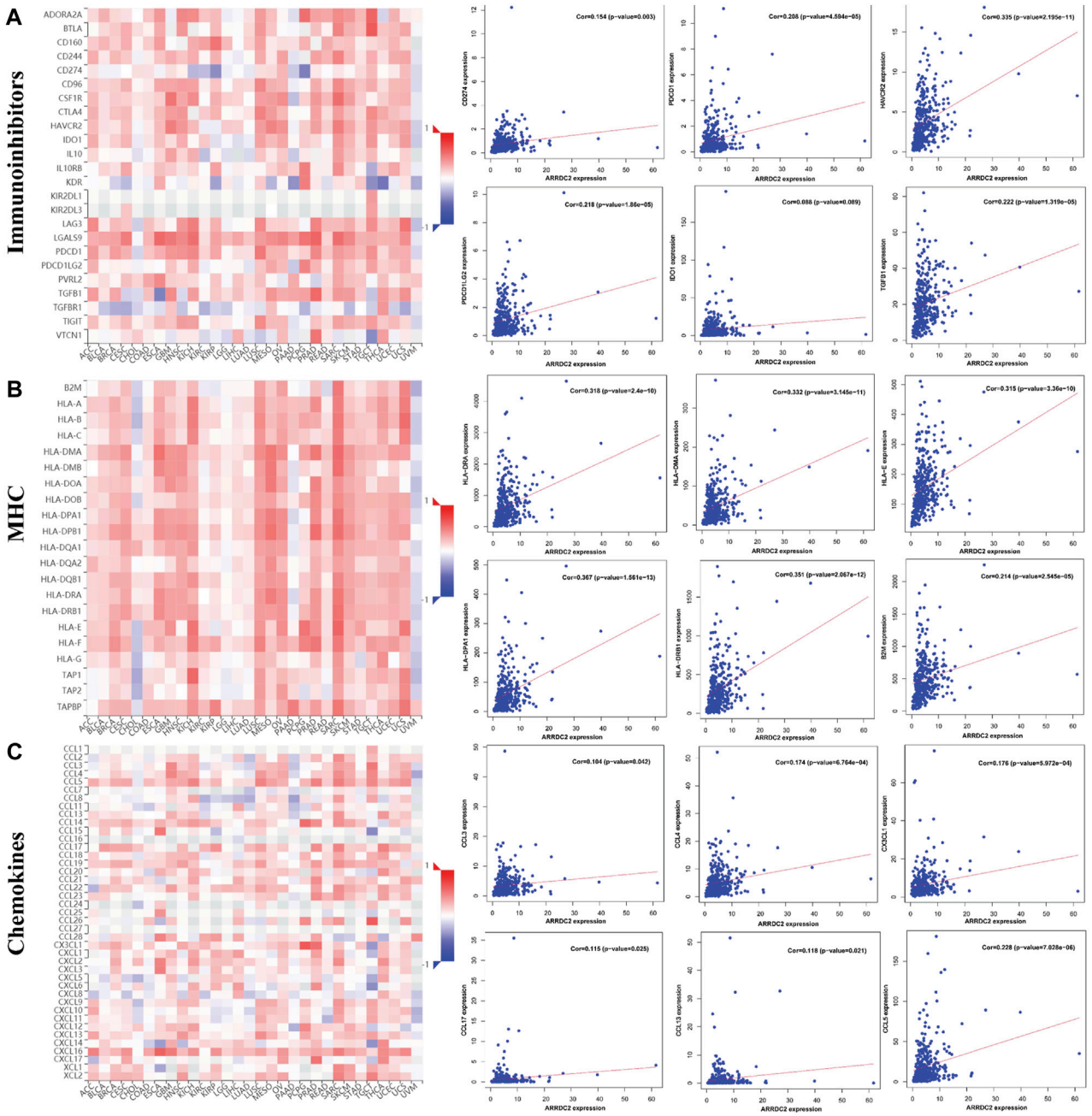


FIGURE 5 | The relationship between the expression of ARRDC2 and immune checkpoints, MHC and chemokines in the TISIDB database. **(A)** The relationship between immune checkpoint (CD274, HAVCR2, IDO1, PDCD1, PDCD1LG2, TGFB1) and ARRDC2 gene expression. **(B)** The relationship between MHC (B2M, HLA-DMA, HLA-DPA1, HLA-DRA, HLA-DRB1, HLA-E) and ARRDC2 gene expression. **(C)** The relationship between Chemokines (CCL17, CCL13, CCL5, CCL3, CCL4, CX3CL1) and ARRDC2 gene expression. Spearman correlation analysis was applied ($p < 0.05$).

methylation. Two sites were subsequently found to be aberrantly hypermethylated (cg23548920 and cg07374145, **Supplementary Figure S2B**), the two methylated CpG sites were negatively correlated with ARRDC2 expression (**Supplementary Figures S2C,D**). In conclusion, DNA methylation of ARRDC2 may be responsible for the difference in ARRDC2 expression levels in normal and tumor tissues.

Effects of ARRDC2 Knockdown on Malignant Biological Behavior of Ovarian Cancer Cells

On the basis of big data analysis, we used PCR and Western blot experiments to verify that the expression level of ARRDC2 in normal ovarian cells (IOSE80) was significantly lower than that in

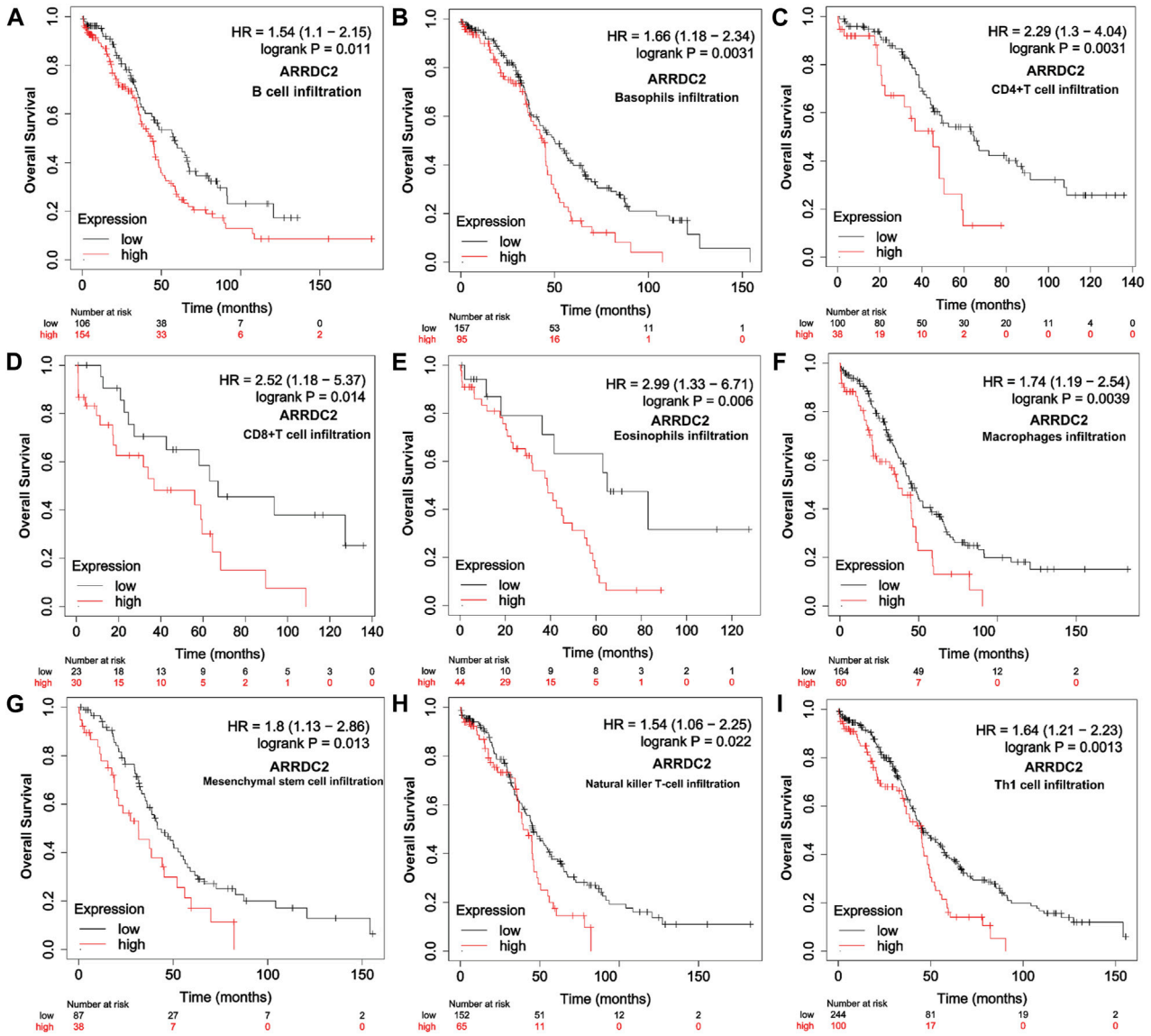
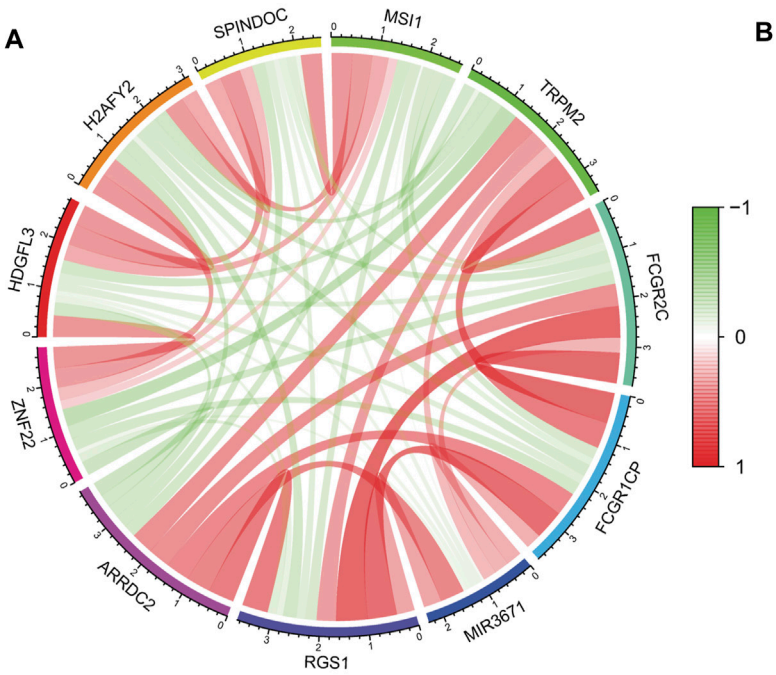


FIGURE 6 | Relationship between overall survival and immune cell infiltration and ARRDC2 gene expression in OV patients. **(A)** B Cell infiltration. **(B)** Basophils infiltration. **(C)** CD4+ T Cell infiltration. **(D)** CD8+ T Cell infiltration. **(E)** Eosinophils infiltration Neutrophil. **(F)** Macrophage infiltration. **(G)** Mesenchymal stem cell infiltration. **(H)** Natural killer T-cell infiltration. **(I)** Th1cell infiltration.

ovarian cancer cells (A2780 and SKOV3) (**Figures 8A,B**). After verifying the knockdown efficiency of ARRDC2 (**Figures 8C,D**), a series of *in vitro* experiments were performed with si-ARRDC2-516 to investigate the effect of ARRDC2 knockdown on the malignant biological behavior of OC cells. First, the CCK8 assay results showed that the optical density of the si-NC group was higher than that of the si-ARRDC2 group at 12, 24, and 48 h in A2780 and SKOV3 cell lines (**Figure 9A**). Thereafter, the results of colony formation assay and Ki-67 immunofluorescence assay showed that in OC cells, the number of colonies formed (**Figure 9E**) and the percentage of Ki-67 fluorescence positive cells were higher in the si-NC group compared with the si-ARRDC2 group. high (**Figure 9B**). Therefore, these results implied that proliferation was inhibited upon ARRDC2 knockdown in OC cells. The wound healing assay results showed that among OC cells, the 24-h wound healing rate of cells in the si-NC group was significantly higher than that in the si-ARRDC2 group (**Figure 9C**). In addition, transwell assay results showed that in OC cells, the number of migrating cells in the si-NC group was higher than that in the si-ARRDC2 group (**Figure 9D**). All the above results also confirmed that migration was inhibited after ARRDC2 knockdown in OC cells. Overall, the results of the *in vitro* experiments strongly suggested that biological behaviors such as proliferation and migration were inhibited after ARRDC2 knockdown in OC cells.



B

Gene	Correlation coefficient	P-value
ARRDC2	1	0.00E+00
RGS1	0.548	4.33E-31
MIR3671	0.514	5.77E-27
FCGR1CP	0.499	2.78E-25
FCGR2C	0.473	1.58E-22
TRPM2	0.472	1.79E-22
ZNF22	-0.229	2.99E-09
HDGFL3	-0.275	5.31E-08
H2AFY2	-0.269	1.05E-07
MSI1	-0.267	1.28E-07
SPINDOC	-0.267	1.32E-07

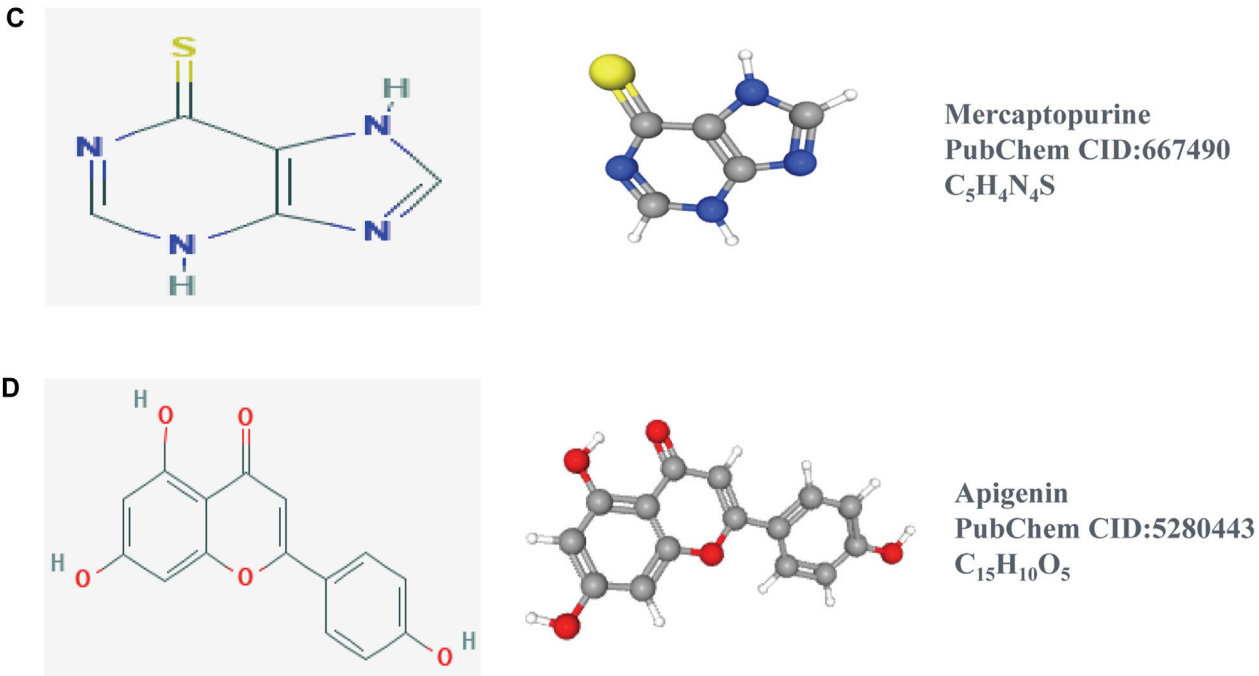


FIGURE 7 | Co-expression analysis of ARRDC2. **(A)** The ten most significant genes of positive and negative correlating with ARRDC2. **(B)** The Correlation coefficients and *p* values of the ten most significant genes of positive and negative correlating with ARRDC2. **(C,D)** Screening of gene therapy drugs for ARRDC2 based on the CMap and Pubchem database (Drug name, chemical structure, 2D structure and 3D structure). **(C)** Mercaptopurine. **(D)** Apigenin.

DISCUSSION

The survival outcome of ovarian cancer patients is not promising due to the lack of early screening markers and multiple alternative treatment options. The key role of ARRDC2, an important member of the arrestin protein family related to immunity, in OV needs to be studied urgently. In this study, we tried to use a variety of databases to explore and verify the expression level of

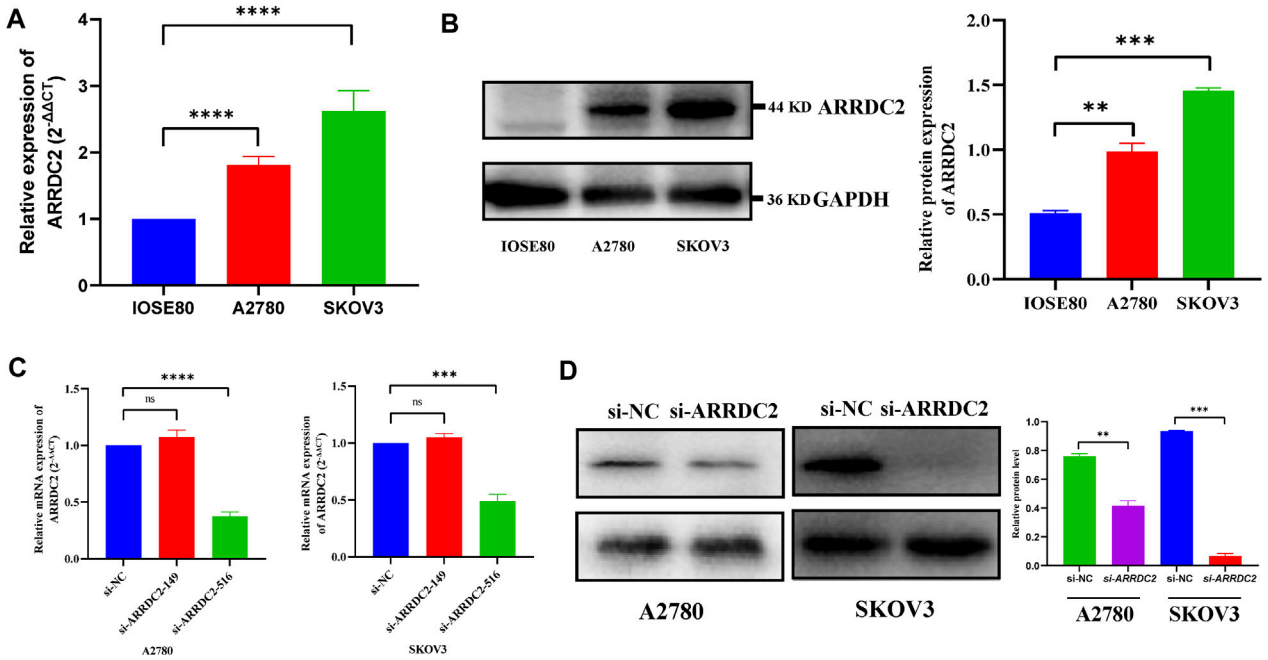


FIGURE 8 | Differential expression and the knockdown efficiency of ARRDC2 at the cellular level. **(A)** RT-qPCR experimental results showed that the expression of ARRDC2 in ovarian cancer cells is higher than that in normal ovarian cells. **(B)** Western blot experimental results showed that the expression of ARRDC2 in ovarian cancer cells is higher than that in normal ovarian cells. The differences in ARRDC2 gene expression levels in OC were compared at the cell levels by Wilcoxon rank sum test. **(C)** RT-qPCR after cell transfection. **(D)** Western blot image and corresponding statistics after cell transfection (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

ARRDC2, the relationship between ARRDC2 and clinical features and prognosis, potential molecular mechanisms and impact on tumor immune microenvironment. And on the basis of big data analysis, it was verified by *in vitro* cell experiments. The schematic diagram of the flow of this study was shown in **Supplementary Figure S1**.

The expression level of ARRDC2 in OV was firstly explored. As shown in **Figure 1A**, ARRDC2 was found to be significantly overexpressed in a variety of malignancies using the TIMER database (**Figure 1A**). In view of the above pan-cancer results, we tried to verify the expression level of ARRDC2 in OV by other means. Firstly, verify it in the GSE data set related to ovarian cancer. Two GEO datasets (GSE29450 and GSE10970) revealed that ARRDC2 showed significantly higher expression in OV compared to normal controls (**Figures 1B,C**). In parallel, we performed experimental PCR and Western blot to validate the expression level of ARRDC2 (**Figures 8A,B**). Secondly, our study also revealed that the abnormally high expression of ARRDC2 may be associated with abnormal hypomethylation of DNA, as shown in **Supplementary Figures S2A–D**. Numerous studies have confirmed that DNA methylation in epigenetics plays an important role in the malignancy, metastasis and recurrence of ovarian cancer (Papakonstantinou et al., 2020; Reid and Fridley, 2020). Epigenetic regulation, represented by DNA methylation, is essential for the regulation of oncogenes (Oliveira et al., 2021). The above results indicated that ARRDC2 was highly expressed in OV, suggesting that ARRDC2 is a potential oncogene and is affected by methylation.

After discovering that ARRDC2 was a potential oncogene, we tried to explore its impact on the prognosis of OV patients through retrospective studies based on gene expression and clinical information based on TCGA data and GSE19829. The results of the correlation analysis of clinical features showed that the expression level of ARRDC2 in OV increased with increasing FIGO stage and was highly expressed in Asian race groups (**Figures 1D,E**). And numerous studies have confirmed that the higher the FIGO stage, the worse the prognosis (Gąsowska-Bajger et al., 2021). In view of the above phenomena, it was of interest to us whether ARRDC2 might contribute to the poor prognosis of OV patients. Immediately after, the Kaplan-Meier survival analysis and survival meta-analysis of this study showed that the overall survival of patients in the high-expression group of ARRDC2 was shorter than that of the low-expression group (**Figures 2A–C**). In addition, survival meta-analysis further improved the scientific validity and rigor of Kaplan-Meier survival analysis. However, to exclude the effect of chance factors, univariate cox and multivariate cox analyses were used to confirm that ARRDC2 could serve as an independent risk factor for poor prognosis in patients with OV (**Figures 2D,E**). A great many studies have shown that genes in the ARRDC family contributed to the progression of various malignancies such as gastric, cervical and colorectal cancers, and were strongly associated with poor prognosis (Shen et al., 2018; Xiao et al., 2018; Takeuchi et al., 2019). For instance, ARRDC3, a member of the ARRDC family, served as a biomarker for the diagnosis and prognosis of epithelial ovarian cancer (Chen et al., 2021). From this, it can be boldly speculated that ARRDC2 may likewise act as an oncogene in ovarian cancer

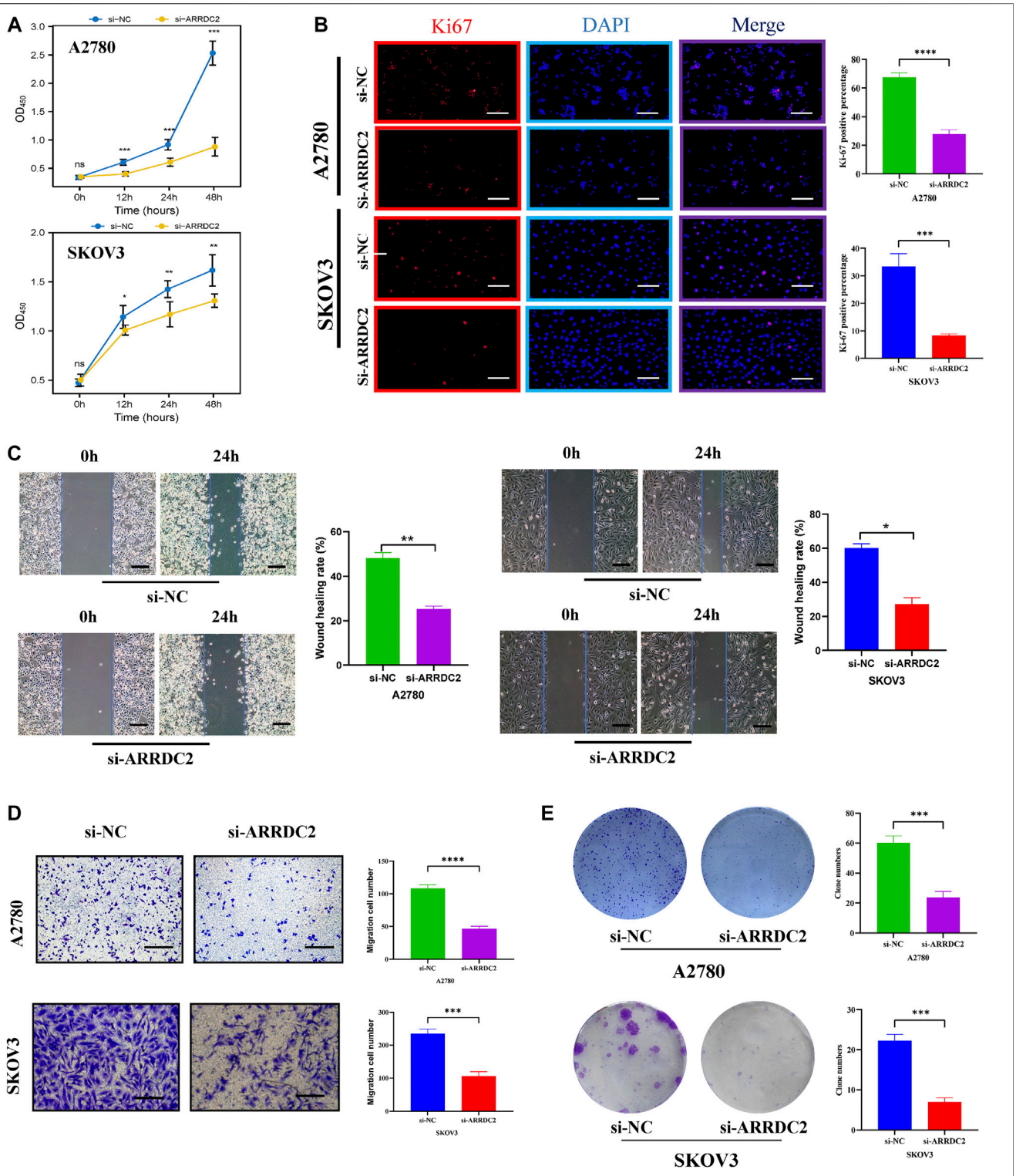


FIGURE 9 | Effects of ARRDC2 gene knockdown on malignant biological behavior of OC cells. **(A)** CCK8 experiment results of A2780 and SKOV3 cell lines after cell transfection. The OD values measured at 450 nm wavelength at 0, 12, 24, and 48 h were displayed, which represented the cell proliferation rate. **(B)** Ki-67 immunofluorescence staining of A2780 and SKOV3 cell lines after cell transfection. **(C)** Wound-healing assay results and statistics of OC cells after cell transfection. The wound-healing rate was measured at 0 and 24 h. **(D)** Transwell assay results and statistics of OC cells after cell transfection. The number of migrating cells was measured at 24 h, which represented the ability to migrate. **(E)** The results and statistics of cell clone formation assay of A2780 and SKOV3 cell lines after cell transfection (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

and lead to poor prognosis, but its possible oncogenic mechanisms need to be further explored.

To further understand the pathological mechanism of poor prognosis of OV due to ARRDC2, we performed GO annotation analysis and enrichment analysis of KEGG cell signaling pathway. As shown in **Figures 4A,B**, ARRDC2 may be involved in immune response regulatory signaling pathway, B cell receptor signaling pathway, T cell signaling pathway, Th17 cell differentiation and other immune-related biological pathways and processes. Meanwhile, using GSEA, we again confirmed that ARRDC2 was highly and consistently enriched in immune-related signaling pathways. Studies have shown that ARRDC3, an important member of the ARRDC family, was closely related to immunity in epithelial ovarian cancer, which supports our research (Chen et al., 2021). The above enrichment analysis suggested that ARRDC2 may play an oncogenic role in ovarian cancer by influencing immune factors in the tumor microenvironment.

Therefore, we further explored the relevance of ARRDC2 to immune cells and immune checkpoints in the tumor immune microenvironment of ovarian cancer using the TIMER and TISIDB databases (Ru et al., 2019). **Figures 4A,B** showed that the mRNA expression levels of ARRDC2 positively correlated with the infiltration of immune cells, including CD8+ T cells, B cells, neutrophils and dendritic cells. Numerous studies have confirmed that immune infiltration is considered as one of the hallmarks of cancer. Considering the importance of immune cell infiltration in tumors, TISIDB further evaluated the abundance ratio of different immune cells in ovarian cancer (**Figure 5A**) (Hanahan and Weinberg, 2011). In addition to this, the TISIDB database also showed a significant positive association of the ARRDC2 gene with immunomodulators, major histocompatibility complex molecules (MHC) and chemokines (**Figures 5B,C**). Previous studies have shown that if upregulation of the MHC-I complex is present, then NK cells can target these cells to send inhibitory signals, leading to long-term survival of tumor cells (Buller et al., 2020). It also corroborated our study that ARRDC2 was positively correlated with the MHC gene and could act as an oncogene in ovarian cancer. Interestingly, we also found a significant positive correlation between ARRDC2 expression and four immune checkpoints (**Figure 4D**). With advances in immunotherapy, particularly antibodies against the immune checkpoints cytotoxic, such as T lymphocyte-associated protein 4 (CTLA-4), programmed death protein 1 (PD-1) and programmed death ligand 1 (PD-L1), have shown clinical efficacy in ovarian cancer (Peyraud and Italiano, 2020; Wan et al., 2021). The degree of immune cell infiltration can be determined by gene expression profiles of immune-related genes, which may help estimate the prognosis of patients (Sui et al., 2020). In this regard, this research showed that the overall survival of patients with high ARRDC2 expression was shortened as seen by the Kaplan-Meier Plotter database under different immune cell infiltration scenarios (**Figures 6A–I**). In conclusion, this study fully confirmed the close correlation between ARRDC2 and important tumor immune microenvironment components such as immune cell infiltration and immune checkpoints. Moreover, the high expression of ARRDC2 could lead to poor prognosis such as reduced overall survival under different immune cell infiltration environments, which suggested that ARRDC2 may be used as a novel immunotherapy target to improve the clinical prognosis of OV patients.

After exploring the aberrant expression of ARRDC2 and its prognostic significance, the effect of ARRDC2 on the biological behavior of OC cells was assessed by a complete set of cytological experiments. The results in **Figures 8C,D** confirmed that the knockdown efficiency of ARRDC2 was sufficient and provided a basis for further studies. Furthermore, the experiments confirmed the inhibitory effect of ARRDC2 knockdown on the proliferation and migration of OC cells (**Figure 9**). Through the combination of big data analysis and cell experiment validation, ARRDC2 was preliminarily confirmed as a potential oncogene. May lead to poor prognosis by promoting the malignant biological behavior of OC cells.

The ultimate goal of the study was to benefit the clinic. Therefore, based on the above research part, the CMap small molecule drug analysis was performed using co-expressed genes positively and negatively associated with ARRDC2, and finally two small molecule compounds with potential therapeutic effects on ovarian cancer were predicted: Mercaptopurine and Apigenin (**Figure 7**). A large number of previous studies have evaluated the reliability of the CMap small molecule drug analysis for drug prediction (Iskar et al., 2010; Cheng et al., 2014). And encouragingly, previous studies have demonstrated the use of Mercaptopurine as an immunomodulatory agent in the treatment of patients with inflammatory bowel disease (Su et al., 2000). In addition, Mercaptopurine is effective in the treatment of immune modulation disorders and acute lymphoblastic leukemia (Marinaki and Arenas-Hernandez, 2020). For Apigenin, it has been reported that Apigenin inhibited various human cancers *in vitro* and *in vivo* as well as stimulates immune responses through multiple biological effects (Yan et al., 2017). Besides, Apigenin limited melanoma growth by inhibiting PD-L1 expression through modulation of tumor and antigen (Xu et al., 2018). Notably, previous studies are highly consistent with our study that ARRDC2 may serve as a new target for immunotherapy and may provide a new direction for subsequent immunopharmacological treatment of ovarian cancer.

The research focus of ovarian cancer treatment is precision immunotherapy or targeted therapy driven by specific biomarkers, which will bring better survival results for patients with ovarian cancer. In addition, it is important to explore new biomarkers not only for prognostic assessment, but also for the exploration of the subtle mechanisms of epigenetic and immunological changes that occur in the tumor immune microenvironment. Based on these circumstances, this study has certain strengths and limitations. The main strength of this study is the originality of the findings. This study brought the role of ARRDC2 as an immune-related prognostic biomarker in ovarian cancer to the public eye for the first time. The discovery and in-depth study of ARRDC2 would enable it to predict the effect of treatment, be used as a new potential therapeutic target and extend the overall survival of patients. However, this study still has some limitations. Although this study clarified the abnormal expression of ARRDC2 and its prognostic significance based on public data and cell experiment data. However, this study did not verify the prognostic significance of ARRDC2 in a clinical cohort. This is the limitation of this study and the direction of further research in the future.

CONCLUSION

Our study suggests that aberrantly expressed ARRDC2 may be a potential prognostic marker for OV. More importantly, it may promote the proliferation and migration of ovarian cancer cells and may be associated with the tumor immune microenvironment. Clinically significant ARRDC2 may be used to assess the clinical prognosis of patients with OV and may also be used as a target for immunotherapy or as a potential marker for checkpoint inhibitor-based immunotherapy.

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**.

ETHICS STATEMENT

The study protocol was approved by The Ethics Committee of the Harbin Medical University Cancer Hospital (Harbin, China). The use of patient samples conformed to the declaration of Helsinki. All patients provided informed written consent.

AUTHOR CONTRIBUTIONS

CW designed the study. MZ and YdL reviewed the raw data and confirm the authenticity of all raw data. SH performed the analysis. HL, YM, and YaL collected data. MZ drafted the manuscript. XC and CW revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Grant Numbers: 81772274).

ACKNOWLEDGMENTS

Thanks for the support of Harbin Medical University Cancer Hospital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.815082/full#supplementary-material

REFERENCES

Buller, C. W., Mathew, P. A., and Mathew, S. O. (2020). Roles of NK Cell Receptors 2B4 (CD244), CS1 (CD319), and LLT1 (CLEC2D) in Cancer. *Cancers (Basel)* 12, 12. doi:10.3390/cancers12071755

Chen, Y., Tian, D., Chen, X., Tang, Z., Li, K., Huang, Z., et al. (2021). ARRDC3 as a Diagnostic and Prognostic Biomarker for Epithelial Ovarian Cancer Based on Data Mining. *Ijgm* Vol. 14, 967–981. doi:10.2147/ijgm.s302012

Cheng, J., Yang, L., Kumar, V., and Agarwal, P. (2014). Systematic Evaluation of Connectivity Map for Disease Indications. *Genome Med.* 6, 540. doi:10.1186/s13073-014-0095-1

Dores, M. R., Lin, H., Grimsey, F., and Trejo, J. (2015). The α-arrestin ARRDC3 Mediates ALIX Ubiquitination and G Protein-Coupled Receptor Lysosomal Sorting. *MBoC* 26, 4660–4673. doi:10.1091/mbc.e15-05-0284

Gąsowska-Bajger, B., Gąsowska-Bodnar, A., Knapp, P., and Bodnar, L. (2021). Prognostic Significance of Survivin Expression in Patients with Ovarian Carcinoma: A Meta-Analysis. *J. Clin. Med.* 10, 10. doi:10.3390/jcm10040879

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Huang, C.-N., Huang, S.-P., Pao, J.-B., Chang, T.-Y., Lan, Y.-H., Lu, T.-L., et al. (2012). Genetic Polymorphisms in Androgen Receptor-Binding Sites Predict Survival in Prostate Cancer Patients Receiving Androgen-Deprivation Therapy. *Ann. Oncol.* 23, 707–713. doi:10.1093/annonc/mdr264

Iskar, M., Campillos, M., Kuhn, M., Jensen, L. J., van Noort, V., and Bork, P. (2010). Drug-induced Regulation of Target Expression. *PLoS Comput. Biol.* 6. doi:10.1371/journal.pcbi.1000925

Marinaki, A. M., and Arenas-Hernandez, M. (2020). Reducing Risk in Thiopurine Therapy. *Xenobiotica* 50, 101–109. doi:10.1080/00498254.2019.1688424

Oliveira, D., Hentze, J., O'Rourke, C. J., Andersen, J. B., Høgdall, C., and Høgdall, E. V. (2021). *DNA Methylation in Ovarian Tumors-A Comparison between Fresh Tissue and FFPE Samples*. Thousand Oaks, Calif): Reproductive sciences.

Papakonstantinou, E., Androutsopoulos, G., Logotheti, S., Adonakis, G., Maroulis, I., and Tzelepi, V. (2020). DNA Methylation in Epithelial Ovarian Cancer: Current Data and Future Perspectives. *Curr. Mol. Pharmacol.* 14 (6), 1013–1027. doi:10.2174/1874467213666200810141858

Peyraud, F., and Italiano, A. (2020). Combined PARP Inhibition and Immune Checkpoint Therapy in Solid Tumors. *Cancers (Basel)* 12. doi:10.3390/cancers12061502

Rafiq, S., Tapper, W., Collins, A., Khan, S., Politopoulos, I., Gerty, S., et al. (2013). Identification of Inherited Genetic Variations Influencing Prognosis in Early-Onset Breast Cancer. *Cancer Res.* 73, 1883–1891. doi:10.1158/0008-5472.can-12-3377

Reid, B. M., and Fridley, B. L. (2020). DNA Methylation in Ovarian Cancer Susceptibility. *Cancers (Basel)* 13. doi:10.3390/cancers13010108

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). TISIDB: an Integrated Repository Portal for Tumor-Immune System Interactions. *Bioinforma. Oxf. Engl.* 35, 4200–4202. doi:10.1093/bioinformatics/btz210

Santoiemma, P. P., Reyes, C., Wang, L.-P., McLane, M. W., Feldman, M. D., Tanyi, J. L., et al. (2016). Systematic Evaluation of Multiple Immune Markers Reveals Prognostic Factors in Ovarian Cancer. *Gynecol. Oncol.* 143, 120–127. doi:10.1016/j.ygyno.2016.07.105

Shen, X., Sun, X., Sun, B., Li, T., Wu, G., Li, Y., et al. (2018). ARRDC3 Suppresses Colorectal Cancer Progression through Destabilizing the Oncoprotein YAP. *FEBS Lett.* 592, 599–609. doi:10.1002/1873-3468.12986

Su, C. G., Stein, R. B., Lewis, J. D., and Lichtenstein, G. R. (2000). Azathioprine or 6-mercaptopurine for Inflammatory Bowel Disease: Do Risks Outweigh Benefits? *Dig. Liver Dis.* 32, 518–531. doi:10.1016/s1590-8658(00)80010-9

Sui, S., An, X., Xu, C., Li, Z., Hua, Y., Huang, G., et al. (2020). An Immune Cell Infiltration-Based Immune Score Model Predicts Prognosis and Chemotherapy Effects in Breast Cancer. *Theranostics* 10, 11938–11949. doi:10.7150/thno.49451

Takeuchi, F., Kukimoto, I., Li, Z., Li, S., Li, N., Hu, Z., et al. (2019). Genome-wide Association Study of Cervical Cancer Suggests a Role forARRDC3gene in Human Papillomavirus Infection. *Hum. Mol. Genet.* 28, 341–348. doi:10.1093/hmg/ddy390

Tian, X., Irannejad, R., Bowman, S. L., Du, Y., Puthenveedu, M. A., von Zastrow, M., et al. (2016). The α-Arrestin ARRDC3 Regulates the Endosomal Residence Time and Intracellular Signaling of the β2-Adrenergic Receptor. *J. Biol. Chem.* 291, 14510–14525. doi:10.1074/jbc.m116.716589

Wan, C., Keany, M. P., Dong, H., Al-Alem, L. F., Pandya, U. M., Lazo, S., et al. (2021). Enhanced Efficacy of Simultaneous PD-1 and PD-L1 Immune Checkpoint Blockade in High-Grade Serous Ovarian Cancer. *Cancer Res.* 81, 158–173. doi:10.1158/0008-5472.CAN-20-1674

Whiteman, D. C., and Wilson, L. F. (2016). The Fractions of Cancer Attributable to Modifiable Factors: A Global Review. *Cancer Epidemiol.* 44, 203–221. doi:10.1016/j.canep.2016.06.013

Xiao, J., Shi, Q., Li, W., Mu, X., Peng, J., Li, M., et al. (2018). ARRDC1 and ARRDC3 Act as Tumor Suppressors in Renal Cell Carcinoma by Facilitating YAP1 Degradation. *Am. J. Cancer Res.* 8, 132–143.

Xu, L., Zhang, Y., Tian, K., Chen, X., Zhang, R., Mu, X., et al. (2018). Apigenin Suppresses PD-L1 Expression in Melanoma and Host Dendritic Cells to Elicit Synergistic Therapeutic Effects. *J. Exp. Clin. Cancer Res.* 37, 261. doi:10.1186/s13046-018-0929-6

Yan, X., Qi, M., Li, P., Zhan, Y., and Shao, H. (2017). Apigenin in Cancer Therapy: Anti-cancer Effects and Mechanisms of Action. *Cell Biosci.* 7, 50. doi:10.1186/s13578-017-0179-x

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.

frontiers | Frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fmolb.2022.895746

Efficient Detection of the Alternative Spliced Human Proteome Using Translatome Sequencing

Chun Wu[1†], Xiaolong Lu[1†], Shaohua Lu[1,2†], Hongwei Wang[1†], Dehua Li[1], Jing Zhao[1], Jingjie Jin[1], Zhenghua Sun[1], Qing-Yu He[1], Yang Chen[1] and Gong Zhang[1]**

[1]Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes and MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, Jinan University, Guangzhou, China, [2]State Key Laboratory of Respiratory Disease, School of Basic Medical Sciences, Sino-French Hoffmann Institute, Guangzhou Medical University, Guangzhou, China

Alternative splicing (AS) isoforms create numerous proteoforms, expanding the complexity of the genome. Highly similar sequences, incomplete reference databases and the insufficient sequence coverage of mass spectrometry limit the identification of AS proteoforms. Here, we demonstrated full-length translating mRNAs (ribosome nascent-chain complex-bound mRNAs, RNC-mRNAs) sequencing (RNC-seq) strategy to sequence the entire translating mRNA using next-generation sequencing, including short-read and long-read technologies, to construct a protein database containing all translating AS isoforms. Taking the advantage of read length, short-read RNC-seq identified up to 15,289 genes and 15,906 AS isoforms in a single human cell line, much more than the Ribo-seq. The single-molecule long-read RNC-seq supplemented 4,429 annotated AS isoforms that were not identified by short-read datasets, and 4,525 novel AS isoforms that were not included in the public databases. Using such RNC-seq-guided database, we identified 6,766 annotated protein isoforms and 50 novel protein isoforms in mass spectrometry datasets. These results demonstrated the potential of full-length RNC-seq in investigating the proteome of AS isoforms.

Keywords: alternative splicing, translatome sequencing, mass spectrometry, proteome, isoform, human proteome project

OPEN ACCESS

Edited by:
Andy T. Y. Lau,
Shantou University, China

Reviewed by:
Yingwei Hu,
Johns Hopkins University, United States
Søren Lykke-Andersen,
Aarhus University, Denmark

****Correspondence:***
Gong Zhang
zhanggong-uni@qq.com
Yang Chen
chanyang.cy@qq.com

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

Specialty section:
This article was submitted to Protein Biochemistry for Basic and Applied Sciences, a section of the journal Frontiers in Molecular Biosciences

Received: *14 March 2022*
Accepted: *28 April 2022*
Published: *02 June 2022*

Citation:
Wu C, Lu X, Lu S, Wang H, Li D, Zhao J, Jin J, Sun Z, He Q-Y, Chen Y and Zhang G (2022) Efficient Detection of the Alternative Spliced Human Proteome Using Translatome Sequencing. Front. Mol. Biosci. 9:895746. doi: 10.3389/fmolb.2022.895746

INTRODUCTION

A single human gene can produce a variety of alternative splicing (AS) isoforms, which may be translated into protein isoforms with different localizations, structures and functions, which dramatically diversify the transcriptome and proteome. Proteoforms from AS can have different functional domains, such as enzymatic active sites or protein-binding sites, and participate in various important physiological and pathological processes. A lot of evidence shows that AS disorders lead to various diseases (Wang et al., 2008; Baralle and Giudice, 2017). Therefore, a major add-on challenge of Human Proteome Project (HPP) over Human Genome Project is to discover protein AS isoforms (Paik et al., 2012).

With the development of next-generation sequencing technology, more than 100,000 AS isoforms have been found in human genome (Pan et al., 2008). However, it has been shown that different AS transcripts are translated in different efficiency, some of which are not translated (Wang et al., 2013). Identifying protein products of the AS isoforms and characterizing their functions remain a huge

challenge because of two main reasons: 1) Protein isoforms from AS are usually highly similar in sequences and thus may be difficult to distinguish using mass spectrometry (MS) techniques. High sequence homologies and similarity physical-chemical properties of protein isoforms make it difficult for them to be effectively separated by pre-fractionation steps, and unique peptides generated by the digestion of low-expression proteins is hard to be identified by MS. The average sequence coverage of a recent near-complete yeast proteome is only 29% when using trypsin digestion (Gao et al., 2021). The sequence coverage of human proteome is 14–25%, which hinders the discovery of unique peptides of AS isoforms (Wang et al., 2019). 2) Many public proteome reference databases tend to include only the canonical isoforms. The AS isoforms are largely missing or incomplete (Sulakhe et al., 2019). Moreover, protein isoforms sequences are largely inconsistent across all the commonly used databases. In comparison with the canonical isoforms, the sequence features in the human alternatively splicing isoforms always be lost or modified (Frankish et al., 2015).

We previously exhibited that the translatome sequencing, i.e. next-generation sequencing on the translating mRNAs, provides a powerful tool to investigate the proteins which are being synthesized (Zhong et al., 2014). Due to the high throughput of translatome sequencing techniques, it is relatively easy to achieve near-complete sequence coverage of translating mRNAs, thus provides a solid basis of the analysis on protein isoforms and single amino-acid polymorphisms. There are two major translatome sequencing techniques, RNC-seq and Ribo-seq (reviewed in (Zhao et al., 2019)). In brief, Ribo-seq use ribonuclease to digest mRNA excluded by ribosomes into ribosome protected fragments (RPFs), also known as ribosome footprints (RFPs), ~28 nt in average in eukaryotic cells. It is proposed to identify the non-canonical translation initiation or termination, the truncation or extension of reading frame and uORF, etc. However, Ribo-seq seems to show high false positives in ORF detection in practice (Guttman et al., 2013; Lu et al., 2019). In contrast, RNC-seq sequences the entire mRNA in the ribosome nascent-chain complex. Since the mRNA is intact, the sequencing library can be of any size. Therefore, RNC-seq has some following advantages in detecting translating AS isoforms. 1) The long reads can easily exclude the small ribosome-engaged RNA fragment contaminants by the longer library length and poly(A)-enrichment strategy, minimizing the false positives. 2) The long reads have a greater chance of span across the splice junction and thus enable alignments to process reads across junctions more accurately, which means RNC-seq can theoretically detect more AS isoforms. 3) The unlimited length of insert fragment can reveal novel isoforms and guide the discovery of novel proteins isoforms by long-read sequencing. 4) The experimental process of RNC-seq without enzymic digestion is simpler. Therefore, the result of RNC-seq has the advantages of better stability and reproducibility. RNC-seq has found a "hidden proteome", i.e. a large number of proteins encoded by "non-coding" RNAs (ncRNAs), demonstrating that RNC-seq is an effective method to guide new protein identification (Lu et al., 2019).

In this work, we systematically compared RNC-seq and Ribo-seq in the context of proteome identification, especially when identifying protein isoforms from AS. We also demonstrated that the single-molecule long read sequencing technique identified thousands of new splice variants and guided the MS identifications of new protein isoforms.

RESULTS

Translating AS Detection Efficiency of RNC-Seq and Ribo-Seq

Both RNC-seq and Ribo-seq need to extract ribosome fraction. In human cells, rRNAs account for 80–85% of the total RNA, while coding mRNAs account for only 1–5% (Zhao et al., 2018). The RNC-seq selects mRNA using poly-dT oligos, which effectively avoid sequencing rRNA. In contrast, the ribosomal footprints in Ribo-seq lacks polyA tail. The rRNA can only be removed by hybridization, which is trickier and often inadequate. For example, in MHCC97H and HeLa cells, our RNC-seq datasets contained only ~1% rRNA reads, while Ribo-seq datasets contained 16–83% rRNA reads (**Figure 1A**). We analyzed 775 RNC-seq and Ribo-seq datasets in the TranslatomeDB (Liu et al., 2018). The Ribo-seq datasets contained much higher fraction of rRNA reads than RNC-seq (**Figure 1B**). This remarkably decreased the effective mRNA reads, which undermined the AS detection efficiency of Ribo-seq.

When detecting known splice variants, it is efficient to map short sequencing reads to RNA reference sequences. The uniquely mapped reads represent the specific splice variants. When normalized against the non-rRNA read number, the RNC-seq datasets (100 nt read length) yielded approximately doubled uniquely mapped reads than the Ribo-seq datasets (~28 nt reads) (**Figure 1C**). This suggests that RNC-seq can improve the efficiency of AS isoforms identification not only by the better enrichment strategies, but also by the longer read lengths, since longer reads are more likely to span across the specific splice junctions and specific exons.

In addition, longer reads also provides more information for the algorithms to identify the reads across junctions more accurately, thereby facilitates the discovery of splice junctions including novel splice events (the splice events which are not annotated in the database). As a validation, we used STAR algorithm to map the rRNA-filtered clean reads to human genome GRCh38. The RNC-seq identified 3.2–8.7 × more canonical junctions and 3.5 × more novel junctions than Ribo-seq (**Figure 1D**).

We then evaluated the number of expressed genes and isoforms identified by RNC-seq and Ribo-seq. With the increasing number of raw reads, both RNC-seq and Ribo-seq can identify more expressed genes and isoforms (**Figure 1E**). However, using the same throughput of raw reads, RNC-seq identified 17–27% more translating genes and 1.6–2.3 × more known splice isoforms than Ribo-seq. For example, RNC-seq identified 15,906 isoforms by isoform-specific and uniquely mapped reads in MHCC97H cell line, while Ribo-seq

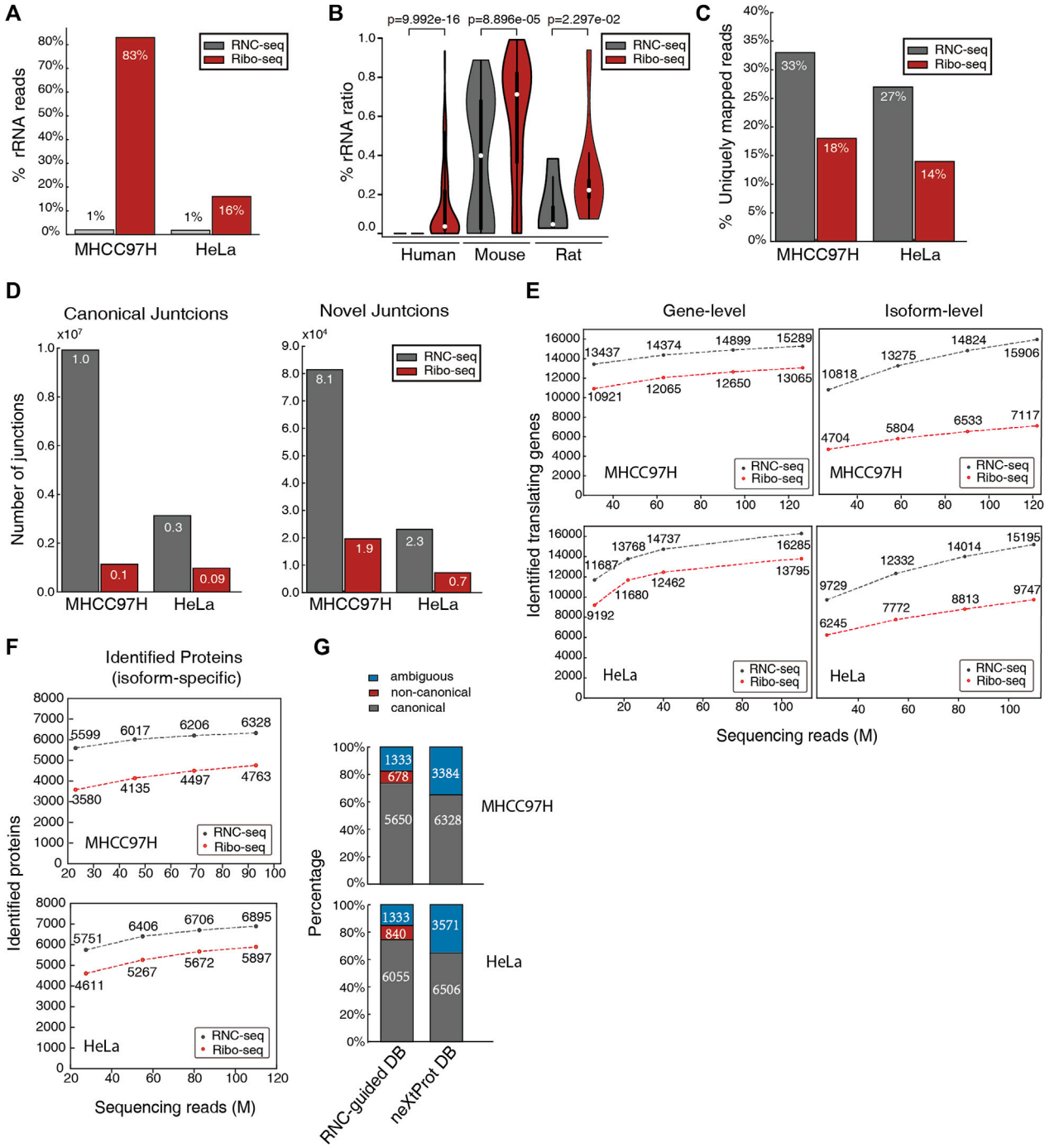


FIGURE 1 | RNC-seq's advantages over Ribo-seq in detecting splicing isoforms with higher efficiency. **(A)** Percentage of rRNA reads in RNC-seq and Ribo-seq datasets of MHCC97H and HeLa cells. **(B)** Percentage of ribosomal RNA reads in RNC-seq and Ribo-seq datasets in human, mouse, and rat. The *p*-values were obtained from Kolmogorov-Smirnov test. **(C)** Percentage of uniquely mapping reads (uni-mapped reads/non-rRNA reads) in RNC-seq and Ribo-seq. **(D)** Number of identified canonical junctions (left) and novel junctions (right) using same number of non-rRNA reads (25M). **(E)** Number of identified genes and isoforms supported by uniquely mapped reads under the same throughput of raw reads. **(F)** Number of protein isoforms identified by using RNC-seq and Ribo-seq guided protein database. **(G)** Proportion of identified distinct proteins by RNC-seq guided database and neXtProt. Ambiguous: these proteins share all their peptides with other proteins, and thus cannot be unambiguously identified.

identified only 7,117 isoforms. This trend is also valid when considering the same number of non-rRNA reads (**Supplementary Figure S1**).

Using the identified translating isoforms (with isoform-specific reads) to build protein databases to identify proteins in mass spectrometry datasets, the RNC-seq

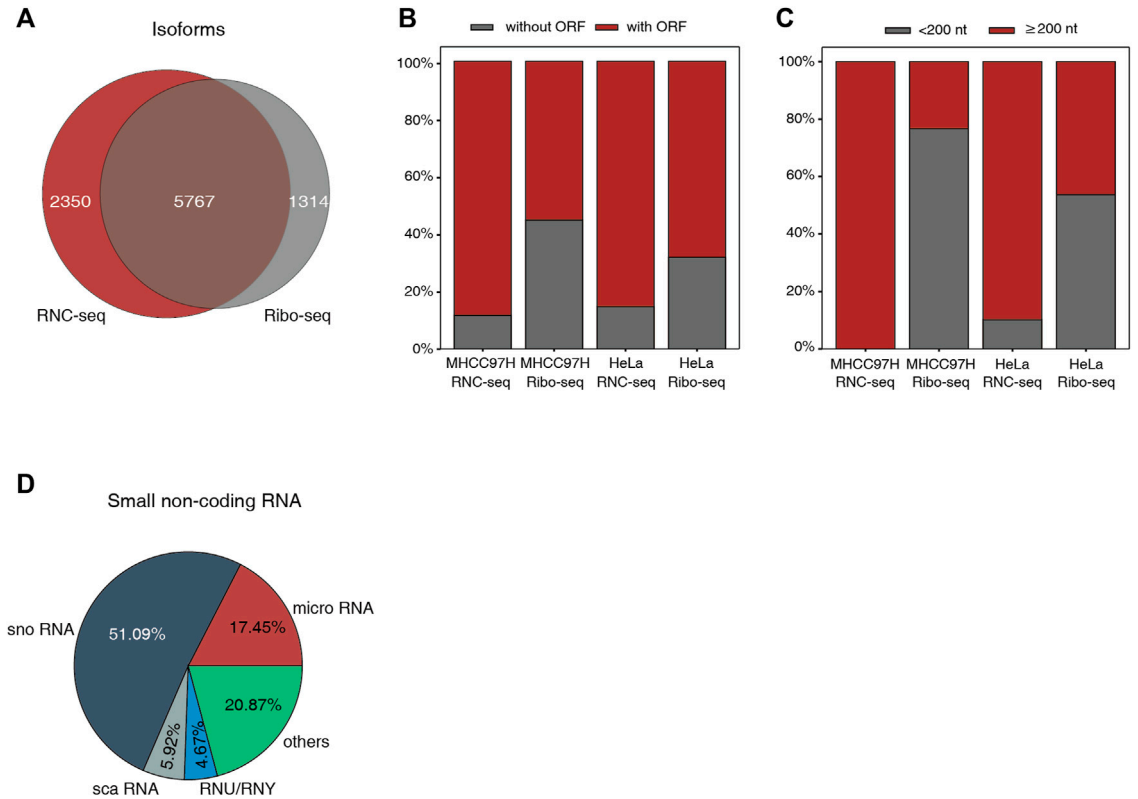


FIGURE 2 | The difference of identified isoforms between RNC-seq and Ribo-seq. **(A)** Venn diagram of isoforms identified by RNC-seq and Ribo-seq using 8 million uniquely mapped reads, respectively. **(B)** Distribution of identified ncRNA with and without ORFs. **(C)** Length distribution of the ncRNA without canonical ORF. **(D)** Types of small ncRNA identified by Ribo-seq of MHCC97H.

database identifies 17–56% more protein than Ribo-seq, following the criteria of HPP Guideline 3.0 (**Figure 1F**). Compared to the standard neXtProt (with isoforms) database, more than 75% of the identified isoforms were canonical ones, and detected 678–840 non-canonical proteins which are not included in neXtProt database (**Figure 1G**). Besides these uniquely identified proteins, only 1,333 proteins were ambiguously identified because all their identified peptides were shared with other proteins. In contrast, 3,384–3,571 proteins were ambiguously identified using neXtProt database (**Figure 1G**). This indicated that the RNC-seq-guided database provided more concise identifications.

We then demonstrated the total number of identified proteins by using the identified translating genes (with gene-specific reads) and translating isoforms (with isoform-specific reads) to guide protein identification. For example, by using RNC-guided database (constructed by 132M reads of MHCC97H and 110M reads of HeLa, respectively) we identified 10,887 and 11,308 proteins (**Supplementary Figure S2**), while only 6,328 and 6,506 proteins identified by using the neXtProt database. This result demonstrated that the RNC-seq-guided database strategy significantly optimized the protein identification efficiency.

In sum, RNC-seq has a distinct advantage in detecting translating AS isoforms and novel protein isoforms.

The Translation Potential of Identified Non-coding RNAs Between RNC-Seq and Ribo-Seq

It is known that many "ncRNAs" can be translated into proteins in canonical way. We evaluated the translating potential of these ncRNAs identified by RNC-seq and Ribo-seq under the same number of uniquely mapping reads (**Figure 2A**). In independently identification of RNC-seq, about 86% of the translating "ncRNAs" contain canonical AUG-started open reading frames (ORFs) of at least 50 aa in length, while in Ribo-seq, 55–68% of the translating "ncRNAs" contains canonical ORFs (**Figure 2B**). This raised a doubt whether such non-canonical "new proteins" identified by Ribo-seq were real. Indeed, 54–77% of the noncoding isoforms without canonical ORF identified by Ribo-seq were classical small noncoding RNAs (**Figure 2C**), mainly snoRNAs, which are unable to encode proteins ≥ 50 aa (**Figure 2D**). Previous studies have shown that ribosomes interact extensively with snoRNAs, such as ribosome biogenesis (Reichow et al., 2007), and the interaction of 80S ribosomes and pre-mRNA with

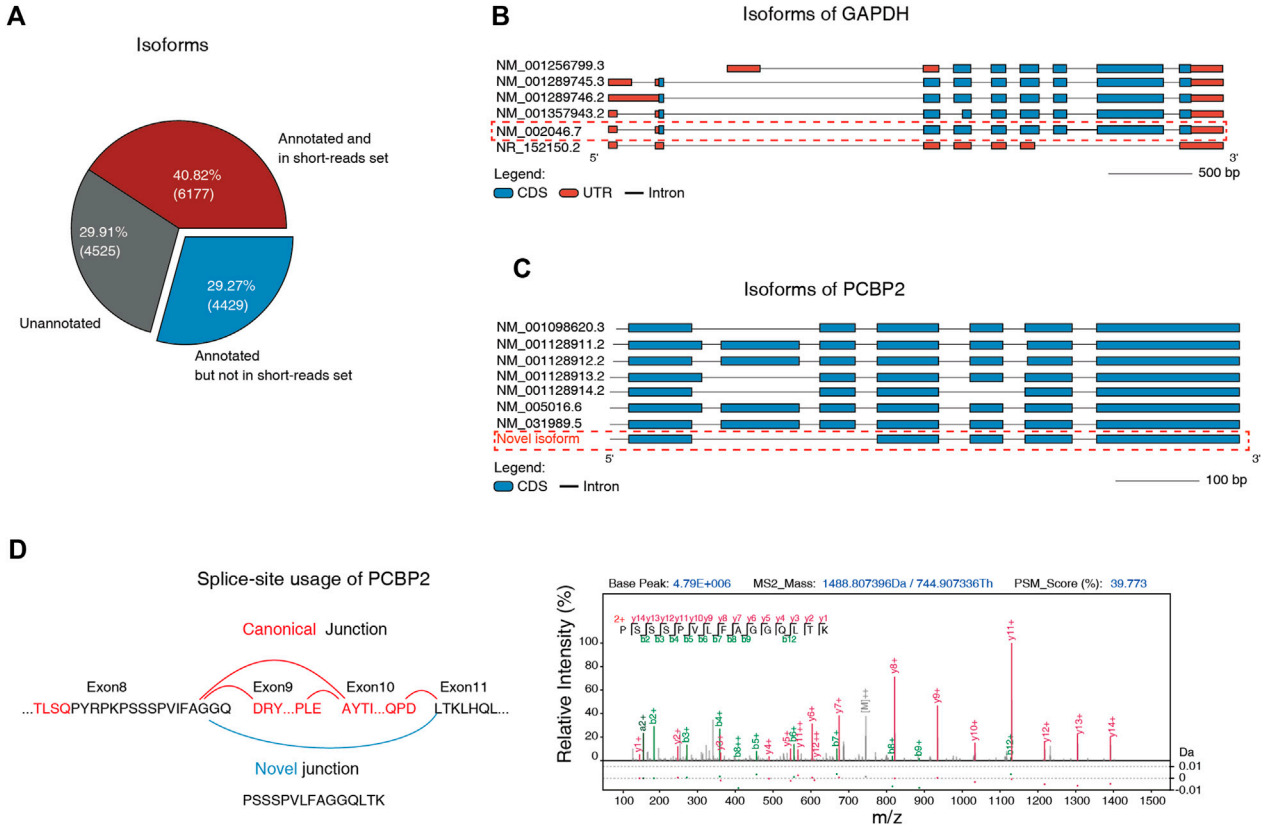


FIGURE 3 | Single-molecule full-length RNC-seq improved the efficiency of isoforms identification. **(A)** Three types of identified isoforms by Nanopore RNC-seq. "Annotated" means that the isoform is included in NCBI human Refseq database. "in short-read set" means that the isoform can be identified by short-read RNC-seq. **(B)** All isoforms of GAPDH. The marked isoform was identified by full-length RNC-seq but not by short-read RNC-seq. **(C)** All isoforms of PCBP2. Nanopore full-length RNC-seq identified a novel isoform that was absent in Refseq and neXtProt databases. **(D)** The MS-detected specific junction peptide (PSSSPVLFAGGQLTK) that supported the novel junction of PCBP2.

snoRNAs induces degradation to generate mature mRNA and functional snoRNAs. Therefore, snoRNAs can cross the sucrose cushion with ribosomes during ultracentrifuging, which would lead to the detection of snoRNAs in final sequencing data. Due to the digestion of RNase, it is difficult for Ribo-seq to exclude the non-RFPs of small ncRNAs and degraded fragments (Smith and Steitz, 1998; Guttman et al., 2013).

Direct Full-Length RNC Sequencing Reveals Isoform Complexity

It is difficult to determine exon arrangement solely by short reads. Single-molecule, long-read sequencing techniques (Iso-seq), such as PacBio or Nanopore, can solve this problem by sequencing the entire mRNA molecule (Rhoads and Au, 2015; Byrne et al., 2017). This provides a more detailed picture of the transcriptome and a powerful tool to detect novel AS isoforms. Using these techniques on RNC-seq, we can sequence the entire ribosome-bound mRNA to accurately determine the translating AS isoforms, as different AS isoforms are translated in different efficiency (Wang et al., 2013).

We performed the direct full-length RNC-seq on MHCC97H using Nanopore MinION sequencer. Due to the high error rate (~8.47%) of the single-molecule Nanopore sequencer, we corrected the sequences by more accurate short read RNC-seq data (error rate ~0.61%) corresponding to ensure the correction of novel isoforms identification. After filtering out novel singletons (Sessegolo et al., 2019), the full-length RNC-seq identified in total 15,131 unique AS isoforms. Among these isoforms, 4,525 (29.91%) were absent in NCBI Refseq mRNA reference database, and 4,429 (29.27%) were annotated in RefSeq database but cannot be identified by the short-read RNC-seq (**Figure 3A**). For example, we identified an isoform NM_002046 of gene GAPDH with the unique combination of exons by long reads of Nanopore sequencing, but would be missing in short-read sequencing because it has no unique splice junction compared to other splice isoforms (**Figure 3B**).

Next, we detected novel isoforms at protein level. We built a protein database by 3-frame-translation of all detected isoforms. This database included 18,200 protein sequences was used for mass spectrometry-based proteome identification. This database is much smaller than the neXtProt database with isoforms (41,653 entries). Therefore, the sensitivity was expected to be better. We detected 6,766 isoforms with at least 1 isoform-unique peptide ≥ 9 aa.

After filtering out the proteins whose unique peptides shared sequences with canonical proteins in neXtProt (with isoforms), we finally identified 50 novel protein isoforms which were not included in RefSeq and neXtProt reference with FDR < 0.01. When we controlled the FDR of protein-level and peptide-level to 0.001, there are still 43 novel isoforms could be identified with stricter quality control (**Supplementary Table S1, S2**, **Supplementary Figure S3**).

For example, we detected a novel isoform of PCBP2 gene with a unique peptide (**Figures 3C,D**). Compared with other annotated isoforms, this isoform alternative spans across the exons 8 and 11, and skips exons 9 and 10. These results indicates that the long-read RNC sequencing is efficient to reveal novel human protein isoforms.

DISCUSSION

Protein AS isoform is of great importance in proteome studies and has not been specifically and thoroughly investigated in the context of HPP because of the difficulties of detecting isoform-specific peptides in large-scale MS data (see the Introduction section). Previously, the cell line-specific protein sequence database based on transcriptome. However, it seems that the efficiency of such approaches was not satisfying. The Human Proteoform Atlas database collected 3,055 protein isoforms in human proteome, which was identified in 2011–2014 from all studies they could collect (Hollas et al., 2022). A recent study on 19 cell types from human blood reconstructed 95,979 transcripts from transcriptome sequencing, but identified only in total 4,344 proteoforms using such database in 19 cell types (Melani et al., 2022). In contrast, we identified 6,766 isoforms with unique peptides from one HeLa cell line using our RNC-seq-guided database strategy, much more efficient than the transcriptome-guided database strategy. One possible reason is that the transcriptome-based protein reference database contained too many false entries (i.e., the protein sequences that were actually not translated into proteins): the 3-frame-translation of RNA generates many false entries, and many mRNAs were not translated, especially some AS transcripts (Wang et al., 2013). Excessive number of false entries largely expands the database and decrease the sensitivity and confidence of database search under the widely-used Target-Decoy scheme (Khatun et al., 2013). RNC-seq detects only the translating RNAs and thus creates a minimal protein database, which would solve the problem. At steady state, the translating RNA corresponds to proteins (Zhong et al., 2014). At non-steady state, there might be proteins which were not being synthesized but not fully degraded. Using RNC-seq-guided database may lead to false negatives, but it can still provide confident identifications of those proteins which are being synthesized.

Unique peptides are required to evidence the existence of a proteoform. However, 11,757 proteins in neXtProt database were predicted to have no unique peptides that allows isoform identification (Jeong et al., 2018). The major reason is that most of these isoforms do not have specific junction or exons that can be distinguished from other isoforms. They are unique just due to the unique combination of exons. We can distinguish such isoforms by single-molecule full-length RNC-seq, but the short peptides prevent unique identification at protein level. In such cases, RNC-seq provides indirect evidence of these isoforms. Protein evidence need advances in other experimental strategies, for example, top-down proteome methods.

Long-read RNC-seq also found 4,525 isoforms that were not included in RefSeq databases, and we identified 50 proteins out of these isoforms at protein level. This suggests a hidden proteome from these newly discovered isoforms. It should be noted that our Nanopore RNC-seq yielded only 1.43M reads. When elevating the throughput, considerably more "new" isoforms are expected to be discovered.

In sum, RNC-seq is an efficient and economical way (compared to Ribo-seq) to investigate the proteome of AS variants, and facilitates the functional studies of these isoforms.

MATERIALS AND METHODS

Cell Lines and Reagents

The human hepatocellular carcinoma MHCC97H cell line was kindly provided by Professor Yinkun Liu, Fudan University. MHCC97H cells were cultured in the DMEM (Life Technologies, Carlsbad, CA, United States) medium supplemented with 10% fetal bovine serum (Life Technologies, Carlsbad, CA, United States), 1% penicillin/streptomycin (Life Technologies, Carlsbad, CA, United States) and 10 μg/ml ciprofloxacin, and both of cells were detected free of *mycoplasma* during maintenance and upon experiments.

Ribosome-Nascent Chain Complex Isolation

The method of ribosome-nascent chain complex (RNC) isolation was generated as described before (Wang et al., 2013). In brief, MHCC97H cells were pre-treated with 100 μg/ml cycloheximide (Acmec, Shanghai, China) for 10 min at 37°C, followed by 5 ml pre-colded PBS (Beyotime, Shanghai, China) washes twice and lysis for 30 min on ice by 2 ml pre-cooled human cell lysis buffer (20 mM Tris-HCl, 5 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, 100 μg/ml cycloheximide, 25 units/mL Turbo DNase I, 1% Triton X-100). Cell lysates were clarified by centrifuge at 17000 × g at 4°C for 15 min, supernatants were transferred on the surface of 14.5 ml sucrose cushion (30% sucrose, 20 mM Tris-HCl, 5 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, 100 μg/ml cycloheximide). RNCs were purified by ultra-centrifugation in a Type 70Ti rotor (Beckman Coulter, Brea, CA, United States) at 185,000 × g for 5 h at 4°C.

RNA Extraction and mRNA Sequencing Library Construction

Total MHCC97H RNC-RNA were isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, United States). 1 μg of total RNC-RNA were subjected for library construction. Briefly, PolyA + mRNAs were isolated using VAHTS mRNA Capture Beads

(Vazyme, Jiangsu, Nanjing, China). The sequencing library was constructed by MGIEasy RNA library Preparation kit (MGItech, Guangdong, Shenzhen, China) following the manufacturer's instructions. Libraries were sequenced in a BGI-Seq 500 (MGItech, Guangdong, Shenzhen, China) sequencer at SE100 mode. The raw data of RNA sequencing can be found below: NCBI SRA Bio Project, accession no: GSE198624.

Ribosome Profiling

The method of ribosome profiling was generated as described before (Ingolia et al., 2012) with some modification. In brief, MHCC97H cells were pre-treated with 100 μg/ml cycloheximide for 10 min at 37°C, and then washed twice using 5 ml pre-cooled PBS. The cells were lyzed for 30 min on ice in 2 ml pre-colded human cell lysis buffer. Cell lysates were clarified by centrifuge at 17000 × g at 4°C for 15 min. Purified lysates were treated with 7 units RNase I (Thermo Fisher, Waltham, MA, United States) per OD260. The RNase digestion was performed at 4°C with gentle mixing for 60 min, and then stopped by adding 10 μL of SUPERase-in RNase inhibitor (Thermo Fisher, Waltham, MA, United States). Monosomes were pelleted using ultra-centrifugation in a Type 70Ti rotor at 185,000 × g for 3 h at 4°C.

RNA was isolated using TRIzol reagent. RNA less than 200 nt was isolated using Zymo RNA clean and concentrator kit (Zymo Research, Orange, CA, United States). RNAs with the length of 17–200 nt were loaded on a 15% (w/v) Urea PAGE and resolved by gel electrophoresis. Ribosome footprints were purified by gel recovery according to the Zymo small-RNA PAGE Recovery kit (Zymo Research, Orange, CA, United States). The 3′ dephosphorylation reaction was performed using T4 PNK (New England Biolab, Hitchin, Hertfordshire, United Kingdom) without ATP. The 5' phosphorylation reaction was performed using T4 PNK with 1 mM ATP. RNA was precipitated by adding 39 μL of nuclease free water, 1.0 μL of GlycoBlue coprecipitant (Invitrogen, Carlsbad, CA, United States) and 10 μL of 3 M sodium acetate (pH5.5), 150 μL isopropanol, at −80°C overnight. Ribosome footprints sequencing library was constructed by using MGIEasy Small RNA library preparation kit. Libraries were sequenced in a BGI-Seq 500 sequencer at SE50 mode. The raw data of RNA sequencing can be found below: NCBI SRA Bio Project, accession no: GSE198624.

Protein Extraction and Protein Digestion

MHCC97H Cell line was cultured to 80–90% coverage and treated with 1% SDS lysis buffer (Beyotime, Shanghai, China) and the protein concentration was measured using a BCA kit (Thermo Fisher, Waltham, MA, United States).

The protein digestion was performed by filter-aided sample preparation (FASP) (Wisniewski et al., 2009) as we previously described (Lu et al., 2019). In brief, firstly, 1 mg protein samples were reduced and alkylated using dithiothreitol solution (DTT) (Solarbio, Beijing, China) and iodoacetamide solution (IAA) (I6125, Merck, Kenilworth, NJ, United States) at a final concentration of 4 M urea (8 M urea in 0.1 M Tris-HCl, pH 8.5). Secondly, all the solution was transferred to a 10KD ultrafiltration tube (Merck, Kenilworth, NJ, United States) and centrifuged at 12000 g, and then washed 3 times with 50 mM TEAB (Thermo Fisher, Waltham, MA, United States). Thirdly, trypsin (V5280, Promega, Madison, WI, United States) was added in a ratio of 1:40, and incubated overnight at 37°C. The peptides were collected into low-binding collection tube (Thermo Scientific™, Waltham, MA, United States) and then measured the concentration using Pierce Quantitative Fluorometric Peptide Assay (Thermo Scientific™, Waltham, MA, United States). Finally, the peptides were freeze-dried and stored at −80°C.

Data-Dependent Acquisition Mass Spectrometry

Firstly, the total peptides were fractionated using high-pH reverse-phase liquid chromatography (RPLC). Specifically, 600 μg peptide was re-dissolved in 100 μL buffer A (2% (v/v) ACN, pH 10), and loaded onto the C18 column (4.6 × 250 mm, C18, 3 μm, 186003581, Waters, Milford, MA, United States). The elution gradient was buffer B (98% ACN, pH 10; flow rate, 800 μL/min) for 65 min, the elution gradient was as follows: 5% B, 0 min; 5% B, 6 min; 37% B, 28 min; 46% B, 45 min; 90% B, 46 min; 90% B, 54 min; 95% B, 55 min; 95% B, 65 min. The eluted peptides were collected every minute from the 6th minute until the 54th minute, and then the front, middle and rear fractions were combined into 16 fractions with equal peak area, finally, the fractionated peptides were freeze-dried.

Secondly, the 16 fractionated peptides were redissolved in 0.5% (V/V) trifluoroacetic acid (TFA) (Macklin, Shanghai, China) and were desalted using a Mono tip C18 columns (Shimadzu, Kyoto, Japan) following the manufacturer's instructions and freeze-dried.

Finally, the 16 fractions of the desalted peptide were re-dissolved in 0.1% (V/V) formic acid (FA) (Thermo Scientific™, Waltham, MA, United States) and then preformed DDA analysis by using Orbitrap Fusion Lumos mass spectrometer equipped with EASY-nLC 1200 system (Thermo Scientific™, Waltham, MA, United States). 2 μg of each fractions peptides were loaded on a nano trap column (C18, 150 μm × 20 mm, 1.9 μm, homemade), and then separated onto an analytical column (C18, 150 μm × 300 mm, 1.9 μm, homemade) using a 120 min linear gradient (solvent A: 98% H2O, 2% ACN, 0.1% FA; solvent B: 98% ACN, 2% H2O, 0.1% FA) at a flow rate of 600 nL/min. The detailed solvent gradient was as follows: 5–12% B, 28 min; 12–24% B, 58 min; 24–38% B, 25 min; 38–95% B, 1 min; 95% B, 8 min. The MS1 scan was acquired from 350 to 1500 m/z with a resolution of 120 k, the MS2 scans were performed at a resolution of 15 k with an isolation window of 1.6 m/z, the cycle time was set to 3s with a dynamic exclusion of 30s. All the MS raw data for DDA are publicly available in ProteomeXchange with identifier PXD032201.

AS Event Calling

The HeLa Ribo-seq datasets (accession SRR3306589) (Park et al., 2016) and HeLa RNC-seq datasets (accession SRR6929904) (Li et al., 2018) were obtained from NCBI. We extract the ribosomal RNA sequences in the NCBI RefSeq-RNA database (downloaded

on 6 December 2019) according to refFlat annotation (downloaded on 17 January 2020) as a human rRNA reference dataset. For full-length translating (RNC) mRNA-seq datasets, reads were mapped to rRNA reference sequences using FANSe3 (Zhang et al., 2021) (Release version 3.13) with the parameters–S12 –E4 in HeLa (50bp read lengths), and–S14 -E4 in MHCC97H (100bp read lengths). For Ribo-seq datasets, reads were mapped to rRNA reference sequences using FANSe3 with the parameters–S10 -E2 -U1. Reads that can be aligned to rRNA reference sequences will be considered rRNA reads and discarded. The filtered reads from each sample were mapped to the NCBI RefSeq-RNA database using FANSe3 with unique mode based on the same parameters as above. The UCSC refFlat annotation was used for isoform calling. Isoforms with uniquely-mapped read count≥10 were considered as detected.

Splice Junction Analysis

RNC-seq and Ribo-seq were mapped to the GRCh38 no-alt analysis set (accession GCA_00001405.15) using STAR (2.5.0a) (Dobin et al., 2013) with the gtf option (GRCh38 full analysis set, accession GCA_000001405.15). The total number of splice junctions detected for each sample was taken from the log.final.out file printed by the aligner STAR (Dobin et al., 2013). We defined "Annotated (SJDB) Juntcion" as the identified canonical splice junctions, and "non-canonical junction" as potential novel junctions.

Protein Isoforms Identification

The MHCC97H and HeLa custom protein isoforms databases were built by translating identified AS isoforms into protein sequences. neXtProt (release 2020-07-17) database was used as a negative control. pFind (version3.1.4) (Wang et al., 2007) was utilized to search protein isoforms in Hela (accession PXD004452) (Bekker-Jensen et al., 2017) and MHCC97H mass spectrometric datasets. The FDR threshold was set to 0.01 at both peptide level and protein level. The carbamidomethyl [C] was set as fixed modification, and oxidation [M] as variable modification during the search. The product ion tolerance was set as default parameters and precursor mass tolerance was set to 10 ppm. The missed cleavage was set to 2 for each peptide.

Full-Length Single-Molecule RNC-Seq

Direct RNA Sequencing Kit (Oxford Nanopore Technologies plc, Oxford, United Kingdom) was used for full-length RNC-seq library preparation. The prepared library was load into a MinION flow cell (Oxford Nanopore Technologies plc, Oxford, United Kingdom) and sequenced on MinION device (Oxford Nanopore Technologies plc, Oxford, United Kingdom). Base-calling was performed with MinKNOW (V 3.6.5). Reads were aligned to GRCh38 no-alt analysis set (accession GCA_00001405.15) using minimap2 v2.7-r654 (Li, 2018) in spliced alignment mode with the command: minimap2 -ax splice -uf -k14 -secondary = no. FLAIR *correct* (v1.5) (Tang et al., 2020) was used to correct the splice-site boundaries of reads. All splice sites were assessed for validity by checking for support in genome annotation file (GRCh38 full analysis set, accession GCA_000001405.15). Splice junctions were extracted from matched RNC-seq data, and only the junctions supported by > 5 uniquely mapped short reads were considered valid. Incorrect splice sites were replaced with the nearest valid splice site within a 10-nt window. Isoforms were assembled and identified using FLAIR *collapse* (v1.5) (Tang et al., 2020) with the default settings. We filtered out the single exon transcripts to increase the confidence of the identification results. The raw data of RNA sequencing can be found below: NCBI SRA Bio Project, accession no: GSE198624.

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 *Materials and Methods*.

AUTHOR CONTRIBUTIONS

CW: analyzing data, writing the initial manuscript, designing experiment; XL: conducting ribo-seq experiments with assistance by JJ, designing experiments; SL: conducting cell culture and MS experiments with assistance by ZS, revised the manuscript; HW: analyzing data, writing the initial manuscript, revised the manuscript; DL: analyzing data; JZ: revising the manuscript; Q-YH: supervision of the study, providing MS resource; YC: concept of the study, designing experiments, conducting RNC-seq experiments; GZ: concept of the study, designing experiments, writing the final manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Ministry of Science and Technology of China, National Key Research and Development Program (2017YFA0505001/2017YFA0505101/ 2018YFC0 910201/2018YFC0910202), the National Natural Science Funds of China (81802916/82002949), Guangdong Key R&D Program (2019B020226001), State Key Laboratory of Respiratory Disease, Guangdong-Hong Kong-Macao Joint Laboratory of Respiratory Infectious Disease (02-000-2101-5061) and the Fundamental Research Funds for the Central Universities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.895746/full#supplementary-material

REFERENCES

Baralle, F. E., and Giudice, J. (2017). Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27

Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., et al. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* 4, 587–599. doi:10.1016/j.cels.2017.05.009

Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., et al. (2017). Nanopore Long-Read RNAseq Reveals Widespread Transcriptional Variation Among the Surface Receptors of Individual B Cells. *Nat. Commun.* 8, 16027. doi:10.1038/ncomms16027

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Frankish, A., Uszczynska, B., Ritchie, G. R., Gonzalez, J. M., Pervouchine, D., Petryszak, R., et al. (2015). Comparison of GENCODE and RefSeq Gene Annotation and the Impact of Reference Geneset on Variant Effect Prediction. *BMC Genomics* 16 (Suppl. 8), S2. doi:10.1186/1471-2164-16-S8-S2

Gao, Y., Ping, L., Duong, D., Zhang, C., Dammer, E. B., Li, Y., et al. (2021). Mass-Spectrometry-Based Near-Complete Draft of the *Saccharomyces cerevisiae* Proteome. *J. Proteome Res.* 20, 1328–1340. doi:10.1021/acs.jproteome.0c00721

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell* 154, 240–251. doi:10.1016/j.cell.2013.06.009

Hollas, M. A. R., Robey, M. T., Fellers, R. T., LeDuc, R. D., Thomas, P. M., and Kelleher, N. L. (2022). The Human Proteoform Atlas: a FAIR Community Resource for Experimentally Derived Proteoforms. *Nucleic Acids Res.* 50, D526–D533. doi:10.1093/nar/gkab1086

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., and Weissman, J. S. (2012). The Ribosome Profiling Strategy for Monitoring Translation *In Vivo* by Deep Sequencing of Ribosome-Protected mRNA Fragments. *Nat. Protoc.* 7, 1534–1550. doi:10.1038/nprot.2012.086

Jeong, S.-K., Kim, C.-Y., and Paik, Y.-K. (2018). ASV-ID, a Proteogenomic Workflow to Predict Candidate Protein Isoforms on the Basis of Transcript Evidence. *J. Proteome Res.* 17, 4235–4242. doi:10.1021/acs.jproteome.8b00548

Khatun, J., Yu, Y., Wrobel, J. A., Risk, B. A., Gunawardena, H. P., Secrest, A., et al. (2013). Whole Human Genome Proteogenomic Mapping for ENCODE Cell Line Data: Identifying Protein-Coding Regions. *BMC Genomics* 14, 141. doi:10.1186/1471-2164-14-141

Li, D., Lu, S., Liu, W., Zhao, X., Mai, Z., and Zhang, G. (2018). Optimal Settings of Mass Spectrometry Open Search Strategy for Higher Confidence. *J. Proteome Res.* 17, 3719–3729. doi:10.1021/acs.jproteome.8b00352

Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191

Liu, W., Xiang, L., Zheng, T., Jin, J., and Zhang, G. (2018). TranslatomeDB: a Comprehensive Database and Cloud-Based Analysis Platform for Translatome Sequencing Data. *Nucleic Acids Res.* 46, D206–D212. doi:10.1093/nar/gkx1034

Lu, S., Zhang, J., Lian, X., Sun, L., Meng, K., Chen, Y., et al. (2019). A Hidden Human Proteome Encoded by 'non-Coding' Genes. *Nucleic Acids Res.* 47, 8111–8125. doi:10.1093/nar/gkz646

Melani, R. D., Gerbasi, V. R., Anderson, L. C., Sikora, J. W., Toby, T. K., Hutton, J. E., et al. (2022). The Blood Proteoform Atlas: A Reference Map of Proteoforms in Human Hematopoietic Cells. *Science* 375, 411–418. doi:10.1126/science.aaz5284

Paik, Y.-K., Jeong, S.-K., Omenn, G. S., Uhlen, M., Hanash, S., Cho, S. Y., et al. (2012). The Chromosome-Centric Human Proteome Project for Cataloging Proteins Encoded in the Genome. *Nat. Biotechnol.* 30, 221–223. doi:10.1038/nbt.2152

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing. *Nat. Genet.* 40, 1413–1415. doi:10.1038/ng.259

Park, J.-E., Yi, H., Kim, Y., Chang, H., and Kim, V. N. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol. Cell* 62, 462–471. doi:10.1016/j.molcel.2016.04.007

Reichow, S. L., Hamma, T., Ferre-D'Amare, A. R., and Varani, G. (2007). The Structure and Function of Small Nucleolar Ribonucleoproteins. *Nucleic Acids Res.* 35, 1452–1464. doi:10.1093/nar/gkl1172

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and its Applications. *Genomics, Proteomics Bioinforma.* 13, 278–289. doi:10.1016/j.gpb.2015.08.002

Sessegolo, C., Cruaud, C., Da Silva, C., Cologne, A., Dubarry, M., Derrien, T., et al. (2019). Transcriptome Profiling of Mouse Samples Using Nanopore Sequencing of cDNA and RNA Molecules. *Sci. Rep.* 9, 14908. doi:10.1038/s41598-019-51470-9

Smith, C. M., and Steitz, J. A. (1998). Classification of Gas5 as a Multi-Small-Nucleolar-RNA (snoRNA) Host Gene and a Member of the 5′-Terminal Oligopyrimidine Gene Family Reveals Common Features of snoRNA Host Genes. *Mol. Cell Biol.* 18, 6897–6909. doi:10.1128/mcb.18.12.6897

Sulakhe, D., D'Souza, M., Wang, S., Balasubramanian, S., Athri, P., Xie, B., et al. (2019). Exploring the Functional Impact of Alternative Splicing on Human Protein Isoforms Using Available Annotation Sources. *Brief. Bioinform* 20, 1754–1768. doi:10.1093/bib/bby047

Tang, A. D., Soulette, C. M., van Baren, M. J., Hart, K., Hrabeta-Robinson, E., Wu, C. J., et al. (2020). Full-length Transcript Characterization of SF3B1 Mutation in Chronic Lymphocytic Leukemia Reveals Downregulation of Retained Introns. *Nat. Commun.* 11, 1438. doi:10.1038/s41467-020-15171-6

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., et al. (2019). A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues. *Mol. Syst. Biol.* 15, e8503. doi:10.15252/msb.20188503

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* 456, 470–476. doi:10.1038/nature07509

Wang, L.-h., Li, D.-Q., Fu, Y., Wang, H.-P., Zhang, J.-F., Yuan, Z.-F., et al. (2007). pFind 2.0: a Software Package for Peptide and Protein Identification *via* Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 21, 2985–2991. doi:10.1002/rcm.3173

Wang, T., Cui, Y., Jin, J., Guo, J., Wang, G., Yin, X., et al. (2013). Translating mRNAs Strongly Correlate to Proteins in a Multivariate Manner and Their Translation Ratios Are Phenotype Specific. *Nucleic Acids Res.* 41, 4743–4754. doi:10.1093/nar/gkt178

Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* 6, 359–362. doi:10.1038/nmeth.1322

Zhang, G., Zhang, Y., and Jin, J. (2021). The Ultrafast and Accurate Mapping Algorithm FANSe3: Mapping a Human Whole-Genome Sequencing Dataset within 30 minutes. *Phenomics* 1, 22–30. doi:10.1007/s43657-020-00008-5

Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T., and Zhang, G. (2019). Translatomics: The Global View of Translation. *Int. J. Mol. Sci.* 20. doi:10.3390/ijms20010212

Zhao, S., Zhang, Y., Gamini, R., Zhang, B., and von Schack, D. (2018). Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical RNA Sequencing: polyA+ Selection versus rRNA Depletion. *Sci. Rep.* 8, 4781. doi:10.1038/s41598-018-23226-4

Zhong, J., Cui, Y., Guo, J., Chen, Z., Yang, L., He, Q.-Y., et al. (2014). Resolving Chromosome-Centric Human Proteome with Translating mRNA Analysis: a Strategic Demonstration. *J. Proteome Res.* 13, 50–59. doi:10.1021/pr4007409

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.

frontiers | Frontiers in Genetics

MINI REVIEW
published: 22 June 2022
doi: 10.3389/fgene.2022.918509

The Epigenetic Regulation of Nonhistone Proteins by SETD7: New Targets in Cancer

Chengyao Chiang[1,2†], Heng Yang[2†], Lizhi Zhu[2†], Chunlan Chen[2], Cheng Chen[2], You Zuo[1] and Duo Zheng[2*]*

[1]Southern University of Science and Technology, Yantian Hospital, Shenzhen, China, [2]Guangdong Provincial Key Laboratory of Regional Immunity and Diseases, Department of Cell Biology and Genetics, Department of Pharmacy, Shenzhen University International Cancer Center, School of Medicine, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital (Shenzhen Institute of Translational Medicine), Shenzhen University, Shenzhen, China

Epigenetic modifications are essential mechanism by which to ensure cell homeostasis. One such modification is lysine methylation of nonhistone proteins by SETD7, a monomethyltransferase containing SET domains. SETD7 methylates over 30 proteins and is thus involved in various classical pathways. As such, SETD7 has been implicated in both the basic functions of normal tissues but also in several pathologies, such as cancers. In this review, we summarize the current knowledge of SETD7 substrates, especially transcriptional-related proteins and enzymes, and their putative roles upon SETD7-mediated methylation. We focus on the role of SETD7 in cancers, and speculate on the possible points of intervention and areas for future research.

Keywords: SETD7, epigentics, non-histone substrate, transcriptional factor, protein methylaiton

OPEN ACCESS

Edited by:
Andy T. Y. Lau,
Shantou University, China

Reviewed by:
Kangdong Liu,
Zhengzhou University, China

****Correspondence:***
You Zuo
183401043@qq.com
Duo Zheng
dzheng@szu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *12 April 2022*
Accepted: *27 May 2022*
Published: *22 June 2022*

Citation:
Chiang C, Yang H, Zhu L, Chen C, Chen C, Zuo Y and Zheng D (2022) The Epigenetic Regulation of Nonhistone Proteins by SETD7: New Targets in Cancer. Front. Genet. 13:918509. doi: 10.3389/fgene.2022.918509

1 INTRODUCTION

The alteration of the gene expression profile in somatic cells is the main cause of human diseases. Such alterations can be driven by DNA methylation, posttranscriptional modification (PTM) of proteins, and noncoding RNAs (Esteller, 2007)—otherwise known as epigenetic modifications. Of the various PTMs, phosphorylation and acetylation help to modulate kinase activity and signal transduction. Ubiquitination and sumoylation regulate protein stability, while methylation influences protein interactions, function, stability, activity, structure and subcellular location (Wang et al., 2017). Many nonhistone proteins are also methylated; for example, lysine (K) can be mono-, di- or tri-methylated, while arginine (R) can be mono- or di-methylated (Pek et al., 2012). We are aware of around more than 50 lysine methyltransferases, 20 lysine demethylases (Han et al., 2019) and 10 proteins arginine methyltransferases (Wu et al., 2021) that are involved in protein methylation, either as a "writer" (adding methyl groups), a "reader" (recognizing the methyl signal), or an "eraser" (removing methyl groups). These proteins regulate several biological processes in both health and disease contexts.

SET domain containing lysine methyltransferase 7 (SETD7) is a 40 kDa protein containing 366 amino acids that is responsible for transferring the monomethyl group to lysine of its substrates from cofactor S-adenosylmethionine (AdoMet) (Fick et al., 2016). Similar to most lysine methyltransferases, the SET domain is required for catalysis, with histidine 297 the critical site for its methyltransferase activity (Nishioka et al., 2002). The methylated lysine targeted by SETD7 usually follows after the consensus motif of [K/R]-[A/S/T] (Del Rizzo and Trievel, 2011). SETD7 contains three membrane occupation and recognition nexus (MORN) motifs in the N-terminal

region, which likely mediate SETD7's interaction with the plasma membrane when the protein is not in the nucleus (Bivona et al., 2006).

SETD7, also known as SET7/9, KIAA1717, or KMT7, was first identified as a histone H3-lysine 4-specific (H3K4) methyltransferase that changes the affinity between histone 3 and double-stranded DNA to regulate gene expression (Wang et al., 2001). Since then, other have showed that SETD7 depletion has little impact on H3K4 methylation status in certain circumstances (Ea and Baltimore, 2009; Gaughan et al., 2011; Lehnertz et al., 2011), implying a more critical role of SETD7 on nonhistone proteins. SETD7 can modify many substrates, including histones and nonhistone transcription factors, transcriptional coactivators, hormone receptors, DNA methyltransferases, and other histone methyltransferases. The role of SETD7 is determined by the function of its substrates. Indeed, more than a dozen SETD7 nonhistone substrates have now been discovered (Keating and El-Osta, 2013).

The fates of the proteins modified by SETD7-mediated lysine methylation are diverse. They range from nucleus to cytoplasm and are implicated in gene transactivation, signaling transduction and regulation of hemostasis. Here, we discuss the known SETD7 substrates and their putative roles when methylated by SETD7 in regulation of cell cycle, apoptosis and response to external stimulation in human cancers.

2 SETD7-MEDIATED SUBSTRATE MODIFICATIONS AND THEIR ROLE IN CANCER

2.1 Cell Cycle and Apoptosis Regulation

Cell cycle progression and cell apoptosis are coupled intimately. These important decisions of cell proliferation or cell death are likely to be controlled by more than one signal and are necessary to ensure a proper cellular response. Some proteins can involve in both cell division and programmed cell death, such as p53, pRb, E2F, which are responsible for reacting cellular stresses and regulating checkpoint-associated proteins, including CDK2 (Engeland, 2018). Basically, cell cycle is regulated by checkpoints which link the cell cycle to apoptotic pathways and ensure that cell cycle events toward the correct order, otherwise initiating cell apoptosis. Data thus far, programmed cell death and cell cycle share common molecular mechanisms, which are modulated by SETD7 via its methyltransferase activity.

TP53 and SIRTUIN 1 (SIRT1)

TP53, a tumor suppressor regulating cell cycle and controlling cell fate, is highly frequent loss-of-function in most of cancers, which is facilitated to cancer progression (Blagih et al., 2020). The TP53 protein can directly binds with transcription factors, including Sp1, TBP and NF-Y, to suppress genes expression (Liebl and Hofmann, 2021). Additionally, TP53 influences CDK-cyclin interaction through up-regulation of its downstream genes, such as *CDKN1A* (encode $p21^{Cip1/Waf1}$ protein), resulting in down-regulation of cell cycle-associated genes (Hu et al., 2021). On the other side, several pro-apoptotic BCL-2 family members, including *BAX*, *BBC3* (*PUMA*) and *PMAIP1* (*NOXA*) are activated by TP 53, which leads to cell apoptosis (Parrales and Iwakuma, 2015). SETD7-mediated methylation of TP53 K372 potentiates apoptosis and facilitates the transcriptional initiation of TP53-downstream genes $p21^{Cip1/Waf1}$ (Chuikov et al., 2004), to decelerate cell cycle progression. Others showed that murine TP53 K369 is also a potential target residue for methylation by SETD7. Methylation of this residue promotes TP53 acetylation by Tip60 and potentiates the expression of downstream genes, including $p21^{Cip1/Waf1}$ and PUMA, *in vivo* (Kurash et al., 2008; Campaner et al., 2011). Methylation-dependent TP53 activation indicates a tumor suppressor role for SETD7 in cancer cells in both humans and mice.

Some epigenetic modifiers of TP53 are also reported as a substrate of SETD7, such as SIRT1 which is a nicotinamide adenine dinucleotide-dependent deacetylase, involving in various cell metabolic processes (Chen et al., 2021). SIRT1 is generally considered as an oncoproteins in leukemia and prostate cancer due to suppressing several tumor suppressors, such as TP53, via its deacetylase activity (Yousafzai et al., 2021). However, SIRT1-mediated regulation of TP53 is inhibited by SETD7-dependent methylation at K233, K235, K236 and K238 on SIRT1. However, multi-methylation does not influence SIRT1 deacetylase activity, which may induce a conformational change of SIRT1 to avoid TP53 binding (Liu et al., 2011). In addition to showing that SETD7 directly methylates TP53, the researchers showed an alternative way in which the transactivation capacity of TP53 can be enhanced during the DNA damage response. Taken together, SETD7 serves as a tumor suppressor to enhance TP53 activity by a novel manner through abolishment of SIRT1 and TP53 interaction.

E2 Promoter-Binding Factor 1 (E2F1) and Retinoblastoma Tumour Suppressor Protein (pRB)

E2F1 is a transcription factor responsible for the expression of DNA damage-induced genes, such as *CCNE1* which accelerates DNA replication and progression from the G1 to S phase of the cell cycle (Fouad et al., 2020). E2F1 also up-regulates downstream pro-apoptotic genes, including *TP73*, and activates programmed cell death through TP53-independent manner (Udayakumar et al., 2010). K185 on E2F1 is methylated by SETD7, which prevents E2F1 accumulation during DNA damage and activation of its proapoptotic target gene *TP73* via destabilization E2F1 by ubiquitination and degradation (Kontaki and Talianidis, 2010). However, other study reveals that SETD7 and LSD1 regulate E2F1-mediated apoptosis upon DNA damage. Methylation of K185 on E2F1 by SETD7 leads to E2F1 stabilization and up-regulation of proapoptotic genes *TP73* and *BIM*, whereas, SETD7-mediated effects are reversed by LSD1 (Xie et al., 2011). Interestingly, other study showed a negative correlation between E2F1 and SETD7 *in vivo* and in clinical specimens: Overexpression of E2F1 leads to SETD7 downregulation and EGFR and Snail upregulation in breast cancer cells (Montenegro et al., 2016). In the case as regulating its substrate, SETD7 is modulated by E2F1 either, which reveals a novel regulatory mechanism in SETD7 expression. Additionally, the threshold of expression of both E2F1 and SETD7 is indicated as a critical

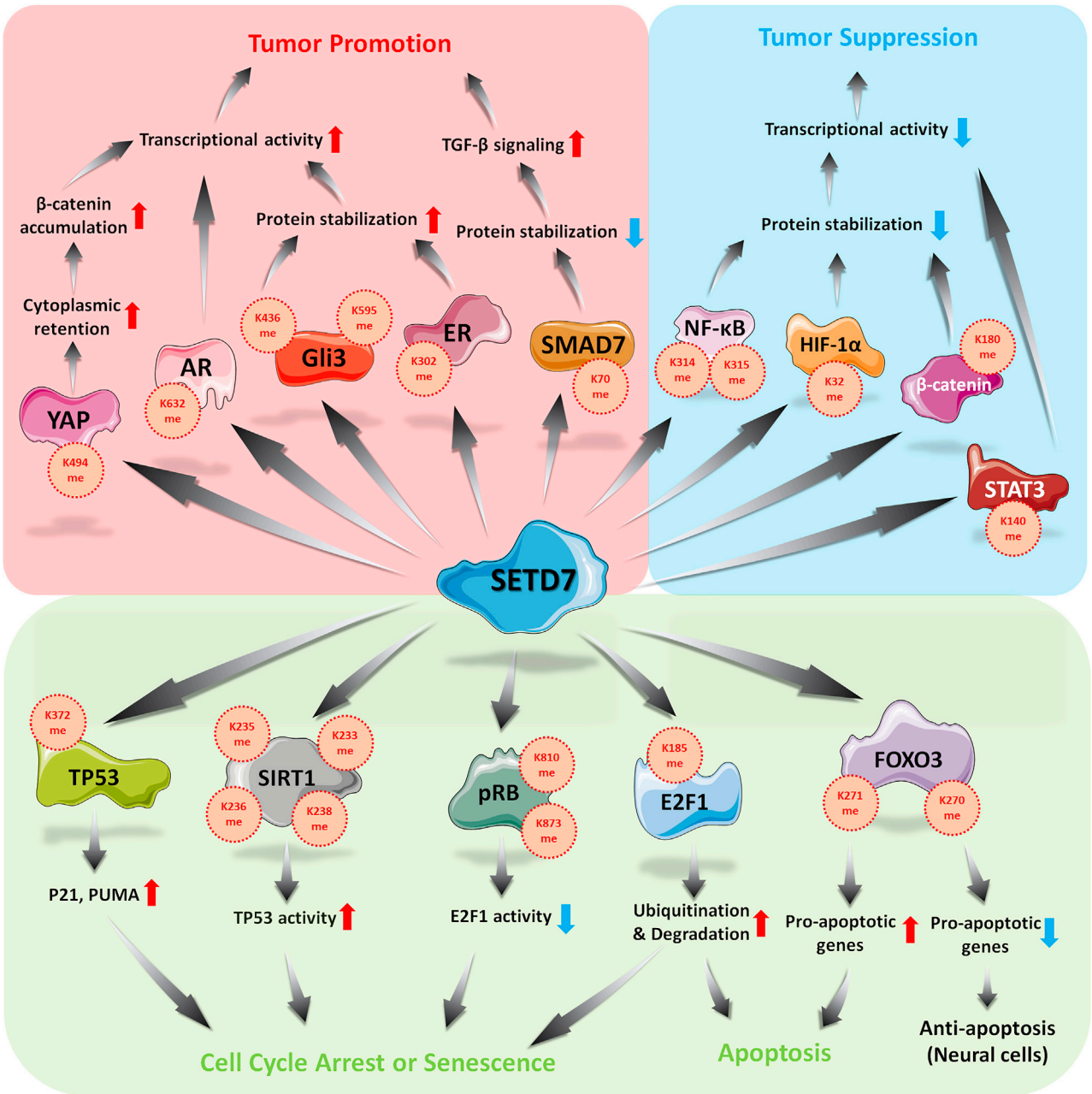


FIGURE 1 | Roles of SETD7 in regulation of cell cycle, apoptosis and external stimulations.

event to control the cell fate (Lezina et al., 2014). Since, the controversial role of E2F1 and its fully activity might also be determined the ubiquitinated level or types after SETD7-mediated methylation.

pRb functions in early cell cycle control by negatively regulating entry into S-phase by suppression of E2F1. In this way, pRb serves as a tumor suppressor, as well as usually being functionally inactivated in retinoblastoma, osteosarcoma, lung, breast and hepatic cancers (Giacinti and Giordano, 2006). Growth control by pRb is influenced by CDK phosphorylation, in which serial phosphorylation events that drive cell cycle transitions regulate pRb-dependent cell cycle progression (Mandigo et al., 2022). SETD7-mediated pRb methylation at K873 is required for pRb-dependent cell cycle arrest, transcriptional repression and pRb-dependent differentiation possibly by enhancing the interaction between pRb and the heterochromatin protein HP1 (Munro et al., 2010). The same group also demonstrated a novel mechanism in the regulation of E2F1 transactivation in which K810 methylation on pRb by SETD7 is essential for impeding cyclin/CDK recognition

and the subsequent phosphorylation of the associated serine residue. As a result, pRb remains in the hypophosphorylated, growth-suppressing state (Carr et al., 2011). These data suggest that SETD7 serves as a tumor suppressor and cooperates with pRb in cell cycle control.

Forkhead Box O3 (FOXO3)

FOXO transcription factors have a critical role in longevity, tumor suppression and oxidative stress-induced neuronal cell death by regulating the expression of various target genes (Fasano et al., 2019). Activation of FOXO3 induces cell cycle arrest and promotes apoptosis in gastric cancer (Li M. et al., 2020), and pancreatic cancer (Usami et al., 2020). Other study shows that FOXO3 interacts with ERα and inhibits its transcriptional activity to suppress breast cancer progression (Zou et al., 2008). In addition, low expression of FOXO3 is associated with poorly prognostic outcome in estrogen-dependent breast cancer (Yin et al., 2020) and colorectal cancers (Bullock et al., 2013). FOXO3-mediated transcription and oxidative stress-induced neuronal apoptosis are negatively regulated by SETD7-dependent K270 methylation, as well as downregulating proapoptotic genes *BIM* (Xie et al., 2012). Interestingly, others showed that K271 on FOXO3 was methylated by SETD7, which decreases FOXO3 protein stability while moderately enhancing FOXO3-dependent activation of pro-apoptotic genes, which may in turn affect FOXO3's ability to promote tumor suppression (Calnan et al., 2012). The role of SETD7 in methylation of K270 and K271 of FOXO3 is opposite and the detail molecular mechanism is demanded more evidences to clarify.

The functions of SETD7 to its substrates and its effects in cell cycle and apoptosis regulation was summarized in **Figure 1**. For some controversial substrates, such as E2F1, FOXO3, whether tissue-specific interacting proteins or cooperation of other epigenetic modifications involved in SETD7-mediated regulation are such interesting issues, which is worth for further investigation in order to elucidate the exactly physiological effects of SETD7-substrates axis.

2.2 Regulation of External Stimulation

The corresponding responses of cells to various stimuli from micro-environment are essential strategies to homeostasis. Generally, transcription factors-mediated genes expression is responsible for implementing such reactions, for example, HIF-1α is activated by hypoxia stress and up-regulates *VEGF* (encode vascular endothelial growth factor) and *EPO* (encode erythropoietin) to overcome hypoxic condition (Plastino et al., 2021). Besides, cellular factors, such as cytokines, activate their corresponding receptors and downstream transcription factors or regulators, including SMADs (Zhang T. et al., 2020), STATs (Verhoeven et al., 2020), or NFκB (Zinatizadeh et al., 2021). Once such cellular responses might play pathological roles if the modulatory mechanisms are dysfunctional, as well as SETD7 showing its significant part among these regulatory processes.

Yes-Associated Protein and Beta-Catenin

YAP, a transcriptional co-activator belonged to Hippo pathway, is required for the growth of embryonic tissues, wound healing, and organ regeneration (Zanconato et al., 2016b). Activated YAP translocates into the nucleus and cooperates with transcriptional co-activator PDZ-binding motif (TAZ) to up-regulate proliferative- and anti-apoptotic-related genes, which is regulated by cell-intrinsic and -extrinsic signals, such as oxidative stress or nutrient-depletion (Koo and Guan, 2018; Moya and Halder, 2019). Unsurprisingly, YAP is hyperactivated in human malignancies (Moroishi et al., 2015), which can reprogram cancer cells into cancer stem cells and promote tumor initiation, progression and metastasis (Nguyen and Yi, 2019). Thus, YAP is emerging as a potentially therapeutic target for clinical application (Zanconato et al., 2016a). The monomethylation of K494 on YAP by SETD7 is critical for YAP cytoplasmic retention. This event thus represents a methylation-dependent checkpoint in the Hippo pathway (Oudhoff et al., 2013). As such, SETD7-dependent methylation of YAP facilitates Wnt-induced nuclear accumulation of β-catenin, linking the Wnt/β-catenin and Hippo/YAP pathways during intestinal regeneration and tumorigenesis (Oudhoff et al., 2016). These data indicate the triple layered regulation and crosstalk of two signaling pathways in an intestinal model.

On the other hand, β-catenin, a positive regulator in the canonical Wnt signaling pathway (Zhang and Wang, 2020), is activated by Wnt protein binding with frizzled receptors and translocates into the nucleus to participate in the transactivated complex (Cheng et al., 2019), promoting cancer progression by upregulation of proliferative-related genes (Zhang and Wang, 2020). Absence of Wnt stimulation, β-catenin is strictly modulated by glycogen synthase kinase-3 beta (GSK-3β) via S33 and S37 phosphorylation, which is recognized by E3 ligase β-TrCP, leading to ubiquitin-dependent proteosomal degradation (Valenta et al., 2012). Accumulating evidences refer oncogenic role of β-catenin in multiple human cancers, including solid tumors and hematological malignancies (Clevers and Nusse, 2012). Under condition of oxidative stress, SETD7 interacts with β-catenin and methylates it at K180, which promotes its phosphorylation by GSK3-β and subsequent degradation. The result is the suppression of downstream c-myc and cyclin D1 and inhibition of cell proliferation (Shen et al., 2015). However, this model currently lacks corresponding animal model and supportive clinical evidence. SETD7 is indeed emerging as a negative regulator of the Wnt/β-catenin pathway depending on the tissue or physical context. Notably, increasing epigenetic modifications on β-catenin is illustrated (Valenta et al., 2012), thus, it is interesting and necessary to be addressed that various of modifications regulate the same protein in certain physiological status.

NF-κB

NF-κB, a critical transcription factor in broad range of physiological functions, including inflammation, cell growth and programmed cell death, mainly retains in cytoplasm. Activated NF-κB, formed by RelA (also called p65) and p50 subunit (Zinatizadeh et al., 2021), translocates into the nucleus and up-regulates target genes after diverse extracellular stimuli, including TNF-α (Hoesel and Schmid, 2013), which might benefit for cell proliferation and survival in leukemia, melanoma, liver,

breast, prostate and colorectal cancers (Dolcet et al., 2005). K314 and K315 on RelA can be methylated by SETD7, leading to destabilization of RelA in a ubiquitination-mediated manner (Yang et al., 2009), which results in downregulation of tumor-associated genes, such as IL-6, IL-8 and NOS-2. A contradictory role of SETD7 in the NF-κB pathway has also been described in which K37 on RelA was methylated, leading to stabilization of the RelA-DNA complex and enhanced expression of NF-κB-regulated genes (Ea and Baltimore, 2009). Similar results were also shown in diabetes models in which SETD7 interacts with RelA, facilitating the nuclear translocation of RelA and promoted function of NF-κB to transactivate downstream genes (Fujimaki et al., 2015; Chokpaisarn et al., 2017). NF-κB regulation by SETD7 might depend on the cellular context, tissue specificity or particular physiological condition, such as in cancer cells or diabetes model.

Hypoxia Inducible Factor

Hypoxia inducible factor-1α (HIF-1α) is a transcription factor involved in adaption of low oxygen concentration. Under normoxia, HIF-1α is strictly modulated by an E3 ligase von Hippel-Lindau (VHL), which induces ubiquitination-dependent proteasomal degradation of HIF-1α (Semenza, 2003). Activated HIF-1α translocates into the nucleus and activates targeting genes, which participates in tumor angiogenesis, metastasis, invasion and glucose homeostasis in various cancer cell lines (Elzakra and Kim, 2021; Satija et al., 2021). Here, K32 methylation of HIF-1α by SETD7 promotes HIF-1α degradation in the nucleus and thus the inhibited expression of downstream genes in a proline hydroxylation-independent manner. This effect can be restored upon exposure to the demethylase LSD1 (Kim et al., 2016). Both HIF-1α and HIF-2α are reported substrates of SETD7 and are methylated on K32 and K29, respectively, due to their homologous of sequence. Interestingly, *SETD7* expression is suppressed under hypoxic conditions (Liu et al., 2015). Others have reported that SETD7 is a negative regulator of HIF-1α and downregulates HIF-1α target genes, such as *GLUT1, LDHA, PGK1, EPO, PKM2* and *VEGF,* which are upregulated after SETD7 inhibition (Li et al., 2021; Xiaoshi et al., 2021).

Esrtogen Receptor and Androgen Receptor

ER and AR, ligand-dependent transcription factors, are activated by sex hormones and responsible for the regulation of cell proliferation, survival and differentiation (Shafi et al., 2013; Berkel and Cacan, 2021) in breast (Anestis et al., 2020) and prostate cancer (Tan et al., 2015), respectively. As most transcription factors, activated ER or AR translocates into the nucleus and recruits other epigenetic enzymes, such as histone acetyltransferase or methyltransferase, to transactivate targeting genes expression (Waddell et al., 2021). Unsurprisingly, aberrant expression ER and AR are risk factors in many cancers, including prostate, breast and lung cancers (Burstein, 2020). Anti-ER or AR approaches thus seem as effective options for such type of cancers. Here, SETD7-mediated methylation of K302 on ER ensures protein stability and promotes DNA binding activity and the expression of ER-downstream genes, such as *PS2* and progesterone receptor (*PgR*), in breast cancer. These results imply that lysine methylation of ER facilitates to prevent ubiquitination on the same residue by E3 ligases (Subramanian et al., 2008). Similarly, SETD7 interacts directly with AR and enhances AR transcriptional activity by methylating its K632 residue (Gaughan et al., 2011), which is not only plays a proliferative role in prostate cancer but is also involved in TNFR and PTEN/PI3K/AKT signaling (Wang et al., 2018). SETD7 thus seems to be a coactivator of hormone receptors, and in this way helps to promote carcinogenesis. Therefore, ER or AR combines with SETD7 might serve as the panel of prognostic markers or therapeutic targets for patients with such cancers.

Gloma-Associated Oncogene Homolog

GLIs, a family of zinc finger transcription factors, serve as nuclear mediators of the Hedgehog pathway and regulate genes essential for various stages of tumor development and progression (Naruse et al., 2010; Katoh, 2019). Without ligand stimulation, GLIs are suppressed by suppressor of fused (SUFU), leading to cytoplasmic retention (Sasai et al., 2019). Aberrant activation of Hedgehog-GLI axis is reported in human malignancies, including breast, pancreatic, lung and ovarian cancers, which resulted in upregulation of oncogenic genes (Matissek and Elsawa, 2020), such as *BCL2*, *CCND1*, *MYCN*, *NANOG*, *SOX2* and *SNAL1*. As such, GLI family members might be therapeutic targets in various cancers (Niewiadomski et al., 2019). In previous study, GLI3 K436 and K595 residues are methylated by SETD7, which stabilizes GLI3 protein and in turn activates the Sonic Hedgehog pathway, resulting in the expression of downstream genes, including *Ptch1*. These genes promote proliferation, invasion and metastasis of non-small-cell lung cancer cells (Fu et al., 2016). To date, however, a clinical correlation between SETD7 and GLI3 expression at protein level is lacking, which is still fuzzy to figure out the significance of GLI3-dependent oncogenesis by SETD7.

Small Mothers Against Decapentaplegic

SMADs, critical regulators participated in transforming growth factor-beta (TGF-β) signaling, have key roles in development, carcinogenesis and fibrogenesis (Derynck and Zhang, 2003). After receptor activating, SMAD2, 3, and 4 translocate into the nucleus to turn on oncogenic genes, including *SNAIL* and *SLUG* (Zhang T. et al., 2020). While SMAD7 serves as a negative modulator to promote degradation of TGF-β receptor by recruiting E3 ligase SMURF1/2 (Smad ubiquitin-related factor1/2) (Colak and Ten Dijke, 2017). Epigenetic modification on K70 of SMAD7 by SETD7-mediated methylation decreases the protein stability of SMAD7 by ubiquitination-dependent manner via Arkadia E3 ligase in mouse models of pulmonary fibrosis. In SETD7-deficient mice, TGF-β-induced lung fibrosis is highly ameliorated (Elkouris et al., 2016), which indicates SETD7 is a positive regulator in TGF-β signaling, even plays an oncogenic role in TGF-β-mediated cancers, such as breast cancer and glioblastoma (Colak and Ten Dijke, 2017). As such, SETD7 might be a potential therapeutic target for lung fibrosis or cancers. Others reported that SETD7 interacts with SMAD3, but not SMAD2, to ensure

TABLE 1 | SETD7-regulated proteins and methylation sites.

Substrate	Methylation Site	Sequence Around Methylation Site											Study Model	Consequence	Role of SETD7	References
		−5	−4	−3	−2	−1	0	1	2	3	4	5				
Transcriptional-related factors																
TP53	K372	S	H	L	K	S	K	K	G	Q	S	T	293F, U2OS (Osteosarcoma), H1299 (NSCLC)	Enhancement of transactivation	Tumor suppressor	Chuikov et al. (2004)
NF-κB (Rel A)	K314, K315	F	K	S	I	M	K	K	S	P	F	S	MEFs (Mouse embryonic fibroblast), U2OS (Osteosarcoma), A549 (NSCLC)	Protein degradation	Tumor suppressor	Yang et al. (2009)
	K37	M	R	F	R	Y	K	C	E	G	R	S	HEK293T	Stabilization of RelA-DNA complex	Oncoprotein	Ea and Baltimore, (2009)
HIF-1α	K32	R	S	R	R	S	K	E	S	E	V	F	Hela (Cervical cancer), RCC4 (Kidney cancer)	Protein degradation	Tumor suppressor	Kim et al. (2016)
HIF-2α	K29	R	C	R	R	S	K	E	T	E	V	F	RCC4 (Kidney cancer)	Suppression of transactivation	Tumor suppressor	Liu et al. (2015)
ER	K302	M	I	K	R	S	K	K	N	S	L	A	Breast cancer	Protein stabilization	Oncoprotein	Subramanian et al. (2008)
AR	K632	G	A	R	K	L	K	K	L	G	N	L	Prostate cancer	Enhancement of transactivation	Oncoprotein	Gaughan et al. (2011)
Gli3	K436	H	N	K	R	S	K	I	K	P	D	E	NSCLC	Protein stabilization	Oncoprotein	Fu et al. (2016)
	K595	H	E	G	C	N	K	A	F	S	N	A				
E2F1	K185	I	A	K	K	S	K	N	H	I	Q	W	NSCLC	Protein stabilization/degradation	Controversial	Lezina et al. (2016)
β-catenin	K180	V	H	Q	L	S	K	K	E	A	S	R	Hela (Cervical cancer)	Protein degradation	Tumor suppressor	Shen et al. (2015)
SMAD7	K70	A	V	R	G	A	K	G	H	H	H	P	Lung fibroblasts, Hela (Cervical cancer)	Protein degradation	Fibrosis suppressor	Elkouris et al. (2016)
YAP	K494	V	L	A	A	T	K	L	D	K	E	S	Mice intestinal tumor	Cytoplasmic retention	Oncoprotein	Oudhoff et al. (2013)
TAF10	K189	S	R	S	K	S	K	D	R	K	Y	T	HEK293, F9 Embryonic carcinoma	Enhancement of TAF10-RNA polymerase II complex	Controversial	Kouskouti et al, (2004)
FOXO3	K270	G	R	A	A	K	K	K	A	A	L	Q	HEK293T	Protein degradation	Neural apoptosis suppressor	Xie et al. (2012)
	K271	R	A	A	K	K	K	A	A	L	Q	A	HEK293T, NIH-3T3	Protein degradation/Moderately enhancement of transactivation	Tumor suppressor	Calnan et al. (2012)
STAT3	K140	A	V	V	T	E	K	Q	Q	M	L	E	DLD1(Colon cancer)	Partial repression of transactivation	Tumor suppressor	Yang et al. (2010)
SOX2	K42	S	P	D	R	V	K	R	P	M	N	A	PA-1 (Ovarian teratocarcinoma)	Protein degradation	Tumor suppressor	Zhang et al, (2018)
	K117	P	R	R	K	T	K	T	L	M	K	K				
pRb	K810	Y	I	S	P	L	K	S	P	Y	K	I	Hela (Cervical cancer), CC42 (Mouse B cell hybridoma), C2C12 (Mouse myoblast), U2OS and SAOS2 (Osteosarcoma)	Protein stabilization	Tumor suppressor	Carr et al. (2011)
	K873	P	P	K	P	L	K	K	L	R	F	D				Munro et al. (2010)

Substrate	Methylation Site	Sequence Around Methylation Site											Study Model	Consequence	Role of SETD7	References
		−5	−4	−3	−2	−1	0	1	2	3	4	5				
Enzymes																
SUV39H1	K105	R	H	H	R	S	K	T	P	R	H	L	MEFs (Mouse embryonic fibroblast), H1299 (NSCLC)	Inhibition of enzyme activity	Tumor suppressor	Wang et al. (2013)
	K123	L	V	Q	K	A	K	Q	R	R	A	L				
ARTD1/ PARP1	K508	L	S	K	K	S	K	G	Q	V	K	E	U2OS (Osteosarcoma), MEFs (Mouse embryonic fibroblast)	Facilitation of DNA repair	Oncoprotein	Kassner et al. (2013)
RIOK1	K411	A	S	Q	R	T	K	E	E	R	S	S	Colorectal and gastric cancers	Protein degradation	Tumor suppressor	Hong et al., 2018
SIRT1	K233	L	S	E	P	P	K	R	K	K	R	K	HEK293T, HCT116 (Colorectal cancer)	Interaction with p53	Tumor suppressor	Liu et al., 2010
	K235	E	P	P	K	R	K	K	R	K	D	I				
	K236	P	P	K	R	K	K	R	K	D	I	N				
	K238	K	R	K	K	R	K	D	I	N	T	I				
PCAF	K89	S	A	P	R	A	K	K	L	E	K	L	HEK293, U2OS (Osteosarcoma)	Nuclear localization	Controversial	Masatsugu and Yamamoto, (2009)
DNMT	K142	T	P	R	R	S	K	S	D	G	E	A	Breast cancer	Protein degradation	Tumor suppressor	Esteve et al. (2009)

protein stability, which is beneficial for increasing of collagen contractility, as well as wound healing in renal fibroblast (Shuttleworth et al., 2018). The effects of SETD7 on various SMAD proteins thus seem to be diverse, which is determined by unique sequence and structure of each protein, or interacting proteins in the same protein family.

Signal Transducer and Activator of Transcription 3

STATs, a family of cytoplasmic transcription factors shared an overall general structure, are responsible for responding to cytokine stimulation (Bose et al., 2020). Among them, STAT3 is involved in numerous biological processes, including cell proliferation, survival, differentiation, and angiogenesis (Xin et al., 2020). STAT3 is hyperactivated in most human cancers, such as prostate, breast, and ovarian cancer (Yu et al., 2014), and is generally associated with a poor clinical prognosis (Zou et al., 2020). Previous study showed that tyrosine phosphorylation is an essential event for K140 methylation on STAT3 by SETD7. Moreover, STAT3 activity and its target gene expression are partially repressed by SETD7-mediated STAT3 methylation when IL-6 stimulation (Yang et al., 2010). In this case, SETD7 seems to serve as both an inflammatory and tumor suppressor.

SETD7 exhibits it impact as either an oncogenic protein or a tumor suppressor (**Figure 1**), thus, SETD7 is emerging as a therapeutic target in YAP-, ER-, AR-, and GLI3-mediated tumorgenesis. Recently, (*R*)-PFI-2 was identified as a first-in-class, potent (Ki^{app} = 0.33 nM), selective, and cell-active inhibitor of the methyltransferase activity of human SETD7 (Barsyte-Lovejoy et al., 2014). (*R*)-PFI-2 exhibits an unusual cofactor-dependent and substrate-competitive inhibitory mechanism by occupying the substrate peptide binding pocket of SETD7, including the catalytic lysine-binding channel, and by making direct contact with AdoMet (Lenstra et al., 2018). (*R*)-PFI-2 showed its activity in breast cancer cell MCF7 and mouse embryonic fibroblast in YAP-related studies (Barsyte-Lovejoy et al., 2014) and thus might be a potential therapeutic option for SETD7-mediated disease progression.

3 PERSPECTIVES AND CONCLUDING REMARKS

Methylation events serve to modulate and fine tune various cellular processes and signaling pathways (Han et al., 2019). As we have outlined in this review, SETD7-mediated methylation of transcription-related factors and enzymes (**Table 1**) has wide-reaching effects in different cell types and contexts. For example, SETD7 may act as either an oncogene or tumor suppressor. Meanwhile, SETD7-mediated methylation at different lysine residues within the same protein can even lead to divergent outcomes in different cancer cells and contexts (Batista and Helguero, 2018). Given the potential implications of intervening on SETD7-mediated methylation in disease contexts, namely cancer, researchers are keen to discover novel SETD7 substrates. Currently, researchers can use online prediction software of putative methylation sites combined with the SETD7 consensus methylation sequence to explore uncharacterized candidate proteins or isoforms of known SETD7 substrates, as exemplified for HIF-1α and HIF-2α (Liu et al., 2015).

Gene expression depends on not only activity of transcription factors, but also heterochromatin status which is regulated by some epigenetic modifiers. Besides SIRT1 we mentioned before, SETD7-mediated methylation has significance in regulation of such modifiers, such as suppressor of variegation 3-9 homolog 1 (SUV39H1) (Wang et al., 2013), p300/CBP-associated factor (PCAF) (Masatsugu and Yamamoto, 2009), ADP-ribosyltransferase diphtheria toxin-like 1 (ARTD1/PARP1) (Kassner et al., 2013) and DNA methyltransferase (DNMT) (Esteve et al., 2009). A multi-layered and -dimension regulatory network of SETD7 reveals the complexity and diversity of genetic modulation in the nucleus.

Data thus far, however, suggest that SETD7 exhibits a suppressive pattern in breast cancer, having a negative correlation with DNMT and E2F1 expression (Montenegro et al., 2016). In addition, low SETD7 expression correlates with a poor prognosis and lower survival rate in patients with gastric cancer (Akiyama et al., 2016), colorectal cancer (Zhang S. L. et al., 2020) and glioma metastasis (Li C. et al., 2020). On the other hand, data from a cohort study showed a positive correlation between SETD7 expression and the staging of cancer progression, which also seems to serve as a serum biomarker in colorectal cancer (Duan et al., 2018). SETD7 is also reported to have an oncogenic character in hepatoma cellular carcinoma, being progressively upregulated according to cancer stage (Gu et al., 2018). Interestingly, strong nuclear staining of SETD7 in high grade patients suggests that its subcellular localization is a significant indicator in the development and progression of prostate cancer (Gaughan et al., 2011), which might associate with the role of AR in nucleus. Due to the complexity of clinical specimens and differences among individuals, verifying the role of SETD7 and its corresponding substrates is unlikely in most cancer types. For this reason, the study of SETD7 is still largely confined to cellular based research or studies conducted in animal models.

Remarkably, SETD7 also acts as a tumor suppressor in certain contexts and indeed is downregulated in some cancers; thus, a method by which to elevate SETD7 expression and increase its activity is also warranted. Berberine, an anticancer agent, is a major botanical alkaloid that can be isolated from the root of *Rhizoma coptidis* (Huanglian) (Khan et al., 2022). Berberine can modulate various methylation- and acetylation-related enzymes that upregulate SETD7 expression in human multiple myeloma U266 cells in a dose-dependent manner (Wang et al., 2016). Moreover, SETD7 upregulation by berberine promotes RelA methylation and suppresses RelA-dependent transactivation of miR-21 in U266 cells (Hu et al., 2013). Unfortunately, due to the multi-bioactivity of berberine, a more specific agonist or inducer needs to be discovered in order to avoid off-target effects.

Going forward, further studies into the physiological and pathological effects of SETD7 are warranted to help develop novel diagnostic, prognostic, and/or therapeutic approaches in the cancer contexts. Although not discussed in this review, SETD7 is also a potential target to ameliorate diabetes,

inflammatory diseases, and aging-associated disorders (Batista and Helguero, 2018). However, as more and more substrates of SETD7 are discovered, researchers have to consider the effects from known substrate of SETD7 in their models when they find a novel candidate of SETD7, which might be as a reason leading to the decreasing of SETD7-related articles in recent years. According to tissue or cellular specificity, SETD7-related studies tend to investigate multi-substrate interactions in the same model to determine the ultimate effects of increasing or reducing various factors. Improving our fundamental knowledge on the mechanism of SETD7-mediated regulation of its substrates will be extremely informative to define tissue and cellular characteristics that are beneficial for SETD7-associated therapies.

AUTHOR CONTRIBUTIONS

CYC and DZ conceptualized this review. CYC wrote the article. DZ, YZ, HY, LZZ, CLC and CC commented on and revised the article. All authors reviewed the manuscript and approved the final version.

FUNDING

This study was supported by grants from the Natural Science Foundation of Guangdong Province (2021A1515011154, 2019A1515010210, 2021A1515011046), and the Shenzhen Municipal Government of China (JCYJ20210324093408024, JCYJ20180507182427559). Guangdong Provincial Science and Technology Program (No. 2017B030301016). Shenzhen Key Medical Discipline Construction Fund (No. SZXK060).

ACKNOWLEDGMENTS

The authors would like to thank Jessica Tamanini for editing the manuscript prior to submission, and Chengli Weng for preparing the tables and bibliography.

REFERENCES

Akiyama, Y., Koda, Y., Byeon, S.-j., Shimada, S., Nishikawaji, T., Sakamoto, A., et al. (2016). Reduced Expression of SET7/9, a Histone Mono-Methyltransferase, Is Associated with Gastric Cancer Progression. *Oncotarget* 7, 3966–3983. doi:10.18632/oncotarget.6681

Anestis, A., Zoi, I., Papavassiliou, A. G., and Karamouzis, M. V. (2020). Androgen Receptor in Breast Cancer-Clinical and Preclinical Research Insights. *Molecules* 25, 358. doi:10.3390/molecules25020358

Barsyte-Lovejoy, D., Li, F., Oudhoff, M. J., Tatlock, J. H., Dong, A., Zeng, H., et al. (2014). (R)-PFI-2 Is a Potent and Selective Inhibitor of SETD7 Methyltransferase Activity in Cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12853–12858. doi:10.1073/pnas.1407358111

Batista, I. d. A. A., and Helguero, L. A. (2018). Biological Processes and Signal Transduction Pathways Regulated by the Protein Methyltransferase SETD7 and Their Significance in Cancer. *Sig Transduct. Target Ther.* 3, 19. doi:10.1038/s41392-018-0017-6

Berkel, C., and Cacan, E. (2021). Estrogen- and Estrogen Receptor (ER)-mediated Cisplatin Chemoresistance in Cancer. *Life Sci.* 286, 120029. doi:10.1016/j.lfs.2021.120029

Bivona, T. G., Quatela, S. E., Bodemann, B. O., Ahearn, I. M., Soskis, M. J., Mor, A., et al. (2006). PKC Regulates a Farnesyl-Electrostatic Switch on K-Ras that Promotes its Association with Bcl-XL on Mitochondria and Induces Apoptosis. *Mol. Cell.* 21, 481–493. doi:10.1016/j.molcel.2006.01.012

Blagih, J., Buck, M. D., and Vousden, K. H. (2020). p53, Cancer and the Immune Response. *J. Cell. Sci.* 133. doi:10.1242/jcs.237453

Bose, S., Banerjee, S., Mondal, A., Chakraborty, U., Pumarol, J., Croley, C. R., et al. (2020). Targeting the JAK/STAT Signaling Pathway Using Phytocompounds for Cancer Prevention and Therapy. *Cells* 9, 1451. doi:10.3390/cells9061451

Bullock, M. D., Bruce, A., Sreekumar, R., Curtis, N., Cheung, T., Reading, I., et al. (2013). FOXO3 Expression during Colorectal Cancer Progression: Biomarker Potential Reflects a Tumour Suppressor Role. *Br. J. Cancer* 109, 387–394. doi:10.1038/bjc.2013.355

Burstein, H. J. (2020). Systemic Therapy for Estrogen Receptor-Positive, HER2-Negative Breast Cancer. *N. Engl. J. Med.* 383, 2557–2570. doi:10.1056/nejmra1307118

Calnan, D. R., Webb, A. E., White, J. L., Stowe, T. R., Goswami, T., Shi, X., et al. (2012). Methylation by Set9 Modulates FoxO3 Stability and Transcriptional Activity. *Aging* 4, 462–479. doi:10.18632/aging.100471

Campaner, S., Spreafico, F., Burgold, T., Doni, M., Rosato, U., Amati, B., et al. (2011). The Methyltransferase Set7/9 (Setd7) Is Dispensable for the P53-Mediated DNA Damage Response *In Vivo*. *Mol. Cell.* 43, 681–688. doi:10.1016/j.molcel.2011.08.007

Carr, S. M., Munro, S., Kessler, B., Oppermann, U., and La Thangue, N. B. (2011). Interplay between Lysine Methylation and Cdk Phosphorylation in Growth Control by the Retinoblastoma Protein. *EMBO J.* 30, 317–327. doi:10.1038/emboj.2010.311

Chen, Y., Zhou, F., Liu, H., Li, J., Che, H., Shen, J., et al. (2021). SIRT1, a Promising Regulator of Bone Homeostasis. *Life Sci.* 269, 119041. doi:10.1016/j.lfs.2021.119041

Cheng, X., Xu, X., Chen, D., Zhao, F., and Wang, W. (2019). Therapeutic Potential of Targeting the Wnt/β-Catenin Signaling Pathway in Colorectal Cancer. *Biomed. Pharmacother.* 110, 473–481. doi:10.1016/j.biopha.2018.11.082

Chokpaisarn, J., Urao, N., Voravuthikunchai, S. P., and Koh, T. J. (2017). Quercus Infectoria Inhibits Set7/NF-Kb Inflammatory Pathway in Macrophages Exposed to a Diabetic Environment. *Cytokine* 94, 29–36. doi:10.1016/j.cyto.2017.04.005

Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., et al. (2004). Regulation of P53 Activity through Lysine Methylation. *Nature* 432, 353–360. doi:10.1038/nature03117

Clevers, H., and Nusse, R. (2012). Wnt/β-Catenin Signaling and Disease. *Cell.* 149, 1192–1205. doi:10.1016/j.cell.2012.05.012

Colak, S., and Ten Dijke, P. (2017). Targeting TGF-β Signaling in Cancer. *Trends Cancer* 3, 56–71. doi:10.1016/j.trecan.2016.11.008

Del Rizzo, P. A., and Trievel, R. C. (2011). Substrate and Product Specificities of SET Domain Methyltransferases. *Epigenetics* 6, 1059–1067. doi:10.4161/epi.6.9.16069

Derynck, R., and Zhang, Y. E. (2003). Smad-dependent and Smad-independent Pathways in TGF-β Family Signalling. *Nature* 425, 577–584. doi:10.1038/nature02006

Dolcet, X., Llobet, D., Pallares, J., and Matias-Guiu, X. (2005). NF-kB in Development and Progression of Human Cancer. *Virchows Arch.* 446, 475–482. doi:10.1007/s00428-005-1264-9

Duan, B., Bai, J., Qiu, J., Wang, J., Tong, C., Wang, X., et al. (2018). Histone-lysine N-Methyltransferase SETD7 Is a Potential Serum Biomarker for Colorectal Cancer Patients. *EBioMedicine* 37, 134–143. doi:10.1016/j.ebiom.2018.10.036

Ea, C.-K., and Baltimore, D. (2009). Regulation of NF-Kb Activity through Lysine Monomethylation of P65. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18972–18977. doi:10.1073/pnas.0910439106

Elkouris, M., Kontaki, H., Stavropoulos, A., Antonoglou, A., Nikolaou, K. C., Samiotaki, M., et al. (2016). SET9-Mediated Regulation of TGF-β Signaling Links Protein Methylation to Pulmonary Fibrosis. *Cell. Rep.* 15, 2733–2744. doi:10.1016/j.celrep.2016.05.051

Elzakra, N., and Kim, Y. (2021). HIF-1α Metabolic Pathways in Human Cancer. *Adv. Exp. Med. Biol.* 1280, 243–260. doi:10.1007/978-3-030-51652-9_17

Engeland, K. (2018). Cell Cycle Arrest through Indirect Transcriptional Repression by P53: I Have a DREAM. *Cell. Death Differ.* 25, 114–132. doi:10.1038/cdd.2017.172

Esteller, M. (2007). Cancer Epigenomics: DNA Methylomes and Histone-Modification Maps. *Nat. Rev. Genet.* 8, 286–298. doi:10.1038/nrg2005

Estève, P.-O., Chin, H. G., Benner, J., Feehery, G. R., Samaranayake, M., Horwitz, G. A., et al. (2009). Regulation of DNMT1 Stability through SET7-Mediated Lysine Methylation in Mammalian Cells. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5076–5081. doi:10.1073/pnas.0810362106

Fasano, C., Disciglio, V., Bertora, S., Lepore Signorile, M., and Simone, C. (2019). FOXO3a from the Nucleus to the Mitochondria: A Round Trip in Cellular Stress Response. *Cells* 8, 1110. doi:10.3390/cells8091110

Fick, R. J., Kroner, G. M., Nepal, B., Magnani, R., Horowitz, S., Houtz, R. L., et al. (2016). Sulfur-Oxygen Chalcogen Bonding Mediates AdoMet Recognition in the Lysine Methyltransferase SET7/9. *ACS Chem. Biol.* 11, 748–754. doi:10.1021/acschembio.5b00852

Fouad, S., Hauton, D., and D'angiolella, V. (2020). E2F1: Cause and Consequence of DNA Replication Stress. *Front. Mol. Biosci.* 7, 599332. doi:10.3389/fmolb.2020.599332

Fu, L., Wu, H., Cheng, S. Y., Gao, D., Zhang, L., and Zhao, Y. (2016). Set7 Mediated Gli3 Methylation Plays a Positive Role in the Activation of Sonic Hedgehog Pathway in Mammals. *Elife* 5, 15690. doi:10.7554/eLife.15690

Fujimaki, K., Ogihara, T., Morris, D. L., Oda, H., Iida, H., Fujitani, Y., et al. (2015). SET7/9 Enzyme Regulates Cytokine-Induced Expression of Inducible Nitric-Oxide Synthase through Methylation of Lysine 4 at Histone 3 in the Islet β Cell. *J. Biol. Chem.* 290, 16607–16618. doi:10.1074/jbc.m115.661777

Gaughan, L., Stockley, J., Wang, N., Mccracken, S. R. C., Treumann, A., Armstrong, K., et al. (2011). Regulation of the Androgen Receptor by SET9-Mediated Methylation. *Nucleic Acids Res.* 39, 1266–1279. doi:10.1093/nar/gkq861

Giacinti, C., and Giordano, A. (2006). RB and Cell Cycle Progression. *Oncogene* 25, 5220–5227. doi:10.1038/sj.onc.1209615

Gu, Y., Wang, X., Liu, H., Li, G., Yu, W., and Ma, Q. (2018). SET7/9 Promotes Hepatocellular Carcinoma Progression through Regulation of E2F1. *Oncol. Rep.* 40, 1863–1874. doi:10.3892/or.2018.6621

Han, D., Huang, M., Wang, T., Li, Z., Chen, Y., Liu, C., et al. (2019). Lysine Methylation of Transcription Factors in Cancer. *Cell. Death Dis.* 10, 290. doi:10.1038/s41419-019-1524-2

Hoesel, B., and Schmid, J. A. (2013). The Complexity of NF-Kb Signaling in Inflammation and Cancer. *Mol. Cancer* 12, 86. doi:10.1186/1476-4598-12-86

Hong, X., Huang, H., Qiu, X., Ding, Z., Feng, X., and Zhu, Y. (2018). Targeting Posttranslational Modifications of RIOK1 Inhibits the Progression of Colorectal and Gastric Cancers. *Elife* 7, e29511.

Hu, H.-y., Li, K.-p., Wang, X.-j., Liu, Y., Lu, Z.-g., Dong, R.-h., et al. (2013). Set9, NF-Kb, and microRNA-21 Mediate Berberine-Induced Apoptosis of Human Multiple Myeloma Cells. *Acta Pharmacol. Sin.* 34, 157–166. doi:10.1038/aps.2012.161

Hu, J., Cao, J., Topatana, W., Juengpanich, S., Li, S., Zhang, B., et al. (2021). Targeting Mutant P53 for Cancer Therapy: Direct and Indirect Strategies. *J. Hematol. Oncol.* 14, 157. doi:10.1186/s13045-021-01169-0

Kassner, I., Andersson, A., Fey, M., Tomas, M., Ferrando-May, E., and Hottiger, M. O. (2013). SET7/9-dependent Methylation of ARTD1 at K508 Stimulates Poly-ADP-Ribose Formation after Oxidative Stress. *Open Biol.* 3, 120173. doi:10.1098/rsob.120173

Katoh, M. (2019). Genomic Testing, Tumor Microenvironment and Targeted Therapy of Hedgehog-Related Human Cancers. *Clin. Sci. (Lond)* 133, 953–970. doi:10.1042/cs20180845

Keating, S., and El-Osta, A. (2013). Transcriptional Regulation by the Set7 Lysine Methyltransferase. *Epigenetics* 8, 361–372. doi:10.4161/epi.24234

Khan, S., Hussain, A., Attar, F., Bloukh, S. H., Edis, Z., Sharifi, M., et al. (2022). A Review of the Berberine Natural Polysaccharide Nanostructures as Potential Anticancer and Antibacterial Agents. *Biomed. Pharmacother.* 146, 112531. doi:10.1016/j.biopha.2021.112531

Kim, Y., Nam, H. J., Lee, J., Park, D. Y., Kim, C., Yu, Y. S., et al. (2016). Methylation-dependent Regulation of HIF-1α Stability Restricts Retinal and Tumour Angiogenesis. *Nat. Commun.* 7, 10347. doi:10.1038/ncomms10347

Kontaki, H., and Talianidis, I. (2010). Lysine Methylation Regulates E2F1-Induced Cell Death. *Mol. Cell.* 39, 152–160. doi:10.1016/j.molcel.2010.06.006

Koo, J. H., and Guan, K.-L. (2018). Interplay between YAP/TAZ and Metabolism. *Cell. Metab.* 28, 196–206. doi:10.1016/j.cmet.2018.07.010

Kouskouti, A., Scheer, I., Staub, A., Tora, L., and Talianidis, I. (2010). Gene-Specific Modulation of TAF10 Function by SET9-Mediated Methylation. *Mol. Cell.* 14, 175–182.

Kurash, J. K., Lei, H., Shen, Q., Marston, W. L., Granda, B. W., Fan, H., et al. (2008). Methylation of P53 by Set7/9 Mediates P53 Acetylation and Activity *In Vivo*. *Mol. Cell.* 29, 392–400. doi:10.1016/j.molcel.2007.12.025

Lehnertz, B., Rogalski, J. C., Schulze, F. M., Yi, L., Lin, S., Kast, J., et al. (2011). p53-dependent Transcription and Tumor Suppression Are Not Affected in Set7/9-Deficient Mice. *Mol. Cell.* 43, 673–680. doi:10.1016/j.molcel.2011.08.006

Lenstra, D. C., Damen, E., Leenders, R. G. G., Blaauw, R. H., Rutjes, F. P. J. T., Wegert, A., et al. (2018). Structure-Activity Relationship Studies on (R)-PFI-2 Analogues as Inhibitors of Histone Lysine Methyltransferase SETD7. *ChemMedChem* 13, 1405–1413. doi:10.1002/cmdc.201800242

Lezina, L., Aksenova, V., Ivanova, T., Purmessur, N., Antonov, A. V., Tentler, D., et al. (2014). KMTase Set7/9 Is a Critical Regulator of E2F1 Activity upon Genotoxic Stress. *Cell. Death Differ.* 21, 1889–1899. doi:10.1038/cdd.2014.108

Li, C., Feng, S. Y., and Chen, L. (2020). SET7/9 Promotes H3K4me3 at lncRNA DRAIC Promoter to Modulate Growth and Metastasis of Glioma. *Eur. Rev. Med. Pharmacol. Sci.* 24, 12241–12250. doi:10.26355/eurrev_202012_24016

Li, M., Ning, J., Wang, J., Yan, Q., Zhao, K., and Jia, X. (2021). SETD7 Regulates Chondrocyte Differentiation and Glycolysis via the Hippo Signaling Pathway and HIF-1α. *Int. J. Mol. Med.* 48, 210. doi:10.3892/ijmm.2021.5043

Li, M., Wang, Y., Liu, X., Zhang, Z., Wang, L., and Li, Y. (2020). miR-629 Targets FOXO3 to Promote Cell Apoptosis in Gastric Cancer. *Exp. Ther. Med.* 19, 294–300. doi:10.3892/etm.2019.8168

Liebl, M. C., and Hofmann, T. G. (2021). The Role of P53 Signaling in Colorectal Cancer. *Cancers (Basel)* 13, 2125. doi:10.3390/cancers13092125

Liu, X., Chen, Z., Xu, C., Leng, X., Cao, H., Ouyang, G., et al. (2015). Repression of Hypoxia-Inducible Factor α Signaling by Set7-Mediated Methylation. *Nucleic Acids Res.* 43, 5081–5098. doi:10.1093/nar/gkv379

Liu, X., Wang, D., Zhao, Y., Tu, B., Zheng, Z., Wang, L., et al. (2011). Methyltransferase Set7/9 Regulates P53 Activity by Interacting with Sirtuin 1 (SIRT1). *Proc. Natl. Acad. Sci. U.S.A.* 108, 1925–1930. doi:10.1073/pnas.1019619108

Mandigo, A. C., Tomlins, S. A., Kelly, W. K., and Knudsen, K. E. (2022). Relevance of pRB Loss in Human Malignancies. *Clin. Cancer Res.* 28, 255–264. doi:10.1158/1078-0432.ccr-21-1565

Masatsugu, T., and Yamamoto, K. (2009). Multiple Lysine Methylation of PCAF by Set9 Methyltransferase. *Biochem. Biophysical Res. Commun.* 381, 22–26. doi:10.1016/j.bbrc.2009.01.185

Matissek, S. J., and Elsawa, S. F. (2020). GLI3: a Mediator of Genetic Diseases, Development and Cancer. *Cell. Commun. Signal* 18, 54. doi:10.1186/s12964-020-00540-x

Montenegro, M. F., Sánchez-del-Campo, L., González-Guerrero, R., Martínez-Barba, E., Piñero-Madrona, A., Cabezas-Herrera, J., et al. (2016). Tumor Suppressor SET9 Guides the Epigenetic Plasticity of Breast Cancer Cells and Serves as an Early-Stage Biomarker for Predicting Metastasis. *Oncogene* 35, 6143–6152. doi:10.1038/onc.2016.154

Moroishi, T., Hansen, C. G., and Guan, K.-L. (2015). The Emerging Roles of YAP and TAZ in Cancer. *Nat. Rev. Cancer* 15, 73–79. doi:10.1038/nrc3876

Moya, I. M., and Halder, G. (2019). Hippo-YAP/TAZ Signalling in Organ Regeneration and Regenerative Medicine. *Nat. Rev. Mol. Cell. Biol.* 20, 211–226. doi:10.1038/s41580-018-0086-y

Munro, S., Khaire, N., Inche, A., Carr, S., and La Thangue, N. B. (2010). Lysine Methylation Regulates the pRb Tumour Suppressor Protein. *Oncogene* 29, 2357–2367. doi:10.1038/onc.2009.511

Naruse, I., Ueta, E., Sumino, Y., Ogawa, M., and Ishikiriyama, S. (2010). Birth Defects Caused by Mutations in humanGLI3and mouseGli3genes. *Congenit. Anom. (Kyoto)* 50, 1–7. doi:10.1111/j.1741-4520.2009.00266.x

Nguyen, C. D. K., and Yi, C. (2019). YAP/TAZ Signaling and Resistance to Cancer Therapy. *Trends Cancer* 5, 283–296. doi:10.1016/j.trecan.2019.02.010

Niewiadomski, P., Niedziolka, S. M., Markiewicz, L., Uspienski, T., Baran, B., and Chojnowska, K. (2019). Gli Proteins: Regulation in Development and Cancer. *Cells* 8, 147. doi:10.3390/cells8020147

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P., et al. (2002). Set9, a Novel Histone H3 Methyltransferase that Facilitates Transcription by Precluding Histone Tail Modifications Required for Heterochromatin Formation. *Genes. Dev.* 16, 479–489. doi:10.1101/gad.967202

Oudhoff, M. J., Braam, M. J. S., Freeman, S. A., Wong, D., Rattray, D. G., Wang, J., et al. (2016). SETD7 Controls Intestinal Regeneration and Tumorigenesis by Regulating Wnt/β-Catenin and Hippo/YAP Signaling. *Dev. Cell.* 37, 47–57. doi:10.1016/j.devcel.2016.03.002

Oudhoff, M. J., Freeman, S. A., Couzens, A. L., Antignano, F., Kuznetsova, E., Min, P. H., et al. (2013). Control of the Hippo Pathway by Set7-dependent Methylation of Yap. *Dev. Cell.* 26, 188–194. doi:10.1016/j.devcel.2013.05.025

Parrales, A., and Iwakuma, T. (2015). Targeting Oncogenic Mutant P53 for Cancer Therapy. *Front. Oncol.* 5, 288. doi:10.3389/fonc.2015.00288

Pek, J. W., Anand, A., and Kai, T. (2012). Tudor Domain Proteins in Development. *Development* 139, 2255–2266. doi:10.1242/dev.073304

Plastino, F., Pesce, N. A., and André, H. (2021). MicroRNAs and the HIF/VEGF axis in Ocular Neovascular Diseases. *Acta Ophthalmol.* 99, e1255–e1262. doi:10.1111/aos.14845

Sasai, N., Toriyama, M., and Kondo, T. (2019). Hedgehog Signal and Genetic Disorders. *Front. Genet.* 10, 1103. doi:10.3389/fgene.2019.01103

Satija, S., Kaur, H., Tambuwala, M. M., Sharma, P., Vyas, M., Khurana, N., et al. (2021). Hypoxia-Inducible Factor (HIF): Fuel for Cancer Progression. *Cmp* 14, 321–332. doi:10.2174/1874467214666210120154929

Semenza, G. L. (2003). Targeting HIF-1 for Cancer Therapy. *Nat. Rev. Cancer* 3, 721–732. doi:10.1038/nrc1187

Shafi, A. A., Yen, A. E., and Weigel, N. L. (2013). Androgen Receptors in Hormone-dependent and Castration-Resistant Prostate Cancer. *Pharmacol. Ther.* 140, 223–238. doi:10.1016/j.pharmthera.2013.07.003

Shen, C., Wang, D., Liu, X., Gu, B., Du, Y., Wei, F. Z., et al. (2015). SET7/9 Regulates Cancer Cell Proliferation by Influencing β-catenin Stability. *FASEB J.* 29, 4313–4323. doi:10.1096/fj.15-273540

Shuttleworth, V. G., Gaughan, L., Nawafa, L., Mooney, C. A., Cobb, S. L., Sheerin, N. S., et al. (2018). The Methyltransferase SET9 Regulates TGFB1 Activation of Renal Fibroblasts via Interaction with SMAD3. *J. Cell. Sci.* 131, jcs207761. doi:10.1242/jcs.207761

Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D. R., Collins, R. E., Sharma, D., et al. (2008). Regulation of Estrogen Receptor α by the SET7 Lysine Methyltransferase. *Mol. Cell.* 30, 336–347. doi:10.1016/j.molcel.2008.03.022

Tan, M. E., Li, J., Xu, H. E., Melcher, K., and Yong, E.-l. (2015). Androgen Receptor: Structure, Role in Prostate Cancer and Drug Discovery. *Acta Pharmacol. Sin.* 36, 3–23. doi:10.1038/aps.2014.18

Udayakumar, T., Shareef, M. M., Diaz, D. A., Ahmed, M. M., and Pollack, A. (2010). The E2F1/Rb and p53/MDM2 Pathways in DNA Repair and Apoptosis: Understanding the Crosstalk to Develop Novel Strategies for Prostate Cancer Radiotherapy. *Seminars Radiat. Oncol.* 20, 258–266. doi:10.1016/j.semradonc.2010.05.007

Usami, M., Kikuchi, S., Takada, K., Ono, M., Sugama, Y., Arihara, Y., et al. (2020). FOXO3a Activation by HDAC Class IIa Inhibition Induces Cell Cycle Arrest in Pancreatic Cancer Cells. *Pancreas* 49, 135–142. doi:10.1097/mpa.0000000000001462

Valenta, T., Hausmann, G., and Basler, K. (2012). The Many Faces and Functions of β-catenin. *EMBO J.* 31, 2714–2736. doi:10.1038/emboj.2012.150

Verhoeven, Y., Tilborghs, S., Jacobs, J., De Waele, J., Quatannens, D., Deben, C., et al. (2020). The Potential and Controversy of Targeting STAT Family Members in Cancer. *Seminars Cancer Biol.* 60, 41–56. doi:10.1016/j.semcancer.2019.10.002

Waddell, A. R., Huang, H., and Liao, D. (2021). CBP/p300: Critical Co-activators for Nuclear Steroid Hormone Receptors and Emerging Therapeutic Targets in Prostate and Breast Cancers. *Cancers (Basel)* 13, 2872. doi:10.3390/cancers13122872

Wang, C., Sargsyan, D., Zhang, C., Wu, R., Yang, Y., and Kong, A.-N. (2018). Transcriptomic Analysis of Histone Methyltransferase Setd7 Knockdown and Phenethyl Isothiocyanate in Human Prostate Cancer Cells. *Anticancer Res.* 38, 6069–6083. doi:10.21873/anticanres.12957

Wang, D., Zhou, J., Liu, X., Lu, D., Shen, C., Du, Y., et al. (2013). Methylation of SUV39H1 by SET7/9 Results in Heterochromatin Relaxation and Genome Instability. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5516–5521. doi:10.1073/pnas.1216596110

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., et al. (2001). Purification and Functional Characterization of a Histone H3-Lysine 4-specific Methyltransferase. *Mol. Cell.* 8, 1207–1217. doi:10.1016/s1097-2765(01)00405-1

Wang, Q., Wang, K., and Ye, M. (2017). Strategies for Large-Scale Analysis of Non-histone Protein Methylation by LC-MS/MS. *Analyst* 142, 3536–3548. doi:10.1039/c7an00954b

Wang, Z., Liu, Y., Xue, Y., Hu, H., Ye, J., Li, X., et al. (2016). Berberine Acts as a Putative Epigenetic Modulator by Affecting the Histone Code. *Toxicol. Vitro* 36, 10–17. doi:10.1016/j.tiv.2016.06.004

Wu, Q., Schapira, M., Arrowsmith, C. H., and Barsyte-Lovejoy, D. (2021). Protein Arginine Methylation: from Enigmatic Functions to Therapeutic Targeting. *Nat. Rev. Drug Discov.* 20, 509–530. doi:10.1038/s41573-021-00159-8

Xiaoshi, J., Maoquan, L., Jiwei, W., Jinqiu, N., and Ke, Z. (2021). SETD7 Mediates the Vascular Invasion in Articular Cartilage and Chondrocytes Apoptosis in Osteoarthriis. *FASEB J.* 35, e21283. doi:10.1096/fj.202000373RRRR

Xie, Q., Bai, Y., Wu, J., Sun, Y., Wang, Y., Zhang, Y., et al. (2011). Methylation-mediated Regulation of E2F1 in DNA Damage-Induced Cell Death. *J. Recept. Signal Transduct.* 31, 139–146. doi:10.3109/10799893.2011.552914

Xie, Q., Hao, Y., Tao, L., Peng, S., Rao, C., Chen, H., et al. (2012). Lysine Methylation of FOXO3 Regulates Oxidative Stress-induced Neuronal Cell Death. *EMBO Rep.* 13, 371–377. doi:10.1038/embor.2012.25

Xin, P., Xu, X., Deng, C., Liu, S., Wang, Y., Zhou, X., et al. (2020). The Role of JAK/STAT Signaling Pathway and its Inhibitors in Diseases. *Int. Immunopharmacol.* 80, 106210. doi:10.1016/j.intimp.2020.106210

Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., et al. (2010). Reversible Methylation of Promoter-Bound STAT3 by Histone-Modifying Enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21499–21504. doi:10.1073/pnas.1016147107

Yang, X.-D., Huang, B., Li, M., Lamb, A., Kelleher, N. L., and Chen, L.-F. (2009). Negative Regulation of NF-Kb Action by Set9-Mediated Lysine Methylation of the RelA Subunit. *EMBO J.* 28, 1055–1066. doi:10.1038/emboj.2009.55

Yin, Z., Wang, W., Qu, G., Wang, L., Wang, X., and Pan, Q. (2020). MiRNA-96-5p Impacts the Progression of Breast Cancer through Targeting FOXO3. *Thorac. Cancer* 11, 956–963. doi:10.1111/1759-7714.13348

Yousafzai, N. A., Jin, H., Ullah, M., and Wang, X. (2021). Recent Advances of SIRT1 and Implications in Chemotherapeutics Resistance in Cancer. *Am. J. Cancer Res.* 11, 5233–5248.

Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting STAT3 Signalling in Cancer: New and Unexpected Biological Functions. *Nat. Rev. Cancer* 14, 736–746. doi:10.1038/nrc3818

Zanconato, F., Battilana, G., Cordenonsi, M., and Piccolo, S. (2016a). YAP/TAZ as Therapeutic Targets in Cancer. *Curr. Opin. Pharmacol.* 29, 26–33. doi:10.1016/j.coph.2016.05.002

Zanconato, F., Cordenonsi, M., and Piccolo, S. (2016b). YAP/TAZ at the Roots of Cancer. *Cancer Cell.* 29, 783–803. doi:10.1016/j.ccell.2016.05.005

Zhang, C., Hoang, N., Leng, F., Saxena, L., Lee, L., and Alejo, S. (2018). LSD1 Demethylase and the Methyl-Binding Protein PHF20L1 Prevent SET7 Methyltransferase-Dependent Proteolysis of the Stem-Cell Protein SOX2. *J. Biol. Chem.* 293, 3663–3674.

Zhang, S. L., Du, X., Tan, L. N., Deng, F. H., Zhou, B. Y., Zhou, H. J., et al. (2020). SET7 Interacts with HDAC6 and Suppresses the Development of Colon Cancer through Inactivation of HDAC6. *Am. J. Transl. Res.* 12, 602–611.

Zhang, T., Wang, X.-F., Wang, Z.-C., Lou, D., Fang, Q.-Q., Hu, Y.-Y., et al. (2020). Current Potential Therapeutic Strategies Targeting the TGF-β/

Smad Signaling Pathway to Attenuate Keloid and Hypertrophic Scar Formation. *Biomed. Pharmacother.* 129, 110287. doi:10.1016/j.biopha.2020.110287

Zhang, Y., and Wang, X. (2020). Targeting the Wnt/β-Catenin Signaling Pathway in Cancer. *J. Hematol. Oncol.* 13, 165. doi:10.1186/s13045-020-00990-3

Zinatizadeh, M. R., Schock, B., Chalbatani, G. M., Zarandi, P. K., Jalali, S. A., and Miri, S. R. (2021). The Nuclear Factor Kappa B (NF-kB) Signaling in Cancer Development and Immune Diseases. *Genes. & Dis.* 8, 287–297. doi:10.1016/j.gendis.2020.06.005

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

Zou, Y., Tsai, W.-B., Cheng, C.-J., Hsu, C., Chung, Y. M., Li, P.-C., et al. (2008). Forkhead Box Transcription Factor FOXO3a Suppresses Estrogen-dependent Breast Cancer Cell Proliferation and Tumorigenesis. *Breast Cancer Res.* 10, R21. doi:10.1186/bcr1872

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

TYPE Mini Review
PUBLISHED 09 August 2022
DOI 10.3389/fgene.2022.934223

OPEN ACCESS

EDITED BY
Yan-Ming Xu,
Shantou University, China

REVIEWED BY
Xuyu Gu,
Southeast University, China

*CORRESPONDENCE
Lizhi Zhu,
lzzhu86@pku.edu.cn
Duo Zheng,
dzheng@szu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 02 May 2022
ACCEPTED 11 July 2022
PUBLISHED 09 August 2022

CITATION
Yang H, Chiang C, Luo Q, Chen C, Huang J, Zhu L and Zheng D (2022), YT521-B homology domain family proteins as N6-methyladenosine readers in tumors.
Front. Genet. 13:934223.
doi: 10.3389/fgene.2022.934223

YT521-B homology domain family proteins as N6-methyladenosine readers in tumors

Heng Yang[1†], Chengyao Chiang[1,2†], Qinhong Luo[1†], Chunlan Chen[1], Junrong Huang[1], Lizhi Zhu[1*] and Duo Zheng[1*]

[1]Guangdong Provincial Key Laboratory of Genome Stability and Disease Prevention, Shenzhen University International Cancer Center, Department of Cell Biology and Genetics, School of Medicine, Department of Pharmacy, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital (Shenzhen Institute of Translational Medicine), Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Shenzhen, China, [2]Central Laboratory, Southern University of Science and Technology, Yantain Hospital, Shenzhen, China

N6-methyladenosine (m6A) is the most abundant internal chemical modification of eukaryotic mRNA and plays diverse roles in gene regulation. The m6A modification plays a significant role in numerous cancer types, including kidney, stomach, lung, bladder tumors, and melanoma, through varied mechanisms. As direct m6A readers, the YT521-B homology domain family proteins (YTHDFs) play a key role in tumor transcription, translation, protein synthesis, tumor stemness, epithelial–mesenchymal transition (EMT), immune escape, and chemotherapy resistance. An in-depth understanding of the molecular mechanism of YTHDFs is expected to provide new strategies for tumor treatment. In this review, we provide a systematic description of YTHDF protein structure and its function in tumor progression.

KEYWORDS

M6A, YTHDFs, tumor, immune escape, EMT, chemotherapy resistance

Introduction

N6-methyladenosine (m6A) is formed by methylation of adenosine at the N6 position. This plentiful posttranscriptional RNA modification extensively regulates multiple aspects of gene expression (He et al., 2019). It exists widely in eukaryotic mRNA and lncRNA and is involved in cell development, stem cell characteristic maintenance, tumor progression, sperm receiving ability, and T cell homeostasis (Dominissini et al., 2012; Maity et al., 2016; Cui et al., 2017; Ianniello et al., 2018). The dynamic m6A RNA modification is catalyzed by the methyltransferase complex (m6A "writer"), demethylases (m6A "eraser"), and m6A-binding proteins (m6A "readers") (Helm et al., 2017; Meyer et al., 2017). M6A readers, including YT521-B homology (YTH)–containing domain family proteins, insulin-like growth factor binding protein 2, eukaryotic initiation factor, and heterogeneous nuclear ribonucleoproteins (Gilbert et al., 2016), recognize and bind RNA methylation modifications to mediate the translation and degradation of

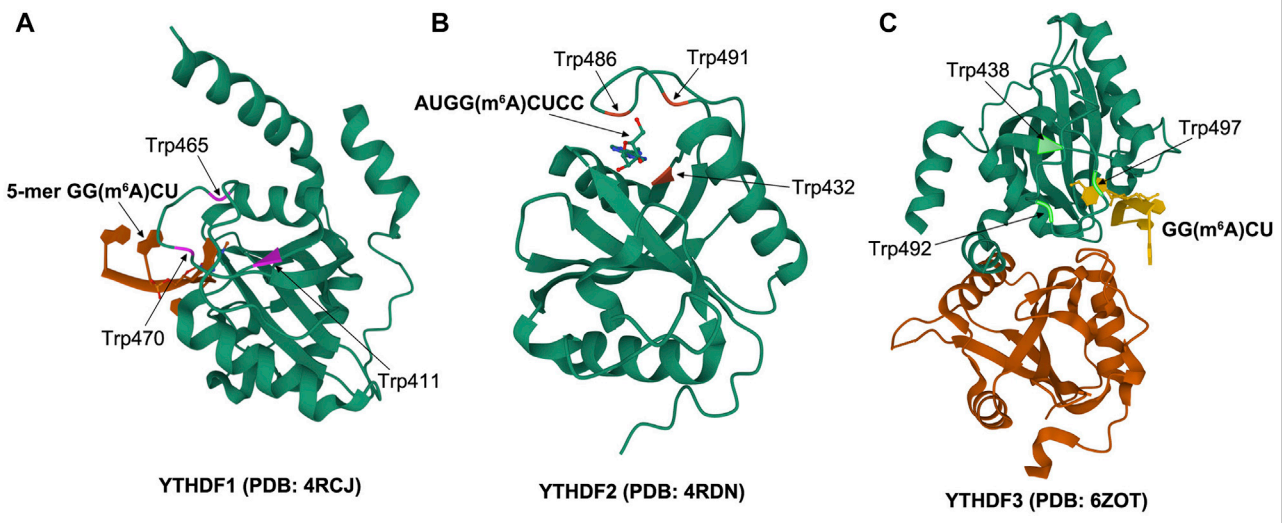


FIGURE 1
Schematic of crystal structure: **(A)** structure of YTHDF1 YTH domain in complex with 5-mer GG (m6A)CU (PDB id: 4RCJ), **(B)** structure of YTHDF2 YTH domain with AUGG (m6A)CUCC (PDB id: 4RDN), **(C)** structure of YTHDF3 YTH domain in complex with GG (m6A)CU (PDB id: 6ZOT).

downstream RNA (Wu et al., 2018). Accumulating evidence supports the influence of the m6A modification on lung cancer, melanoma, gastric cancer, bladder cancer, and liver cancer through various mechanisms (Dai et al., 2018; Burbage et al., 2019; Chen X. et al., 2019; Li et al., 2019). In this review, we focus on the current status of our understanding of m6A readers. We describe the functions of YT521-B homologous domain family proteins (YTHDFs) in tumorigenesis and cancer progression. At last, we depict future research directions to fully characterize YTHDFs.

Biological characteristics of YTHDFs and m6A modification

As the most common m6A readers, YTH domain–containing proteins include the YTHDF and YTHDC subtypes, which specifically recognize m6A *via* an evolutionarily conserved aromatic cage in the YTH domain (Xu et al., 2014; Shi et al., 2017; Liao et al., 2018). Another two proteins, FMRP translational regulator 1, associated with embryonic development, and leucine-rich pentatricopeptide repeat containing, associated with human tumors, can also recognize m6A modification sites (Arguello et al., 2017; Edupuganti et al., 2017). As the main YTHDF proteins, YTHDF1, YTHDF2, and YTHDF3 were first discovered in the cytoplasm (Wang et al., 2014). YTHDFs contain two domains: a YTH domain at the C-terminal and the unstructured N-terminal of most remaining proteins, comprising a large and low-complexity domain that contains mainly Gln, Asn, and Pro residues (Stoilov et al., 2002). YTHDFs are prone to forming liquid–liquid phase separation and protein droplets because of the presence of this type of large low-complexity domain, especially those with RNA binding domains (Aguzzi et al., 2016; Ries et al., 2019). In fact, proteomic analysis revealed that each YTHDF exists in the form of droplet structures, involving stress particles and artificially created RNA particles (Kato et al., 2012; Jain et al., 2016). The m6A binding pocket of the YTHDF1 YTH domain is made up of two α helices (α1 and α2) and four β strands (β1, β2, β4, and β5). To be exact, m6A is properly matched in a pocket composed of Trp411, Trp465, and Trp470, parallel to the loop planes of Trp411 and Trp470 and perpendicular to the loop plane of Trp465 (Xu et al., 2015) (Figure 1A). The residues Lys416 and Arg527 of the YTHDF2 YTH domain play a critical role in binding RNA backbone and hydrophobic vesicles. M6A single nucleotide is located in the aromatic cage composed of three aromatic amino acids Trp432, Trp486, and Trp491, forming π–π interaction. These types of particular binding contribute to the specific recognition of m6A (Ma et al., 2019) (Figure 1B). YTHDF3 possesses 86% sequence identity with the YTH structural domain of both YTHDF1 and YTHDF2. Furthermore, X-ray structural analyses and molecular dynamics simulations demonstrated that all YTHDF YTH domains share essentially equal interactions with the m6A-modified RNA and have an analogous affinity (Figure 1C), suggesting the redundant characters of the three proteins in physiological functions (Li et al., 2020). Different binding proteins selectively recognize m6A-modified RNA and exert a crucial role in regulating target gene expression. YTHDF1 increases the transmission of mRNA transcription complex by binding to m6A at a site around the stop codon and coordinates translation initiation and protein synthesis

TABLE 1 YT521-B homology domain family proteins in tumorigenesis.

Reader	Trend	Tumor	Regulatory pathway/Gene	References
YTHDF1	↑	HCC	Cell cycles	Zhao et al. (2018)
	↑	GC	Wnt/β-catenin, FZD7	Pi et al. (2021)
	↑	CRC	RhoA Signaling, ARHGEF2	Wang et al. (2022)
	↑	NSCLC	CDK2, CDK4, cyclin D1	Shi et al. (2019)
		TNBC	ITGA6	Jin D. et al. (2019)
	↑	GBM		Xu et al. (2020)
YTHDF2	↑	HCC	OCT4, SOCS2	Zhang et al. (2020); Chen et al. (2018)
	↑	GC	Cell cycle, apoptosis	Zhang et al. (2017)
	↑	NSCLC	6-PGD	Sheng et al. (2020)
	↑	BCa	SETD7, KLF4	Zheng et al. (2021); Xie et al. (2020)
	↑	GBM	LXRα, HIVEP2	Fang et al. (2021)
YTHDF3	↑	CRC	GAS5, YAP	Ni et al. (2019)
	↑	TNBC	ZEB1, EMT	Lin et al. (2022)

GBM, glioma; BCa, bladder cancer.

(Wang et al., 2015). YTHDF2 was the first binding protein recognized to regulate mRNA stability and remodeling of RNA structure through m6A modification (Ivanova et al., 2017). YTHDF3 is usually coupled with YTHDF1 and YTHDF2, which control the attenuation or translation of methylated RNA (Li et al., 2017). Although many studies have provided evidence of the importance of the m6A recognition protein in multiple physiological and pathological processes, further studies are still needed to fully clarify its role in mRNA biology.

YTHDFs and tumorigenesis

YTHDFs have been reported to elevate the growth and metastasis of various tumors, such as hepatic cell carcinoma (HCC), gastric cancer (GC), colorectal cancer (CRC), nonsmall cell lung cancer (NSCLC), and triple-negative breast cancer (TNBC) (Table 1). For example, YTHDF1 expression was dramatically upregulated in stage III and IV HCC patients compared with stage II disease, and the prognosis of patients with high YTHDF1 levels was worse (Zhao et al., 2018). Since the potential target genes regulated by YTHDF1 might be associated with amino acid degradation and lipid metabolism in the tumor cell cycle, abnormalities in these physiological functions could lead directly to the occurrence and progression of HCC. YTHDF2 also acts as an oncogene to promote the HCC lung metastasis and cancer stem cell (CSC) phenotype through increasing m6A methylation levels of OCT4 mRNA in the 5′-untranslated region (UTR) and OCT4 protein expression (Zhang et al., 2020). In addition, YTHDF2 attaches to m6A sites to boost mRNA degradation. When the m6A modification level in tumor suppressor gene SOCS2 mRNA was upregulated, the increased number of m6A binding sites in YTHDF2 ultimately promoted the degradation of SOCS2 mRNA, which was beneficial to HCC cell proliferation (Chen et al., 2018). These studies reveal that YTHDFs can accelerate HCC progression by promoting protein translation and regulating the stability of mRNA.

Huo et al. (2021) reported that YTHDF1 promoted upregulation of SPHK2 expression at the protein level *via* its specific structural domain, whereas SPHK2 coupled with KLF2 to induce phosphorylation modifications, thereby mediating KLF2 degradation and promoting the antiapoptotic function of GC cells. Compared with normal tissues, YTHDF2 was significantly upregulated in gastric cancer, Silencing of YTHDF2 suppressed gastric cancer cell multiplication, arresting the cell cycle in the G1/S phase and boosting cell apoptosis (Zhang et al., 2018).

In CRC, c-MYC initiated YTHDF1 expression, and high expression was associated with poor overall survival (OS) (Nishizawa et al., 2018). LncRNA-GAS5 was shown to directly bind to the WW domain of YAP to promote its phosphorylation and subsequently ubiquitin degradation, thus declining YTHDF3 transcription. YTHDF3 tended to selectively combine with GAS5 in methylated m6A site, thereby initiating its decay and the formation of a GAS5-YAP-YTHDF3 negative feedback loop to accelerate CRC development (Ni et al., 2019). This study suggests that targeting the lncRNA GAS5-YAP-YTHDF3 axis is a promising approach for CRC treatment.

In a large-scale genome and transcriptome sequencing, Shi et al. (2019) found that YTHDF1 was amplified in NSCLC from Tibetan domestic mammals. Knockdown of YTHDF1 inhibited NSCLC cell growth and the formation of xenograft tumor by adjusting the translational efficiency of CDK2, CDK4, and cyclin D1 and also restrained *de novo* lung adenocarcinoma

progression. It is of interest that patients with high expression of YTHDF1 in high altitudes have a better clinical prognosis. This is because the knockdown of YTHDF1 activates the Keap1-Nrf2-AKR1C1 axis, which increases the sensitivity of NSCLC patients to cisplatin chemotherapy. The above results indicate that the specific recognition and regulation of YTHDF1 on m6A-modified mRNA may be associated with time and cell environment, which contributes to different functions of YTHDF1 in various tumors. Moreover, YTHDF2 was reported to be upregulated in lung cancer and shown to bind straightly to the m6A modification site in the 3′-UTR of the 6-phosphogluconate dehydrogenase (6-PGD) mRNA. This event promoted the translation of 6-PGD mRNA and enhanced the pentose phosphate pathway flux to facilitate the growth of lung cancer cells (Sheng et al., 2020).

It was found that m6A abundantly existed in ITGA6 transcripts, and methyltransferase-like 3 (METTL3) promoted YTHDF1/YTHDF3 binding to the ITGA6 mRNA 3′-UTR, thereby enhancing ITGA6 mRNA translation and changing the adhesion ability of bladder cancer cells (Laudato et al., 2017; Jin H. et al., 2019). In TNBC patients, YTHDF3 expression was correlated with poorer disease-free survival (DFS) and OS. YTHDF3 improved ZEB1 mRNA stability in an m6A-dependent manner and positively regulated cell migration, invasion, and EMT in TNBC cells (Lin et al., 2022). The above evidence demonstrates that YTHDF3 can not only function alone but also couple with METTL3 to elevate the translation of methylated RNA, playing a vital role in tumor progression.

It is worth noting that some studies have shown that YTHDF1 inhibits tumor progression. Knockdown of YTHDF1 could improve the crosspresentation ability of dendritic cells in mouse melanoma, and when combined with a PD-L1 checkpoint inhibitor, the tumor was almost completely controlled (Han et al., 2019). Another study showed that m6A modification dramatically inhibited ocular melanoma cell growth, and its low expression suggested a poor prognosis. YTHDF1 functions as a tumor suppressor of ocular melanoma by recognizing m6A-modified RNA and accelerating the translation of histidine trinucleotide binding protein 2 (HINT2) (Jia et al., 2019). Hence, as m6A readers, YTHDFs regulate tumor-specific oncogene expression and biological function *via* different mechanisms. Thus, targeting YTHDFs is implicated as a new strategy for cancer treatment.

YTHDFs and tumor immune responses

T cell immune responses play important roles in tumor immunotherapy. Although neoantigens in cancer patients are abundant, thorough removal of tumors is hindered by the failure to produce lasting antitumor immune responses (Yarchoan et al., 2017). Han et al. (2019) reported that YTHDF1 regulated neoantigen-specific immunity *via* m6A modification. YTHDF1 depletion in classical dendritic cells not only increased the mutual expression of tumor antigens but also enhanced $CD8^+$ T cells crosspriming. The transcripts of lysosomal proteases were labeled by m6A and recognized by YTHDF1, thus promoting the translation of lysosomal cathepsin in dendritic cells and significantly inhibiting crosspresentation by wild-type dendritic cells. This study also showed that T cells in YTHDF1-deficient mice produced high levels of IFN-γ, indicating that YTHDF1 knockout in host cells promotes T cell activation at an early stage. In addition, the therapeutic effect of a PD-L1 immune checkpoint inhibitor was heightened in $YTHDF1^{-/-}$ mice, with almost complete inhibition of tumor growth, suggesting the potential importance of YTHDF1 as a therapeutic target for antitumor immune therapy.

Lin et al. (2020) found that YTHDF2 expression was positively associated with the expression of multiple immune checkpoint biomarkers, including CTLA-4, TIM-3, and PD-1, as well as isocitrate dehydrogenase 1, and tumor-associated macrophage factors in lower-grade glioma. Further experiments are needed to confirm this bioinformatic conclusion. In addition, high YTHDF2 transcription was associated with a higher survival rate and increased tumor-infiltrating lymphocytes in clear cell renal cell carcinoma and NSCLC patients, suggesting the antitumor function of YTHDF2 (Su et al., 2021; Tsuchiya et al., 2021). It is of interest that there are two aspects to the regulatory role of YTHDF2 in tumor immunity. Wang et al. (2020) reported that YTHDF2 stabilized STAT1 and IRF1 mRNA in tumors deficient in METTL3 or METTL4, thus promoting the signal transduction of the IFN-γ-STAT1-IRF1 axis and enhancing the response of immunotherapy-resistant colorectal cancer to PD-1 treatment. Methylated YTHDF2 decreased the stability of mRNA transcribed from the melanoma-promoting genes PD-1, CXCR4, and SOX10, to inhibit the growth of melanoma (Yang et al., 2019). A natural killer (NK) cell is a type of lymphocyte that mediates antitumor immune response. Ma et al. (2021) showed that YTHDF2 was upregulated in NK cells to maintain the steady state and terminal maturation of NK cells after activation of the immune system. In a mechanistic manner, YTHDF2 mediated the release of IFN-γ, granzyme B, and perforin in NK cells by forming a STAT5-YTHDF2 positive feedback loop, which finally promoted TARDBP degradation and enhanced NK proliferation and survival. However, YTHDF2 was found to inhibit innate immunity by binding with m6A-modified circZKSCAN1 in mammalian cells, to inhibit RIG-I recognition and k63-PolyUb and promote tumor immune escape (Chen Y. G. et al., 2019). Therefore, YTHDF2 not only enhances immunity but also promotes tumor immune escape, which are functions that may depend on differences in tumor types and the mechanisms of tumorigenesis.

Furthermore, some studies have shown that YTHDF3 contributes to the formation of the tumor microenvironment and could be used as an immune-related marker. High YTHDF3 expression indicates poor survival with a diversity of lymphocyte infiltration in breast cancer, as well as esophageal squamous cell carcinoma and head and neck squamous cell carcinoma (Zhang et al., 2021; Zhao et al., 2021; Pu et al., 2022). The above research shows that YTHDF3 may be a potential therapeutic target related to tumor-infiltrating immune cells.

YTHDFs and EMT

EMT is the key process of tumor cell metastasis, with the deficit of E-cadherin, which is considered to be the most basic event in this process. As a key transcription factor in this process, Snail reduces the expression of E-cadherin, which promotes tumor recurrence, metastasis, and drug resistance (Jang et al., 2019). YTHDF1 promotes cancer progression by regulating EMT. Lin et al. (2019) found that Snail was methylated in the CDS region and that YTHDF1 was significantly enriched in the CDS region of Snail mRNA leading to increased TGF-β-induced snail expression in HeLa cells. In addition, YTHDF1 silencing significantly increased the expression levels of Claudin-1 and zonula occludens protein 1, whereas expression of matrix metalloproteinase-9, matrix metalloproteinase-2, Vimentin, and N-cadherin was inhibited. These findings indicated that YTHDF1-mediated EMT is responsible for promoting the migration and invasion of HCC cells (Luo et al., 2021).

Recent studies have confirmed the involvement of YTHDF2 in tumor EMT and its role in regulating tumor cell migration and invasion. The METTL3/YTHDF2 m6A axis regulates miR-1915-3p expression by inhibiting the activity of the transcription factor KLF4, thereby targeting SET and inhibiting NSCLC cell migration, invasion, and EMT (Pan et al., 2021). Chen et al. (2017b) reported that YTHDF2 orchestrated two cellular processes in pancreatic cancer cells. YTHDF2 promoted pancreatic cancer cell proliferation and inhibited migration and invasion, a phenomenon termed "migration-proliferation dichotomy," and also repressed EMT, probably by negatively regulating the YAP signaling pathway (Supplementary Figure S1). Therefore, YTHDF2 should be seriously considered as a therapeutic target for pancreatic cancer. Other studies showed that YTHDF3 promotes tumor migration and invasion by regulating EMT. Knockdown of YTHDF3 increased E-cadherin expression and reduced the expression of N-cadherin and Vimentin, which finally inhibited the migration and invasion of TNBC cells (Lin et al., 2022). Wang M. et al. (2020) reported that YTHDF3 improved ZEB1 mRNA stability in an m6A-dependent manner, maintained ZEB1 mRNA stability, and finally, promoted the translation and expression of ZEB1.

YTHDFs and drug resistance in tumors

Abnormal AKR1C1 expression induces the development of drug resistance in chemotherapy of cancers including NSCLC, breast cancer, and ovarian cancer (Chen et al., 2017a). YTHDF1 knockout was associated with decreased Keap1 expression, whereas the expression of Nrf2 and its downstream factor AKR1C1 was increased. Furthermore, YTHDF1 knockout reduced the translation efficiency of m6A-modified Keap1 transcripts, resulting in activation of the antioxidant oxygen radical scavenging system (Nrf2-AKR1C1) and resistance of NSCLC to cisplatin chemotherapy, leading to a worse clinical prognosis. This study suggested that the Keap1-Nrf2-AKR1C1 axis was the downstream regulatory target of YTHDF1, emphasizing the important character of YTHDF1 in hypoxia adjustment and NSCLC progression (Shi et al., 2019). In addition, YTHDF1 was recruited to the m6A modification site of TRIM29, promoting TRIM29 translation in cisplatin-resistant ovarian cancer cells. Furthermore, YTHDF1 knockdown inhibited cisplatin-resistant ovarian cancer cells with CSC-like characteristics (Hao et al., 2021), suggesting that YTHDF1 is a promising cancer marker to trace the recurrence of ovarian cancer.

Studies have shown that temozolomide (TMZ) resistance was acquired in recurrent glioblastoma multiforme (GBM) tumors after radiotherapy and chemotherapy. After 60 Gy radiation, YTHDF2 synthesis in GBM cells was significantly upregulated (Pak et al., 2021), suggesting that YTHDF2 plays an important role in TMZ resistance to recurrent glioblastoma. Huang et al. (2019) showed that TUSC7, a strong inhibitory lnc-RNA, was expressed at low levels in NSCLC tissues and inhibited the nondivision stem cells by obstructing Notch signaling. YTHDF2 promoted the renewal of the stem cell–like population in erlotinib-resistant cells by inhibiting TUSC7 and promoted the characteristics of tumor resistance (Li et al., 2021). Furthermore, YTHDF1/3 increased the drug resistance of tumors by promoting the translation of drug resistance–related proteins. Jin et al. (2019) found that METTL3 improved the m6A modification levels of YAP and further accelerated YAP translation by engaging YTHDF1/3 and the eEIF3a translation initiation complex, thereby inducing the NSCLC drug resistance and metastasis. Therefore, there is strong evidence that YTHDFs play an important regulatory role in the responses of tumors to chemotherapy and represent a promising therapeutic target for drug-resistant cancer.

Conclusion and perspective

Based on the current research, YTHDFs mainly recognize m6A-modified target genes, increase translation through

triggering translational initiation and elongation, and then control mRNA stability and target gene expression. YTHDFs affect tumorigenesis, metastasis, tumor immunity, EMT, and chemoresistance by regulating target gene expression. Many studies have shown the various functions of YTHDFs and further revealed the crosstalk and competition between different YTHDFs. Here, we summarized the current status of the structure and biological characteristics of m6A readers—YTHDFs—in human cancers. However, a full and comprehensive understanding of YTHDFs remains a distant prospect for several reasons. First, YTHDFs exert either oncogenic or tumor-suppressive effects that may depend on the type of regulated targets or interactome in specific cancer species. Second, the biological significance of liquid–liquid separation induced by the low-complexity domain of YTHDFs in tumors is still unknown. Third, we need to dissect the interactions among YTHDFs as well as those between YTHDFs and erasers and readers in tumorigenesis, metastasis, and immunity of human cancers. Fourth, we need to determine how the expression and activity of YTHDFs are regulated in cancer and whether the recognition of m6A by YTHDFs is RNA sequence specific. At last, it is unclear whether YTHDFs can regulate the occurrence and development of cancer by recognizing m6A-modified noncoding RNA, such as microRNA or lncRNA. The functions of YTHDF mutants and isoforms in tumor molecular diagnosis also represent an important topic for future research.

Several groups have reported that targeting m6A-modifying enzymes and regulatory proteins, such as FTO (Su et al., 2020), METTL3 (Yankova et al., 2021), and ALKBH5 (Fang et al., 2022), is beneficial to the treatment of malignant leukemia and solid tumors. We also found a covalent small molecule drug, which could activate the tumor immune microenvironment and play an antihepatoma role by targeting the m6A recognition protein IGF2BP1 (Liu et al., 2022). Thus, the targeted regulation of YTHDFs by small chemical molecules may be a promising therapeutic strategy for human cancer. Overall, more basic and clinical studies are required to fully elucidate the mechanism underlying the functions of YTHDFs, which will provide new hope for cancer diagnosis and targeted treatment.

Author contributions

DZ and LZ contributed equally to the design and coordination of the study. HY, CYC, and QL collected the data with help from the CLC, and HY wrote the manuscript. All authors reviewed and edited the manuscript.

Funding

This study was supported by grants from the Natural Science Foundation of Guangdong Province (2021A1515011154, 2019A1515010210, 2021A1515011046, 2021A1515010996), Regional Joint Fund of Guangdong Province (2019B1515120080), the Shenzhen Municipal Government of China (JCYJ20210324093408024, JCYJ20180507182427559), and Shenzhen Key Medical Discipline Construction Fund (No. SZXK060). The authors would like to thank Jessica Tamanini for editing the manuscript prior to submission.

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.2022.934223/full#supplementary-material

References

Aguzzi, A., and Altmeyer, M. (2016). Phase separation: Linking cellular compartmentalization to disease. *Trends Cell. Biol.* 26 (7), 547–558. doi:10.1016/j.tcb.2016.03.004

Arguello, A. E., DeLiberto, A. N., and Kleiner, R. E. (2017). RNA chemical proteomics reveals the N6-methyladenosine (m6A)-regulated protein-RNA interactome. *J. Am. Chem. Soc.* 139 (48), 17249–17252. doi:10.1021/jacs.7b09213

Burbage, M., Gros, M., and Amigorena, S. (2019). Translate less, prime better, to improve anti-tumor responses. *Nat. Immunol.* 20 (5), 518–520. doi:10.1038/s41590-019-0371-8

Chen, J., Solomides, C., Simpkins, F., and Simpkins, H. (2017a). The role of Nrf2 and ATF2 in resistance to platinum-based chemotherapy. *Cancer Chemother. Pharmacol.* 79 (2), 369–380. doi:10.1007/s00280-016-3225-1

Chen, J., Sun, Y., Xu, X., Wang, D., He, J., Zhou, H., et al. (2017b). YTH domain family 2 orchestrates epithelial-mesenchymal transition/proliferation dichotomy in pancreatic cancer cells. *Cell. Cycle* 16 (23), 2259–2271. doi:10.1080/15384101.2017.1380125

Chen, M., Wei, L., Law, C. T., Tsang, F. H., Shen, J., Cheng, C. L., et al. (2018). RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer

progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. *Hepatology* 67 (6), 2254–2270. doi:10.1002/hep.29683

Chen, X., Zhang, J., and Zhu, J. (2019). The role of m6A RNA methylation in human cancer. *Mol. Cancer* 18 (1), 103. doi:10.1186/s12943-019-1033-z

Chen, Y. G., Chen, R., Ahmad, S., Verma, R., Kasturi, S. P., Amaya, L., et al. (2019). N6-Methyladenosine modification controls circular RNA immunity. *Mol. Cell.* 76 (1), 96–109. doi:10.1016/j.molcel.2019.07.016

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). M6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell. Rep.* 18 (11), 2622–2634. doi:10.1016/j.celrep.2017.02.059

Dai, D., Wang, H., Zhu, L., Jin, H., and Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell. Death Dis.* 9 (2), 124. doi:10.1038/s41419-017-0129-x

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485 (7397), 201–206. doi:10.1038/nature11112

Edupuganti, R. R., Geiger, S., Lindeboom, R., Shi, H., Hsu, P. J., Lu, Z., et al. (2017). N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24 (10), 870–878. doi:10.1038/nsmb.3462

Fang, R., Chen, X., Zhang, S., Shi, H., Ye, Y., Shi, H., et al. (2021). EGFR/SRC/ERK-stabilized YTHDF2 promotes cholesterol dysregulation and invasive growth of glioblastoma. *Nat. Commun.* 12 (1), 177. doi:10.1038/s41467-020-20379-7

Fang, Z., Mu, B., Liu, Y., Guo, N., Xiong, L., Guo, Y., et al. (2022). Discovery of a potent, selective and cell active inhibitor of m6A demethylase ALKBH5. *Eur. J. Med. Chem.* 238, 114446. doi:10.1016/j.ejmech.2022.114446

Gilbert, W. V., Bell, T. A., and Schaening, C. (2016). Messenger RNA modifications: Form, distribution, and function. *Science* 352 (6292), 1408–1412. doi:10.1126/science.aad8711

Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour immunity controlled through mRNA m6A methylation and YTHDF1 in dendritic cells. *Nature* 566 (7743), 270–274. doi:10.1038/s41586-019-0916-x

Hao, L., Wang, J., Liu, B., Yan, J., Li, C., Jiang, J., et al. (2021). M6A-YTHDF1-mediated TRIM29 upregulation facilitates the stem cell-like phenotype of cisplatin-resistant ovarian cancer cells. *Biochim. Biophys. Acta. Mol. Cell. Res.* 1868 (1), 118878. doi:10.1016/j.bbamcr.2020.118878

He, L., Li, H., Wu, A., Peng, Y., Shu, G., and Yin, G. (2019). Functions of N6-methyladenosine and its role in cancer. *Mol. Cancer* 18 (1), 176. doi:10.1186/s12943-019-1109-9

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: Predict and validate. *Nat. Rev. Genet.* 18 (5), 275–291. doi:10.1038/nrg.2016.169

Huang, G., Wang, M., Li, X., Wu, J., Chen, S., Du, N., et al. (2019). TUSC7 suppression of Notch activation through sponging MiR-146 recapitulated the asymmetric cell division in lung adenocarcinoma stem cells. *Life Sci.* 232, 116630. doi:10.1016/j.lfs.2019.116630

Huo, F. C., Zhu, Z. M., Zhu, W. T., Du, Q. Y., Liang, J., and Mou, J. (2021). METTL3-mediated m6A methylation of SPHK2 promotes gastric cancer progression by targeting KLF2. *Oncogene* 40 (16), 2968–2981. doi:10.1038/s41388-021-01753-1

Ianniello, Z., and Fatica, A. (2018). N6-Methyladenosine role in acute myeloid leukaemia. *Int. J. Mol. Sci.* 19 (8), 2345. doi:10.3390/ijms19082345

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., et al. (2017). The RNA m6A reader YTHDF2 is essential for the post-transcriptional regulation of the maternal transcriptome and oocyte competence. *Mol. Cell.* 67 (6), 1059–1067. doi:10.1016/j.molcel.2017.08.003

Jain, S., Wheeler, J. R., Walters, R. W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated stress granules contain a diverse proteome and substructure. *Cell.* 164 (3), 487–498. doi:10.1016/j.cell.2015.12.038

Jang, D., Kwon, H., Choi, M., Lee, J., and Pak, Y. (2019). Sumoylation of Flotillin-1 promotes EMT in metastatic prostate cancer by suppressing Snail degradation. *Oncogene* 38 (17), 3248–3260. doi:10.1038/s41388-018-0641-1

Jia, R., Chai, P., Wang, S., Sun, B., Xu, Y., Yang, Y., et al. (2019). M6A modification suppresses ocular melanoma through modulating HINT2 mRNA translation. *Mol. Cancer* 18 (1), 161. doi:10.1186/s12943-019-1088-x

Jin, D., Guo, J., Wu, Y., Du, J., Yang, L., Wang, X., et al. (2019). M6A mRNA methylation initiated by METTL3 directly promotes YAP translation and increases YAP activity by regulating the MALAT1-miR-1914-3p-YAP axis to induce NSCLC drug resistance and metastasis. *J. Hematol. Oncol.* 12 (1), 135. doi:10.1186/s13045-019-0830-6

Jin, H., Ying, X., Que, B., Wang, X., Chao, Y., Zhang, H., et al. (2019). N6-methyladenosine modification of ITGA6 mRNA promotes the development and progression of bladder cancer. *EBioMedicine* 47, 195–207. doi:10.1016/j.ebiom.2019.07.068

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., et al. (2012). Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell.* 149 (4), 753–767. doi:10.1016/j.cell.2012.04.017

Laudato, S., Patil, N., Abba, M. L., Leupold, J. H., Benner, A., Gaiser, T., et al. (2017). P53-induced miR-30e-5p inhibits colorectal cancer invasion and metastasis by targeting ITGA6 and ITGB1. *Int. J. Cancer* 141 (9), 1879–1890. doi:10.1002/ijc.30854

Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., et al. (2017). Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell. Res.* 27 (3), 444–447. doi:10.1038/cr.2017.10

Li, J., Han, Y., Zhang, H., Qian, Z., Jia, W., Gao, Y., et al. (2019). The m6A demethylase FTO promotes the growth of lung cancer cells by regulating the m6A level of USP7 mRNA. *Biochem. Biophys. Res. Commun.* 512 (3), 479–485. doi:10.1016/j.bbrc.2019.03.093

Li, K., Peng, Z., Gao, S., Wang, Q., Wang, R., Li, X., et al. (2021). M6A associated TSUC7 inhibition contributed to Erlotinib resistance in lung adenocarcinoma through a notch signaling activation dependent way. *J. Exp. Clin. Cancer Res.* 40 (1), 325. doi:10.1186/s13046-021-02137-9

Li, Y., Bedi, R. K., Moroz-Omori, E. V., and Caflisch, A. (2020). Structural and dynamic insights into redundant function of YTHDF proteins. *J. Chem. Inf. Model.* 60 (12), 5932–5935. doi:10.1021/acs.jcim.0c01029

Liao, S., Sun, H., and Xu, C. (2018). YTH domain: A family of N(6)-methyladenosine (m(6)A) readers. *Genomics Proteomics Bioinforma.* 16 (2), 99–107. doi:10.1016/j.gpb.2018.04.002

Lin, X., Chai, G., Wu, Y., Li, J., Chen, F., Liu, J., et al. (2019). RNA m6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. *Nat. Commun.* 10 (1), 2065. doi:10.1038/s41467-019-09865-9

Lin, X., Wang, Z., Yang, G., Wen, G., and Zhang, H. (2020). YTHDF2 correlates with tumor immune infiltrates in lower-grade glioma. *Aging* 12 (18), 18476–18500. doi:10.18632/aging.103812

Lin, Y., Jin, X., Nie, Q., Chen, M., Guo, W., Chen, L., et al. (2022). YTHDF3 facilitates triple-negative breast cancer progression and metastasis by stabilizing ZEB1 mRNA in an m6A-dependent manner. *Ann. Transl. Med.* 10 (2), 83. doi:10.21037/atm-21-6857

Liu, Y., Guo, Q., Yang, H., Zhang, X. W., Feng, N., Wang, J. K., et al. (2022). Allosteric regulation of IGF2BP1 as a novel strategy for the activation of tumor immune microenvironment. *ACS Cent. Sci.* doi:10.1021/acscentsci.2c00107

Luo, X., Cao, M., Gao, F., and He, X. (2021). YTHDF1 promotes hepatocellular carcinoma progression via activating PI3K/AKT/mTOR signaling pathway and inducing epithelial-mesenchymal transition. *Exp. Hematol. Oncol.* 10 (1), 35. doi:10.1186/s40164-021-00227-0

Ma, C., Liao, S., and Zhu, Z. (2019). Crystal structure of human YTHDC2 YTH domain. *Biochem. Biophys. Res. Commun.* 518 (4), 678–684. doi:10.1016/j.bbrc.2019.08.107

Ma, S., Yan, J., Barr, T., Zhang, J., Chen, Z., Wang, L., et al. (2021). The RNA m6A reader YTHDF2 controls NK cell anti-tumor and antiviral immunity. *J. Exp. Med.* 218 (8), e20210279. doi:10.1084/jem.20210279

Maity, A., and Das, B. (2016). N6-methyladenosine modification in mRNA: Machinery, function and implications for health and diseases. *Febs. J.* 283 (9), 1607–1630. doi:10.1111/febs.13614

Meyer, K. D., and Jaffrey, S. R. (2017). Rethinking m6A readers, writers, and erasers. *Annu. Rev. Cell. Dev. Biol.* 33 (1), 319–342. doi:10.1146/annurev-cellbio-100616-060758

Ni, W., Yao, S., Zhou, Y., Liu, Y., Huang, P., Zhou, A., et al. (2019). Long noncoding RNA GAS5 inhibits progression of colorectal cancer by interacting with and triggering YAP phosphorylation and degradation and is negatively regulated by the m6A reader YTHDF3. *Mol. Cancer* 18 (1), 143. doi:10.1186/s12943-019-1079-y

Nishizawa, Y., Konno, M., Asai, A., Koseki, J., Kawamoto, K., Miyoshi, N., et al. (2018). Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression in colorectal cancer. *Oncotarget* 9 (7), 7476–7486. doi:10.18632/oncotarget.23554

Pak, O., Zaitsev, S., Shevchenko, V., Sharma, A., Sharma, H. S., and Bryukhovetskiy, I. (2021). Effectiveness of bortezomib and temozolomide for eradication of recurrent human glioblastoma cells, resistant to radiation. *Prog. Brain Res.* 266, 195–209. doi:10.1016/bs.pbr.2021.06.010

Pan, H., Pan, Z., Guo, F., Meng, F., Zu, L., Fan, Y., et al. (2021). MicroRNA-1915-3p inhibits cell migration and invasion by targeting SET in non-small-cell lung cancer. *BMC Cancer* 21 (1), 1218. doi:10.1186/s12885-021-08961-8

Pi, J., Wang, W., Ji, M., Wang, X., Wei, X., Jin, J., et al. (2021). YTHDF1 promotes gastric carcinogenesis by controlling translation of FZD7. *Cancer Res.* 81 (10), 2651–2665. doi:10.1158/0008-5472.CAN-20-0066

Pu, Y., Lu, X., Yang, X., Yang, Y., Wang, D., Li, M., et al. (2022). Estimating the prognosis of esophageal squamous cell carcinoma based on the Cancer Genome Atlas (TCGA) of m6A methylation-associated genes. *J. Gastrointest. Oncol.* 13 (1), 1–12. doi:10.21037/jgo-21-686

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., et al. (2019). M6A enhances the phase separation potential of mRNA. *Nature* 571 (7765), 424–428. doi:10.1038/s41586-019-1374-1

Sheng, H., Li, Z., Su, S., Sun, W., Zhang, X., Li, L., et al. (2020). YTH domain family 2 promotes lung cancer cell growth by facilitating 6-phosphogluconate dehydrogenase mRNA translation. *Carcinogenesis* 41 (5), 541–550. doi:10.1093/carcin/bgz152

Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., et al. (2017). YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell. Res.* 27 (3), 315–328. doi:10.1038/cr.2017.15

Shi, Y., Fan, S., Wu, M., Zuo, Z., Li, X., Jiang, L., et al. (2019). YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. *Nat. Commun.* 10 (1), 4892. doi:10.1038/s41467-019-12801-6

Stoilov, P., Rafalska, I., and Stamm, S. (2002). Yth: A new domain in nuclear proteins. *Trends biochem. Sci.* 27 (10), 495–497. doi:10.1016/s0968-0004(02)02189-8

Su, G., Liu, T., Han, X., Sun, H., Che, W., Hu, K., et al. (2021). YTHDF2 is a potential biomarker and associated with immune infiltration in kidney renal clear cell carcinoma. *Front. Pharmacol.* 12, 709548. doi:10.3389/fphar.2021.709548

Su, R., Dong, L., Li, Y., Gao, M., Han, L., Wunderlich, M., et al. (2020). Targeting FTO suppresses cancer stem cell maintenance and immune evasion. *Cancer Cell.* 38 (1), 79–96. e11. doi:10.1016/j.ccell.2020.04.017

Tsuchiya, K., Yoshimura, K., Inoue, Y., Iwashita, Y., Yamada, H., Kawase, A., et al. (2021). YTHDF1 and YTHDF2 are associated with better patient survival and an inflamed tumor-immune microenvironment in non–small-cell lung cancer. *OncoImmunology* 10 (1), 1962656. doi:10.1080/2162402X.2021.1962656

Wang, L., Hui, H., Agrawal, K., Kang, Y., Li, N., Tang, R., et al. (2020). M6a RNA methyltransferases METTL3/14 regulate immune responses to anti-PD-1 therapy. *EMBO J.* 39 (20), e104514. doi:10.15252/embj.2020104514

Wang, M., Yang, Y., Yang, J., Yang, J., and Han, S. (2020). Circ_KIAA1429 accelerates hepatocellular carcinoma advancement through the mechanism of m6A-YTHDF3-Zeb1. *Life Sci.* 257, 118082. doi:10.1016/j.lfs.2020.118082

Wang, S., Gao, S., Zeng, Y., Zhu, L., Mo, Y., Wong, C., et al. (2022). N6-Methyladenosine reader YTHDF1 promotes ARHGEF2 translation and RhoA signaling in colorectal cancer. *Gastroenterology* 162 (4), 1183–1196. doi:10.1053/j.gastro.2021.12.269

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505 (7481), 117–120. doi:10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N6-methyladenosine modulates messenger RNA translation efficiency. *Cell.* 161 (6), 1388–1399. doi:10.1016/j.cell.2015.05.014

Wu, X., Sang, L., and Gong, Y. (2018). N6-methyladenine RNA modification and cancers. *Am. J. Cancer Res.* 8 (10), 1957–1966.

Xie, H., Li, J., Ying, Y., Yan, H., Jin, K., Ma, X., et al. (2020). METTL3/YTHDF2 m6A axis promotes tumorigenesis by degrading SETD7 and KLF4 mRNAs in bladder cancer. *J. Cell. Mol. Med.* 24 (7), 4092–4104. doi:10.1111/jcmm.15063

Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., and Min, J. (2015). Structural basis for the discriminative recognition of N6-Methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem.* 290 (41), 24902–24913. doi:10.1074/jbc.M115.680389

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10 (11), 927–929. doi:10.1038/nchembio.1654

Xu, C., Yuan, B., He, T., Ding, B., and Li, S. (2020). Prognostic values of YTHDF1 regulated negatively by mir-3436 in Glioma. *J. Cell. Mol. Med.* 24 (13), 7538–7549. doi:10.1111/jcmm.15382

Yang, S., Wei, J., Cui, Y., Park, G., Shah, P., Deng, Y., et al. (2019). M6A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. *Nat. Commun.* 10 (1), 2782. doi:10.1038/s41467-019-10669-0

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., et al. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* 593 (7860), 597–601. doi:10.1038/s41586-021-03536-w

Yarchoan, M., Johnson, B. A., Lutz, E. R., Laheru, D. A., and Jaffee, E. M. (2017). Targeting neoantigens to augment anti-tumour immunity. *Nat. Rev. Cancer* 17 (4), 209–222. doi:10.1038/nrc.2016.154

Zhang, C., Huang, S., Zhuang, H., Ruan, S., Zhou, Z., Huang, K., et al. (2020). YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation. *Oncogene* 39 (23), 4507–4518. doi:10.1038/s41388-020-1303-7

Zhang, J., Pi, J., Liu, Y., Yu, J., and Feng, T. (2018). Knockdown of YTH N6-methyladenosine RNA binding protein 2 (YTHDF2) inhibits proliferation and promotes apoptosis in MGC-803 gastric cancer cells. *Cell. Mol. Immunol.* 33 (12), 1628–1634.

Zhang, Y., Li, L., Ye, Z., Zhang, L., Yao, N., and Gai, L. (2021). Identification of m6A methyltransferase-related genes predicts prognosis and immune infiltrates in head and neck squamous cell carcinoma. *Ann. Transl. Med.* 9 (20), 1554. doi:10.21037/atm-21-4712

Zhao, G., An, J., Pu, Q., Geng, W., Song, H., Zhao, Q., et al. (2021). Gene signatures and Cancer-Immune phenotypes based on m6A regulators in breast cancer. *Front. Oncol.* 11, 756412. doi:10.3389/fonc.2021.756412

Zhao, X., Chen, Y., Mao, Q., Jiang, X., Jiang, W., Chen, J., et al. (2018). Overexpression of YTHDF1 is associated with poor prognosis in patients with hepatocellular carcinoma. *Cancer Biomark.* 21 (4), 859–868. doi:10.3233/CBM-170791

Zheng, B., Wang, J., Zhao, G., Chen, X., Yao, Z., Niu, Z., et al. (2021). A new m6A methylation-related gene signature for prognostic value in patient with urothelial carcinoma of the bladder. *Biosci. Rep.* 41 (4), BSR20204456. doi:10.1042/BSR20204456

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 10 August 2022
DOI 10.3389/fgene.2022.891214

OPEN ACCESS

EDITED BY
Mee-Hyun Lee,
Dongshin University, South Korea

REVIEWED BY
Xujie Zhao,
St. Jude Children's Research Hospital, United States
Sonali P. Barwe,
Alfred I. duPont Hospital for Children, United States

*CORRESPONDENCE
Jixia Li,
j.li@auckland.ac.nz
Maggie L. Kalev-Zylinska,
m.kalev@auckland.ac.nz

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 07 March 2022
ACCEPTED 11 July 2022
PUBLISHED 10 August 2022

CITATION
Li J and Kalev-Zylinska ML (2022), Advances in molecular characterization of myeloid proliferations associated with Down syndrome.
Front. Genet. 13:891214.
doi: 10.3389/fgene.2022.891214

Advances in molecular characterization of myeloid proliferations associated with Down syndrome

Jixia Li [1,2]* and Maggie L. Kalev-Zylinska [1,3]*

[1]Blood and Cancer Biology Laboratory, Department of Molecular Medicine and Pathology, University of Auckland, Auckland, New Zealand, [2]Department of Laboratory Medicine, School of Medicine, Foshan University, Foshan, China, [3]Haematology Laboratory, Department of Pathology and Laboratory Medicine, Auckland City Hospital, Auckland, New Zealand

Myeloid leukemia associated with Down syndrome (ML-DS) has a unique molecular landscape that differs from other subtypes of acute myeloid leukemia. ML-DS is often preceded by a myeloproliferative neoplastic condition called transient abnormal myelopoiesis (TAM) that disrupts megakaryocytic and erythroid differentiation. Over the last two decades, many genetic and epigenetic changes in TAM and ML-DS have been elucidated. These include overexpression of molecules and micro-RNAs located on chromosome 21, *GATA1* mutations, and a range of other somatic mutations and chromosomal alterations. In this review, we summarize molecular changes reported in TAM and ML-DS and provide a comprehensive discussion of these findings. Recent advances in the development of CRISPR/Cas9-modified induced pluripotent stem cell-based disease models are also highlighted. However, despite significant progress in this area, we still do not fully understand the pathogenesis of ML-DS, and there are no targeted therapies. Initial diagnosis of ML-DS has a favorable prognosis, but refractory and relapsed disease can be difficult to treat; therapeutic options are limited in Down syndrome children by their stronger sensitivity to the toxic effects of chemotherapy. Because of the rarity of TAM and ML-DS, large-scale multi-center studies would be helpful to advance molecular characterization of these diseases at different stages of development and progression.

KEYWORDS

leukemia, acute myeloid leukemia, acute megakaryoblastic leukemia, transient abnormal myelopoiesis, Down syndrome, epigenetics, genomics, GATA1 mutations

Introduction

Myeloid leukemia associated with Down syndrome (ML-DS) is a unique category of acute myeloid leukemia (AML) most often of the megakaryoblastic subtype (i.e., acute megakaryoblastic leukemia (AMKL), formerly known as AML-M7 (Singh et al., 2017). The term ML-DS also includes an antecedent myelodysplastic syndrome (MDS)-like phase. There is no biologic or prognostic difference between MDS (blasts 5–19%) and

AML (blasts ≥20%) in Down syndrome (DS) (Lange et al., 1998), therefore this distinction is not being made for ML-DS in the current pathologic classification (Arber et al., 2016).

ML-DS is frequently preceded by transient abnormal myelopoiesis (TAM), a unique myeloproliferative disorder affecting megakaryocytic and erythroid lineages. TAM is a pre-leukemic condition characterized by reduced platelet and increased leukocyte counts, and the presence of blasts in the peripheral blood. TAM diagnosis requires the presence of *GATA1* mutations together with increased blasts and/or certain clinical features (in particular hepatosplenomegaly) in a neonate with constitutional trisomy 21, which can be mosaic (Arber et al., 2016). TAM may be indistinguishable from ML-DS but there is a wide spectrum of clinical presentation, ranging from asymptomatic to a stormy course and fatal outcome. Typically, TAM presents in neonates 3–7 days after birth but it may present within 2 months from birth (Singh et al., 2017). Overt TAM (blasts >10%) occurs in approximately 10–15% of DS neonates, but a further 10–15% may have *GATA1* mutations detectable only by sensitive methods with no clinical or hematologic manifestations (i.e., silent TAM) (Roberts et al., 2013). Most patients with TAM recover spontaneously within 3 months, but some require cytotoxic therapy. Unfortunately, despite initial TAM resolution, 20–30% of children progress to ML-DS within 4 years (Bombery and Vergilio, 2014).

TAM is extremely rare in neonates without DS but such cases have been well documented (Apollonsky et al., 2008; Tsai et al., 2011; Ono et al., 2015; Yuzawa et al., 2020; Panferova et al., 2021). The molecular pathogenesis and clinical outcomes of TAM in neonates without DS are similar to those with DS (i.e., DS-like). These patients acquire trisomy 21 and *GATA1* mutations in the TAM clone (Yuzawa et al., 2020; Panferova et al., 2021). In addition, *GATA1* mutations may be germline, as recently reported in familial childhood cases of TAM/AMKL, highlighting a unique functional cooperation between these lesions that may be independent of the order of their acquisition (Hasle et al., 2022). The rates of early death and leukemic progression of TAM in non-DS and DS children are similar, emphasizing the importance of making the diagnosis of DS-like TAM to assist appropriate patient management (Yuzawa et al., 2020). Rare cases of TAM without *GATA1* mutations feature in the literature. However, this may be due to technical limitations, in particular prior to the use of sensitive next-generation-sequencing methods (Panferova et al., 2021), small disease clones, or the lack of appropriate diagnostic samples if the condition is not suspected at presentation (Aksu et al., 2020). The expanding use of sensitive sequencing technologies will make the diagnosis of DS-like TAM easier in the future, which should advance our knowledge about this extremely rare condition.

ML-DS often presents with a period of thrombocytopenia reflecting a prodromal MDS-like phase (Lange et al., 1998). ML-DS is characterized by the expansion of megakaryoblasts, frequent bone marrow fibrosis, and the presence of *GATA1* mutations in the blasts that drive expression of a truncated (short) GATA1 protein (GATA1s) (Hasle et al., 2008). The median age of patients with ML-DS is 1–1.8 years (Gamis et al., 2003; Bhatnagar et al., 2016). Majority of patients with ML-DS (72%) also carry other cytogenetic changes in addition to trisomy 21 (Forestier et al., 2008; De Souza et al., 2017). The contribution of these changes to disease development and progression is the subject of active research.

ML-DS pathogenesis is understood to follow a multistep clonal evolution process (Figure 1). Trisomy 21 represents a "primary hit", which alters hematopoiesis during embryonic development; acquisition of somatic *GATA1s* mutations represents a "secondary hit", which promotes hematopoietic deregulation and emergence of TAM in DS newborns; additional mutations predominantly affecting chromatin and epigenetic regulators (e.g., the cohesin complex) and signaling mediators (e.g., Janus kinase 2, JAK2) represent a "tertiary hit", which leads to ML-DS (Labuhn et al., 2019; Garnett et al., 2020; De Castro et al., 2021) (Figure 1). The detailed mechanism of how these events contribute to different stages of disease is still unclear. One of the studies showed that *GATA1s* mutations lead to TAM when introduced into trisomy 21 long-term hematopoietic stem cells (LT-HSCs), where a subset of chromosome 21 microRNAs (miRNAs) influences predisposition toward pre-leukemia initiation (Wagenblast et al., 2021). However, progression to ML-DS was independent of trisomy 21 in this study, but required synergy between mutations in *GATA1* and the cohesin genes, in particular cohesin subunit SA-2 (*STAG2*) knockout occurring in fetal or early postnatal but not adult HSCs (Wagenblast et al., 2021). Our review was motivated by these and other recent advances in the field that will likely open up new lines of research into ML-DS pathogenesis and targeted treatment development. We provide a comprehensive and up-to-date summary of molecular alterations in ML-DS, with an overriding aim to help guide future mechanistic studies into the pathogenesis of this disease. However, this is a rapidly advancing field, so despite our efforts this review may not be complete.

Trisomy 21

Trisomy 21 is associated with defects in hematopoiesis and the immune system. Trisomy 21 fetuses have dysregulated development of megakaryocytic, erythroid and B-cell lineages (Laurent et al., 2020; De Castro et al., 2021). The mechanism through which an extra copy of chromosome 21 perturbs hematopoiesis and then how it cooperates with subsequent mutations to lead to TAM and ML-DS are still uncertain. Findings from a humanized model of pre-malignant and malignant stages of ML-DS demonstrated that trisomy 21 was necessary for pre-leukemia initiation but dispensable for leukemia progression (Wagenblast et al., 2021). The predominant

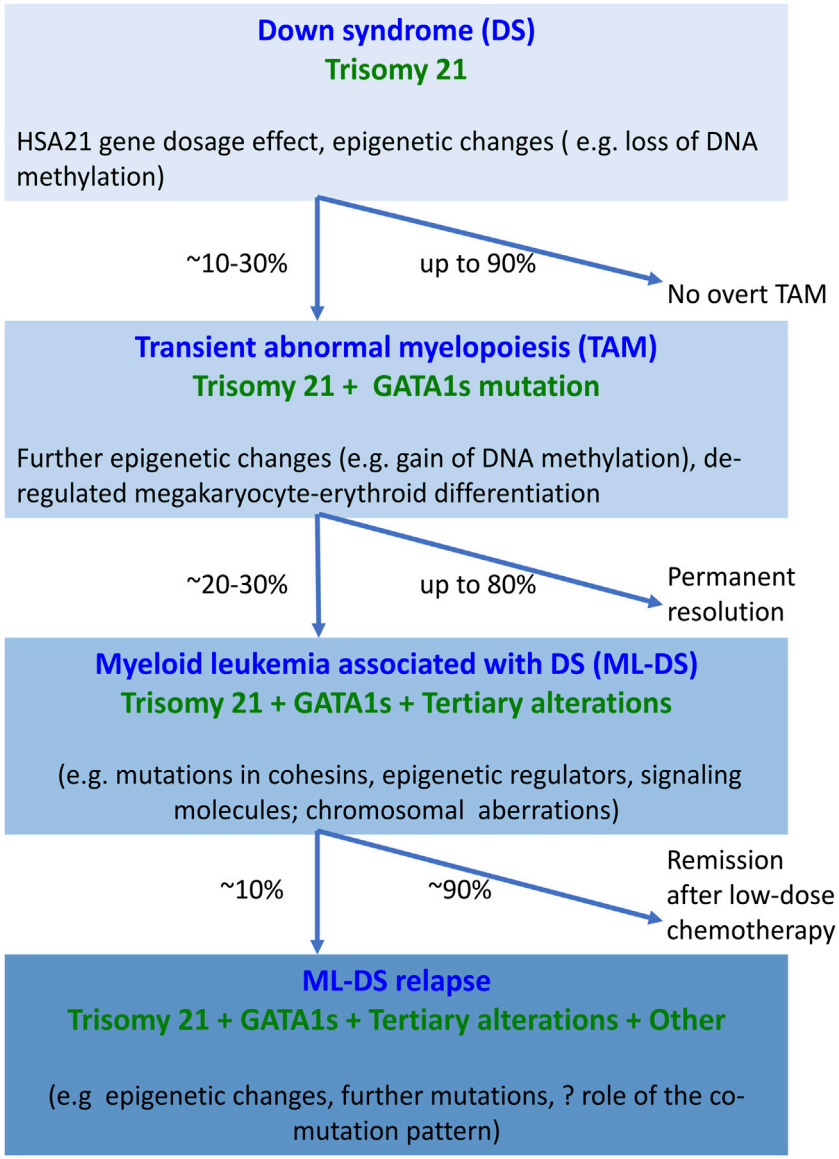


FIGURE 1
Overview of molecular changes reported at different stages of myeloid proliferation associated with Down syndrome. Trisomy 21 alone disturbs hematopoiesis through the increased dosage of HSA21-located genes and alterations in the epigenome, resulting in increased megakaryopoiesis. The combination of trisomy 21 and GATA1s causes expansion of megakaryocytic progenitors. Progression of TAM to ML-DS requires the interaction of GATA1s with additional somatic mutations and chromosomal structural abnormalities. Little is known about the molecular landscape of refractory or relapsed ML-DS. Abbreviations: DS, Down syndrome; GATA1s, GATA1 short; HSA21, human chromosome 21; ML-DS, myeloid leukemia associated with Down syndrome; TAM, transient abnormal myelopoiesis.

current view is that DS-associated myeloproliferations result from deregulation of genes on human chromosome 21 (HSA21) estimated to contain 234 protein-coding genes (Antonarakis, 2017). These include genes critical for myeloid differentiation, such as ETS-related gene (*ERG*), ETS proto-oncogene 2 (*ETS2*), runt-related transcription factor 1 (*RUNX1*), dual specificity tyrosine phosphorylation regulated kinase 1A (*DYRK1A*), regulator of calcineurin 1 (*RCAN1*), chromatin assembly factor 1 subunit B (*CHAF1B*), high mobility group nucleosome binding domain 1 (*HMGN1*), SON DNA and RNA binding protein (*SON*), and a subset of miRNAs (Table 1) (Laurent et al., 2020; Vukadin et al., 2021; Wagenblast et al., 2021). The encoded molecules belong to several functional classes, such as transcription factors, signaling effectors, epigenetic regulators, and miRNAs.

TABLE 1 List of HSA21 genes involved in myeloid proliferation associated with Down syndrome.

HSA21 genes	Classification	Function in hematopoiesis/leukemogenesis	References
ERG	Transcription factor	Causes megakaryoblastic expansion; involved in megakaryocytic leukemia; cooperates with GATA1s to drive TAM/ML-DS	Rainis et al. (2005); Salek-Ardakani et al. (2009); Stankiewicz and Crispino (2009); Carmichael et al. (2012); Stankiewicz and Crispino (2013); Banno et al. (2016)
EST2	Transcription factor	Regulates megakaryopoiesis; cooperates with GATA1s to drive TAM/ML-DS	Rainis et al. (2005); Ge et al. (2008); Stankiewicz and Crispino (2009); Banno et al. (2016)
RUNX1	Transcription factor	Involved in the pathogenesis of megakaryoblastic leukemia; causes abnormal megakaryocytic differentiation in cooperation with ERG, ETS2 and GATA1s; involved in TAM/ML-DS development	Elagib et al. (2003); Yanagida et al. (2005); Banno et al. (2016)
BACH1	Transcription factor	Inhibits megakaryocyte differentiation and platelet production	Toki et al. (2005)
SON	Transcription factor	Regulates hematopoiesis; represses megakaryocytic differentiation in megakaryoblastic leukemia	Belmonte et al. (2021); Vukadin et al. (2021)
C21ORF66	Transcription factor	Unknown	Reymond et al. (2001); Bourquin et al. (2006)
GABPA	Transcription factor	Regulates hematopoiesis and involved in CML development; role in TAM/ML-DS unknown	Yang et al. (2013); Manukjan et al. (2015)
DYRK1A	Signaling effector	Promotes TAM/ML-DS in human and murine models; co-operates with GATA1s to increase megakaryoblastic proliferation through NFAT inhibition	Malinge et al. (2012)
RCAN1	Signaling effector	Promotes megakaryopoiesis by inhibiting calcineurin-NFAT pathway	Zaslavsky et al. (2013)
HMGN1	Epigenetic modulator	Regulates myeloid differentiation; promotes leukemic stem cell activity by increasing H3K27 acetylation	Cabal-Hierro et al. (2020)
CHAF1B	Epigenetic modulator	Regulates hematopoiesis; impairs myeloid differentiation and promotes myeloid leukemia through binding of chromatin and interference with transcription factors such as CEBPA	Volk et al. (2018)
miR-99a	miRNA	Increases predisposition toward TAM but not ML-DS; has oncogenic function	Zhang et al. (2013); Si et al. (2016); Wagenblast et al. (2021)
miR-125b	miRNA	Increases predisposition toward TAM; regulates megakaryopoiesis; has oncogenic function; synergizes with GATA1s to induce megakaryoblastic leukemia	Klusmann et al. (2010); Alejo-Valle et al. (2021); Wagenblast et al. (2021)
miR-155	miRNA	Increases predisposition toward TAM but not ML-DS	Elton et al. (2010); Sas et al. (2020); Wagenblast et al. (2021)

Transcription factors

The roles of ERG, ETS2 and RUNX1 in hematopoiesis and leukemogenesis have been thoroughly studied. ERG, an ETS transcription factor, is a megakaryocytic oncogene; its overexpression facilitates megakaryocytic expansion and promotes lymphoid and erythro-megakaryocytic leukemia *in vitro* and *in vivo* (Rainis et al., 2005; Salek-Ardakani et al., 2009; Stankiewicz and Crispino, 2009; Carmichael et al., 2012). Increased expression of ERG alone contributes to rapid onset of leukemia in mice (Salek-Ardakani et al., 2009). ERG strongly cooperates with the GATA1s mutated protein to immortalize hematopoietic and megakaryocytic progenitors *ex vivo* (Salek-Ardakani et al., 2009; Stankiewicz and Crispino, 2009). ERG and protein kinase B (PKB) also crosstalk, which alters GATA1 function (Stankiewicz and Crispino, 2013). Similar to ERG, ETS2 is an ETS transcription factor and a megakaryocytic oncogene (Ge et al., 2008). ETS2 promotes megakaryopoiesis and collaborates with GATA1s to immortalize hematopoietic progenitor cells (HPCs) (Rainis et al., 2005; Stankiewicz and Crispino, 2009). RUNX1 is a crucial transcription factor involved in the regulation of megakaryopoiesis, and its expression and cooperation with GATA1s facilitates megakaryocytic differentiation (Elagib et al., 2003). In 2005, overexpression of *RUNX1* was reported in bone marrow of ML-DS children (Langebrake et al., 2006). A subsequent report from 2006 showed *RUNX1* expression was lower than anticipated in ML-DS, while it was higher in megakaryoblasts from children with non-DS-AMKL (Bourquin et al., 2006). It appears that *SON*, another HSA21 gene, inhibits *RUNX1* expression (Vukadin et al., 2021), which may neutralize trisomy 21-related overdosage of *RUNX1* effects. Evidence from animal studies indicates that *RUNX1* overexpression in mice shortens the latency of leukemia development displaying enhanced frequency of megakaryoblastic leukemia, which supports that *RUNX1* overexpression is leukemogenic in ML-DS (Yanagida et al., 2005). Data from disease models using human induced pluripotent stem cells (iPSCs) and

genome-editing technologies showed that an extra copy of *RUNX1* is essential for accelerating early hematopoiesis in the context of trisomy 21, leading to HPC expansion and increased myeloid differentiation (Banno et al., 2016). *RUNX1* expression level in trisomy 21 (*GATA1* wild type) iPSCs is increased by ~1.8-fold compared with that in disomy 21 (*GATA1* wild type) iPSCs, which is slightly higher than the expected change in gene dosage (Banno et al., 2016). Abnormal megakaryocyte differentiation in TAM is accelerated by trisomy 21. Trisomy 21 up-regulates *GATA1s* expression leading to aberrant megakaryopoiesis, and the overdosage of *RUNX1*, *ETS2*, and *ERG* accelerates production of aberrantly differentiated cells (Banno et al., 2016). These observations highlight the importance of synergy between trisomy 21 and *GATA1s* in driving myeloid proliferation in DS children.

Other transcription factor encoding genes located on HSA21 are also highly expressed in ML-DS, including BTB domain and CNC homolog 1 (*BACH1*) (1.98-fold), *SON* (1.84-fold), chromosome 21 open reading frame 66 (*C21ORF66*) (1.64-fold) and GA-binding protein alpha chain (*GABPA*) (1.53-fold) (Bourquin et al., 2006). BACH1 acts as a transcriptional repressor of normal megakaryopoiesis and is likely a target of GATA1 and SON (Bourquin et al., 2006). Overexpression of *BACH1* causes maturation arrest of megakaryocytes resulting in marked peripheral thrombocytopenia (Toki et al., 2005). *SON* is a gene with homology to the proto-oncogene MYC family, and an RNA splicing factor regulating transcription of leukemia-associated genes. *SON* is indispensable for proper blood cell formation, as *SON* knockdown results in lower amounts of all myeloid cells and T cells (Belmonte et al., 2021). Megakaryocytic differentiation in AMKL is impaired by *SON* inhibiting expression of *RUNX1* and other megakaryocytic genes (Vukadin et al., 2021). *SON* also negatively regulates the expression of the AP-1 complex subunits JUN, JUNB and FOSB, which suggests that overexpression of *SON* could be pathogenic in ML-DS (Vukadin et al., 2021). *C21ORF66* is known as the GC-rich sequence DNA-binding factor candidate (Reymond et al., 2001), but its function is unknown. Further work is needed to elucidate the role of *C21ORF66* in hematopoiesis and leukemogenesis. *GABPA* has a known role in hematopoiesis (Yang et al., 2013). Deletion of *GABPA* leads to cell cycle arrest in hematopoietic stem cells (HSCs) and profound loss of HPCs (Yang et al., 2013). *GABPA* is necessary for chronic myeloid leukemia (CML) development through its regulation of protein kinase D2 (PRKD2) (Yang et al., 2013). *GABPA* expression positively correlates with the *BCR::ABL1/ABL1* ratio in cells from patients with CML, and influences imatinib sensitivity in leukemic cell lines (TKI-sensitive K-562 and TKI-resistant NALM-1) (Manukjan et al., 2015). However, the function of *GABPA* in the setting of trisomy 21 and *GATA1* mutations is not clear.

Signaling effectors

DYRK1A belongs to the CMGC kinase group named after the initials of its subgroup members, including cyclin-dependent kinases, mitogen-activated protein kinases (MAPK), glycogen synthase kinases and CDK-like kinases. DYRK1A participates in various cellular functions through the phosphorylation of several substrates such as nuclear factor of activated T cells (NFAT) (Lindberg and Meijer, 2021). *DYRK1A* is a potent megakaryoblastic tumor-promoting gene, contributing to leukemogenesis in a mouse model containing 33 gene orthologs of HSA21, a *GATA1s* mutation, and a *MPL* mutation (Malinge et al., 2012). *DYRK1A* overexpression induces a marked megakaryoblastic proliferation through the suppression of NFAT in this model (Malinge et al., 2012). *RCAN1*, also known as Down syndrome critical region gene 1 (*DSCR1*), is an endogenous calcineurin inhibitor. Overexpression of *RCAN1* represses calcineurin-NFAT pathway, which leads to the expansion of megakaryocytes and their progenitors, and a high number of platelets (Zaslavsky et al., 2013). Both *DYRK1A* and *RCAN1* can down-regulate calcineurin-NFAT pathway, but little is known about how these signaling molecules collaborate with other HSA21 genes and *GATA1* mutations to initiate megakaryocytic neoplasia.

Epigenetic modulators

HMGN1 is the chromatin accessibility regulator and a target of recurrent DNA copy gains in leukemia (Cabal-Hierro et al., 2020). *HMGN1* overexpression blocks myeloid differentiation, increases clonal progenitor expansion, enhances HSC activity and leukemic stem cell (LSC) activity in the presence of RUNX1::RUNX1T1 fusion oncoprotein (Cabal-Hierro et al., 2020). In addition, *HMGN1* up-regulation elevates *H3K27* acetylation, and in turn histone acetyltransferase CBP/p300 inhibition reverses the HMGN1-induced differentiation arrest. Another epigenetic modulator coded by a gene on HSA21 is CHAF1B, representing the p60 subunit of the chromatin assembly factor complex (Volk et al., 2018). CHAF1B is essential for normal hematopoiesis, whereas its overexpression promotes leukemia by binding chromatin at discrete sites and interfering with the occupancy of CCAAT enhancer binding protein alpha (CEBPA) (Volk et al., 2018). *CHAF1B* expression is higher in patient cells from ML-DS than those of non-DS-AMKL (Malinge et al., 2012). Reducing CHAF1B activity is sufficient to suppress leukemogenesis in mice without impairing normal hematopoiesis, suggesting CHAF1B is a potential therapeutic target (Volk et al., 2018). Overall, HMGN1 and CHAF1B block myeloid differentiation and promote leukemia growth in other contexts but their roles in the initiation of TAM and progression to ML-DS are not known.

miRNAs

miRNAs, endogenous non-coding RNAs (~23 nucleotides in length), target mRNA of protein-coding genes to regulate expression, through which they control a range of cellular processes, such as cell proliferation, apoptosis, hematopoiesis and tumorigenesis (Brás et al., 2018). A number of HSA21 miRNA genes are up-regulated in DS, including miR-155, miR-802, miR-125b-2, let-7c and miR-99a. Deregulated expression of miRNAs may contribute to a range of phenotypes in patients with DS, not only leukemia but also brain pathology, congenital heart defects, as well as low incidence of solid tumors in DS individuals (Brás et al., 2018). The miR-99a~125b cluster, encoding let-7c, miR-99a and miR-125b, is highly expressed in TAM, ML-DS, and non-DS AMKL (Emmrich et al., 2014). The role of some HSA21 miRNAs in TAM/ML-DS pathogenesis has been partially revealed in recent years. (Alejo-Valle et al., 2021). *GATA1* mutations and miR-99a~125b cluster interact to induce the block in megakaryocytic differentiation that leads to the expansion of megakaryocytic progenitors and AMKL in a mouse model (Alejo-Valle et al., 2021). Another study highlighted the role of three HSA21 miRNAs (miR-99a, miR-125b-2, and miR-155) in the development of TAM, but not ML-DS (Wagenblast et al., 2021). Co-expression of miR-99a, miR-125b-2, and miR-155 in normal fetal liver LT-HSCs recapitulates features of a trisomy 21-like hematopoietic state, while deletion of these miRNAs reduces the blast population in the presence of *GATA1s*. Nevertheless, in the mouse model of ML-DS with and without deletion of HSA21 miRNAs blast numbers are similar (Wagenblast et al., 2021). Other studies suggest that miR-99a plays an oncogenic role through increasing proliferation and colony forming ability, and decreasing apoptosis of hematopoietic progenitors (Zhang et al., 2013; Si et al., 2016). miR-125b-2 is a positive modulator of megakaryopoiesis and an oncogenic miRNA in ML-DS. miR-125b-2 up-regulation promotes proliferation and self-renewal of megakaryocytic and megakaryocytic/erythroid progenitors, while its down-regulation inhibits growth of ML-DS cells (Klusmann et al., 2010). Moreover, miR-125b-mediated repression of the megakaryocytic transcription factor AT-rich interactive domain-containing protein 3A (ARID3A) is a critical event in ML-DS pathogenesis (Alejo-Valle et al., 2021). In the context of miR-125b overexpression and *GATA1s* mutations, *ARID3A* is the main target of miR-125b. Down-regulation of *ARID3A* blocks megakaryocytic differentiation and subsequently AMKL, while restoring *ARID3A* expression reverses megakaryocytic differentiation arrest in AMKL patient-derived xenografts. This suggests that restoration of *ARID3A* could be a promising strategy to inhibit megakaryoblastic leukemia growth. miR-155, a known regulator of the immune system, is also a crucial player in TAM through targeting tumor necrosis factor (TNF) superfamily receptors; miR-155 expression increases 2-fold and 3-fold in DS fetal and adult cells, respectively (Elton et al., 2010; Sas et al., 2020). How miR-155-modulated TNF receptor expression promotes TAM/ML-DS remains unknown.

Other effects of trisomy 21

Beyond the direct impact of HSA21 genes on myeloid proliferation, trisomy 21 also alters non-HSA21 gene expression through modulating genome organization (Letourneau et al., 2014; Liu et al., 2015; Ahlfors et al., 2019). Genome-wide studies showed that trisomy 21 has profound effects on DNA methylation in fetal and neonatal hematopoietic cells (Muskens et al., 2021). How these epigenetic changes influence TAM and ML-DS is not yet known. However, it has been shown that prior to the acquisition of *GATA1* mutations, trisomy 21 causes loss of DNA methylation at genes linked with the regulation of the cardiovascular, neurological, and endocrine organs. ML-DS has a unique epigenetic pattern characterized by gains of DNA methylation at genes correlated with hematopoiesis, cell proliferation, cell death, and cell cycle, which is distinct from other subtypes of pediatric AML, including non-DS-AMKL (Malinge et al., 2013). Significantly, TAM and ML-DS share the identical landscape of epigenetic changes (Malinge et al., 2013). Hence, it is possible that altered DNA methylation contributes towards development of TAM and ML-DS.

GATA1 mutations

Mutations in *GATA1* causing expression of its short isoform (*GATA1s*) are detected in nearly every case of TAM and ML-DS, implying mutated GATA1 deregulation plays a central role in TAM and ML-DS development (Wechsler et al., 2002; Panferova et al., 2021). The lack of detected *GATA1* mutations in ML-DS may be due to technical and sample limitations similar to those listed earlier for TAM. In addition, AMKL is associated with bone marrow fibrosis, which often impacts the quality and quantity of diagnostic bone marrow aspirate samples, in particular blast numbers. Because blasts are the cells that carry *GATA1s* in ML-DS, their paucity may limit detection. Similar to DS-like TAM, ML-DS-like leukaemia may arise in children without DS where *GATA1s* and trisomy 21 are somatically acquired in leukemic blasts (Ono et al., 2015; De Rooij et al., 2017; Panferova et al., 2021), or *GATA1s* mutations may be germline (Hasle et al., 2022). ML-DS-like leukemia is very rare but it shares multiple pathologic and clinical features with ML-DS, including good prognosis (De Rooij et al., 2017), emphasizing the importance of recognizing ML-DS-like leukemia in non-DS children.

GATA1 is encoded by the gene located on chromosome X and acts as a master transcription factor essential for the development of erythroid and megakaryocytic lineages (Pevny et al., 1991). More than 100 types of *GATA1* mutations have been reported in DS. These mutations are predominantly insertions, deletions, or duplications occurring in exon 2 or surrounding

sequences. *GATA1* mutations create an early stop codon that results in an exclusive expression of a short isoform of GATA1 protein (referred to as GATA1s) that lacks the N-terminal activation domain. Rarely, mutations in exon 3 generate GATA1 proteins with internal deletions. GATA1s can bind DNA but fails to initiate transcription, leading to deregulation of many downstream target genes (Wechsler et al., 2002). The cellular stage in which the functional and molecular consequences of *GATA1s* begin in the embryo has been narrowed down to the erythro-megakaryocytic subpopulation of progenitors with the following immunophenotype: $CD34^{+}CD43^{+}CD235$-$CD11b$-$CD71^{+}CD41^{+}$ (Nishinaka-Arai et al., 2021). The identification of this cellular stage should assist further studies into the pathogenesis of both TAM and ML-DS.

GATA1s promotes megakaryocytic progenitor expansion and disrupts megakaryocytic and erythroid differentiation (Shimizu et al., 2009; Chlon et al., 2015; Banno et al., 2016; Juban et al., 2021). This appears to involve synergistic interactions with other leukemogenic molecules; for example, *GATA1s* increases expression of miRNA-486-5p, an erythroid oncogenic miRNA (Shaham et al., 2015). In the presence of trisomy 21, *GATA1s* mutations are sufficient to drive TAM, and these mutations become undetectable when TAM resolves (Shimizu et al., 2009). Evidence from a range of cellular and animal disease models confirmed that TAM is initiated by increased gene dosage from chromosome 21 acting in cooperation with *GATA1s*. *GATA1s* mutation alone disrupts differentiation of megakaryocytes and promotes expansion of myeloid and megakaryocytic progenitors, while production of aberrant megakaryoblasts is strengthened on the background of trisomy 21 (Banno et al., 2016; Juban et al., 2021; Matsuo et al., 2021). TAM requires the synergy between trisomy 21 and *GATA1s* but leukemic transformation may be independent of trisomy 21 (Wagenblast et al., 2021; Arkoun et al., 2022). In contrast, synergy between *GATA1s* and subsequent "tertiary" molecular alterations is critical for progression of TAM to ML-DS. Evidence from sequential longitudinal studies highlights that pre-leukemic and leukemic clones are truly related, due to the fact that identical *GATA1* mutations are found in paired TAM and ML-DS samples (Hitzler et al., 2003; Saida et al., 2013). Although most TAM clones disappear by the age of 3 months, some heterogeneous clones persist during remission, and these carry different leukemia-initiating potential (Saida et al., 2013). ML-DS can be derived from a minor clone with a distinct *GATA1s* in TAM, but novel clones can also arise and become dominant (Xu et al., 2006; Saida et al., 2013; Labuhn et al., 2019).

So far, there is no solid proof of whether the type of *GATA1* mutations, the level of *GATA1s* expression, or the size of dominant *GATA1s*-bearing clones can predict progression from TAM to ML-DS (Alford et al., 2011; Grimm et al., 2021). Kanezaki et al. pointed out that the type of *GATA1* mutations influences expression of the GATA1s protein, and these expression levels are inversely linked with the risk of progression to ML-DS (Kanezaki et al., 2010). Nonetheless, in the clinical setting, persistence of *GATA1s* mutations is the most important risk factor associated with progression to ML-DS, even in cases with high GATA1s protein levels (Massey et al., 2006; Pine et al., 2007). The features used in the clinic to predict TAM progression to ML-DS include detection of minimal residual disease by flow cytometry (blasts >0.1%), persistence of patient-specific *GATA1s* mutation beyond 12 weeks from the initial diagnosis, and the appearance of thrombocytopenia (platelet count less than 100×10^{9}/L) (Klusmann et al., 2008; Flasinski et al., 2018).

Tertiary alterations

It has become well accepted that evolution from TAM to ML-DS relies on the acquisition of tertiary somatic mutations and additional chromosomal structural aberrations in *GATA1s*-mutated cells. Tertiary mutations seen in ML-DS most commonly affect genes encoding the cohesin complex, JAK family kinases, and epigenetic regulators; other mutations occur in genes recurrently mutated in other types of AML, including fms-like tyrosine kinase 3 (*FLT3*) and *TP53* (Table 2) (Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). Patients with TAM usually harbor fewer tertiary mutations than those with ML-DS, at the average of 0.4 and 1.6 variants per sample respectively (Labuhn et al., 2019). Most TAM cases carry only *GATA1s*, while additional somatic mutations are rare. Even if present in TAM, "third hit" mutations appear to be non-functional and un-linked from pre-leukemia or leukemia phenotype (Labuhn et al., 2019). By way of illustration, no autonomous or cytokine-induced signaling was found for *JAK1*, *JAK2*, *JAK3* or *MPL* variants by dual-luciferase assays with a signal transducer and activator of transcription 5 (STAT5) reporter at the TAM stage (Labuhn et al., 2019). During leukemic transformation, two to five additional mutations are found in a murine model of ML-DS. The most frequently altered genes encode signaling pathways (34%), members of the cohesin complex or its associated components (28.5%), and epigenetic regulators (22%) (Labuhn et al., 2019). The authors suggest that ML-DS progression is influenced by the cooperation between activated signaling pathways and deregulated epigenetic processes in the context of trisomy 21 and *GATA1s*. For instance, a remarkable co-occurrence of variants in genes encoding tyrosine kinases (e.g., JAK2-3) and RAS proteins with variants in epigenetic regulators (e.g., enhancer of zeste 2, EZH2) or cohesin genes has been shown in ML-DS mouse models and ML-DS patients (Labuhn et al., 2019). No tertiary mutations were detected in approximately 15–25% of ML-DS patients in relatively large studies reported in the last few years (Labuhn et al., 2019; Panferova et al., 2021). However, it is possible that such mutations will be detected in the future using updated sequencing methodologies. Karyotypic changes other than trisomy 21 may also contribute to ML-DS because such alterations are rarely found at the TAM stage.

TABLE 2 Recurrent somatic mutations reported in myeloid leukemia associated with Down syndrome.

Class	Mutant genes	Frequency of mutations in various studies n (%)	Function in hematopoiesis/ leukemogenesis; pathway to which it contributes	References
Cohesin complex and associated components	*CTCF*	16/141 (11.3); 10/49 (20.4); 5/44 (11.4)	Tumor suppressor; involved in chromatin organization, gene regulation, RNA splicing, myeloid cell growth and differentiation; contributes to leukemogenesis	Bell et al. (1999); Bell and Felsenfeld (2000); Torrano et al. (2005); Xu et al. (2007); Shukla et al. (2011); Yoshida et al. (2013); Zuin et al. (2014); Kim et al. (2017); Labuhn et al. (2019); Mujahed et al. (2020); Wang et al. (2020); Grimm et al. (2021); Panferova et al. (2021); Wagenblast et al. (2021)
	NIPBL	5/141 (3.5); 3/49 (6.1)	Cohesin regulator; regulates myeloid cell differentiation; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Mazzola et al. (2019b); Mazzola et al. (2020); Wagenblast et al. (2021)
	RAD21	16/141 (11.3); 11/49 (22.4); 6/44 (13.6)	Cohesin subunit; regulates gene expression, epigenetic modulation, HSPC self-renewal and differentiation; contributes to leukemogenesis	Yoshida et al. (2013); Fisher et al. (2017); Labuhn et al. (2019); Bisaillon et al. (2020); Panferova et al. (2021); Wagenblast et al. (2021)
	SMC1A	9/141 (6.4); 2/49 (4.1); 1/44 (2.3)	Cohesin subunit; regulates gene expression, genome organization; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Carico et al. (2021); Panferova et al. (2021); Wagenblast et al. (2021)
	SMC3	1/141 (0.7); 1/49 (2.0); 1/44 (2.3); 1/7 (14.3)	Cohesin ATPase subunit; contributes to hematopoietic failure and leukemogenesis	Nikolaev et al. (2013); Yoshida et al. (2013); Labuhn et al. (2019); Wang et al. (2019); Panferova et al. (2021); Rivas et al. (2021); Wagenblast et al. (2021); Arkoun et al. (2022)
	STAG2	19/141 (13.5); 9/49 (18.4); 4/44 (9.1)	Cohesin subunit; regulates gene expression, epigenetic modulation, HSPC self-renewal and differentiation; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Nie et al. (2019); Viny et al. (2019); Ochi et al. (2020; Panferova et al. (2021); Wagenblast et al. (2021); Arkoun et al. (2022); Barwe et al. (2022)
Epigenetic regulators	*ASXL1*	1/49 (2.0); 1/44 (2.3)	Regulates histone modifications; impairs hematopoiesis; involved in leukemogenesis	Yoshida et al. (2013); Nagase et al. (2018); Panferova et al. (2021)
	BCOR	2/141 (1.4); 2/49 (4.1); 1/44 (2.3)	Transcription factor; PRC1 component; leads to myeloid progenitor expansion; regulates myeloid differentiation; contributes to leukemogenesis	Yoshida et al. (2013); Kelly et al. (2019); Labuhn et al. (2019); Panferova et al. (2021)
	DNMT1	1/44 (2.3)	Involved in DNA methylation; regulates hematopoiesis; contributes to leukemogenesis	Panferova et al. (2021); Chattopadhyaya and Ghosal ((2022)
	DNMT3A	1/49 (2.0)	Involved in DNA methylation; causes HSC expansion and impairs differentiation	Yoshida et al. (2013); Izzo et al. (2020)
	EED	1/141 (0.7)	PRC2 subunit; increases HSPC proliferation and impairs differentiation; contributes to leukemogenesis	Ikeda et al. (2016); Labuhn et al. (2019)
	EP300	1/141 (0.7)	Transcriptional cofactor; chromatin modifier; increases HSCs self-renewal and impairs differentiation; contributes to leukemogenesis	Labuhn et al. (2019); Man et al. (2021)
	EZH2	10/141 (7.1); 16/49 (32.7); 1/44 (2.3); 1/7 (14.3)	Tumor suppressor; PRC2 subunit; chromatin modifier; regulates histone modifications; inhibits megakaryocyte differentiation; contributes to leukemogenesis	Ntziachristos et al. (2012); Nikolaev et al. (2013); Yoshida et al. (2013); Labuhn et al. (2019); Mazzi et al. (2021); Panferova et al. (2021)
	KANSL1	17/141 (12.1); 3/49 (6.1)	Regulates histone acetylation; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Wagenblast et al. (2021)
	KAT6A	1/44 (2.3)	Oncogene; regulates histone acetylation; impairs myeloid differentiation; contributes to leukemogenesis	Panferova et al. (2021); Yan et al. (2021)
	KDM6A	1/141 (0.7)	Regulates histone modifications; regulates hematopoiesis; contributes to leukemogenesis	Labuhn et al. (2019); Tian et al. (2021)
	KMT2C	1/141 (0.7)	Tumor suppressor; regulates histone modifications; involved in myelopoiesis; contributes to leukemogenesis	Labuhn et al. (2019); Maurya et al. (2021)
	NAT6	1/141 (0.7)	Regulates actin acetylation	Labuhn et al. (2019); Muffels et al. (2021)

(Continued on following page)

TABLE 2 (*Continued*) Recurrent somatic mutations reported in myeloid leukemia associated with Down syndrome.

Class	Mutant genes	Frequency of mutations in various studies n (%)	Function in hematopoiesis/ leukemogenesis; pathway to which it contributes	References
	SUZ12	9/141 (6.4); 1/49 (2.0); 1/ 44 (2.3)	PRC2 subunit; tumor suppressor; chromatin modifier; regulates histone modifications and HSCs activity; contributes to leukemogenesis	Majewski et al. (2008); Ntziachristos et al. (2012); Yoshida et al. (2013); Labuhn et al. (2019); Panferova et al. (2021)
	TET2	2/141 (1.4); 3/44 (6.8)	Involved in DNA methylation; causes HSC expansion and impairs differentiation	Labuhn et al. (2019); Izzo et al. (2020); Panferova et al. (2021)
Tyrosine kinases	*FLT3*	1/44 (2.3); 1/7 (14.3); 2/7 (28.6)	PI3K-PKB; MAPK; regulates hematopoiesis; contributes to leukemogenesis	Gilliland and Griffin (2002); Malinge et al. (2008); Grafone et al. (2012); Nikolaev et al. (2013); Panferova et al. (2021)
	GNB1	1/141 (0.7)	PI3K-PKB; MAPK	Zimmermannova et al. (2017); Labuhn et al. (2019)
	JAK1	6/141 (4.3); 2/49 (4.1); 3/44 (6.8); 1/7 (14.3)	JAK-STAT; regulates hematopoiesis; contributes to leukemogenesis	Nikolaev et al. (2013); Yoshida et al. (2013); Labuhn et al. (2019); Fasouli and Katsantoni (2021); Panferova et al. (2021)
	JAK2	14/141 (9.9); 4/49 (8.2); 4/44 (9.1); 1/7 (14.3)	JAK-STAT; regulates hematopoiesis; contributes to leukemogenesis	Malinge et al. (2008); Yoshida et al. (2013); Labuhn et al. (2019); Fasouli and Katsantoni (2021); Panferova et al. (2021)
	JAK3	19/141 (13.5); 6/49 (12.2); 12/44 (27.3); 1/7 (14.3); 1/11 (9.1); 1/3 (33.3); 1/14 (7.1)	JAK-STAT; regulates hematopoiesis; contributes to leukemogenesis	Walters et al. (2006); Kiyoi et al. (2007); Klusmann et al. (2007); Malinge et al. (2008); Yoshida et al. (2013); Labuhn et al. (2019); Fasouli and Katsantoni (2021); Panferova et al. (2021)
	KIT	2/141 (1.4)	Kit signaling; regulates hematopoiesis; contributes to leukemogenesis	Stankov et al. (2014); Labuhn et al. (2019)
	MPL	10/141 (7.1); 3/49 (6.1); 1/ 44 (2.3)	MPL signaling; JAK-STAT; regulates megakaryopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Loscocco et al. (2020); Nakamura-Ishizu and Suda (2020); Panferova et al. (2021); Arkoun et al. (2022)
	NTRK3	1/44 (2.3)	Oncogene; JAK-STAT; PI3K/PKB; MAPK; contributes to leukemogenesis	Joshi et al. (2020); Panferova et al. (2021)
	PI3KC2A	1/7 (14.3)	PI3K member; insulin signaling; human cytomegalovirus virions production	Nikolaev et al. (2013); Polachek et al. (2016); Zhuo et al. (2017)
	PTEN	1/141 (0.7)	Tumor suppressor; PI3K/PKB/mTOR; regulates hematopoiesis; contributes to leukemogenesis	Labuhn et al. (2019); Wu et al. (2020)
	PTPRD	1/141 (0.7)	Tumor suppressor; contributes to leukemogenesis	Song et al. (2016); Labuhn et al. (2019)
	SH2B3	4/141 (2.8); 4/49 (8.2); 2/ 44 (4.5)	JAK-STAT, PKB, MAPK; regulates thrombopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Maslah et al. (2017); Labuhn et al. (2019); Panferova et al. (2021)
RAS	*KRAS*	7/141 (5.0); 4/49 (8.2); 2/ 44 (4.5)	Oncogene; RAS signaling; KRAS/RAC1/ROS/ NLRP3/IL-1β; regulates hematopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Sasine et al. (2018); Labuhn et al. (2019); Hamarsheh et al. (2020); Panferova et al. (2021)
	NF1	4/141 (2.8)	Tumor suppressor; RAS signaling; regulates hematopoiesis; contributes to leukemogenesis	Zhang et al. (2001); Labuhn et al. (2019); Vara et al. (2020)
	NRAS	6/141 (4.3); 4/49 (8.2); 4/ 44 (9.1)	Oncogene; RAS signaling; regulates hematopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Gu et al. (2019); Labuhn et al. (2019); Shi et al. (2019); Panferova et al. (2021)
	PTPN11	1/49 (2.0)	Oncogene; RAS signaling; regulates hematopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Pandey et al. (2017)
Transcription factors	*CREBBP*	1/141 (0.7)	Tumor suppressor; transcriptional coactivator; lysine acetyltransferase enzyme; regulates hematopoiesis; contributes to leukemogenesis	Labuhn et al. (2019); Zhang et al. (2021)
	GATA2	1/44 (2.3)	Transcription factor; regulates early hematopoiesis (HSPC generation and function); contributes to leukemogenesis	Fujiwara (2017); Soukup and Bresnick (2020); Panferova et al. (2021)

(Continued on following page)

TABLE 2 (*Continued*) Recurrent somatic mutations reported in myeloid leukemia associated with Down syndrome.

Class	Mutant genes	Frequency of mutations in various studies n (%)	Function in hematopoiesis/ leukemogenesis; pathway to which it contributes	References
	MYC	1/141 (0.7); 1/44 (2.3)	Oncogene; transcription factor; regulates hematopoiesis; contributes to leukemogenesis	Labuhn et al. (2019); Benetatos et al. (2020); Panferova et al. (2021)
	RUNX1	3/141 (2.1)	Transcription factor; master-regulator of hematopoiesis; regulates megakaryopoiesis; contributes to leukemogenesis	Elagib et al. (2003); Yanagida et al. (2005); Banno et al. (2016); Gonzales et al. (2021); Panferova et al. (2021)
	TAL1	1/44 (2.3)	Oncogene; transcription factor; regulates HSC; contributes to leukemogenesis	Panferova et al. (2021); Thoms et al. (2021)
	TP53	5/141 (3.5); 3/49 (6.1); 2/44 (4.5); 2/11 (18.2)	Tumor suppressor; transcription factor; regulates hematopoiesis; contributes to leukemogenesis	Kiyoi et al. (2007); Yoshida et al. (2013); Labuhn et al. (2019); George et al. (2021); Panferova et al. (2021)
	WT1	1/141 (0.7); 2/49 (4.1); 1/44 (2.3)	Transcriptional activator or repressor; regulates hematopoiesis; contributes to leukemogenesis	Yoshida et al. (2013); Labuhn et al. (2019); Panferova et al. (2021); El Hussein et al. (2022)
Others	*CARD11*	1/44 (2.3)	Oncogene; TCR and BCR signaling; regulates hematopoiesis; contributes to leukemogenesis	Lu et al. (2021); Panferova et al. (2021)
	CHEK2	2/44 (4.5)	DNA damage response gene; contributes to leukemogenesis	Bazinet et al. (2021); Panferova et al. (2021); Singh et al. (2022)
	CSF2RB	7/141 (5.0)	Oncogene; JAK-STAT; PI3K-PKB- mTOR; MEK/ERK; regulates megakaryocytic proliferation and differentiation; contributes to leukemogenesis	Labuhn et al. (2019)
	CSF3R	1/44 (2.3)	Oncogene; JAK-STAT; regulates granulocyte progenitor differentiation; contributes to leukemogenesis	Maxson and Tyner (2017); Panferova et al. (2021)
	DCAF7	1/141 (0.7); 2/49 (4.1)	Scaffold protein or adaptor protein; interacts with ERCC1-XPF, DYRK1A, DYRK1B, MEKK1, and HIPK2	Yoshida et al. (2013); Yousefelahiyeh et al. (2018); Kawara et al. (2019); Labuhn et al. (2019)
	DLEC1	1/7 (14.3)	Tumor suppressor	Nikolaev et al. (2013); Hong et al. (2016)
	DHX29	1/7 (14.3)	RNA helicase; RNA co-sensor for anti-encephalomyocarditis virus immunity; regulates translation initiation	Nikolaev et al. (2013); Sweeney et al. (2021); Zhou et al. (2021)
	ETNK1	1/44 (2.3)	Kinase; involved in ethanolamine phosphorylation, ROS production, and DNA damage	Fontana et al. (2020); Panferova et al. (2021)
	PML	1/44 (2.3)	Tumor suppressor; regulates hematopoiesis; contributes to leukemogenesis	Haupt et al. (2013); Panferova et al. (2021)
	POLE	1/7 (14.3)	DNA replication; cancer-predisposing gene	Nikolaev et al. (2013); Magrin et al. (2021)
	PRPF40B	1/44 (2.3)	RNA splicing machinery; contributes to leukemogenesis	Lorenzini et al. (2019); Panferova et al. (2021)
	SF3B1	3/141 (2.1)	RNA splicing machinery; contributes to leukemogenesis	Labuhn et al. (2019); Van Der Werf et al. (2021)
	SRSF2	12/141 (8.5); 1/49 (2.0)	RNA splicing machinery; contributes to leukemogenesis	Hama et al. (2008); Yoshida et al. (2013); Labuhn et al. (2019); Todisco et al. (2021)
	WRN	1/44 (2.3)	Helicases; DNA replication and repair machinery; contributes to leukemogenesis	Moles et al. (2016); Panferova et al. (2021)

Mutations in the cohesin complex and related components

Cohesin is a multi-subunit complex composed of three main structural proteins (structural maintenance of chromosomes protein 1A (SMC1A), structural maintenance of chromosomes protein 3 (SMC3), and double-strand-break repair protein Rad21 homolog (RAD21)), which bind to either cohesin subunit SA-1 (STAG1) or cohesin subunit SA-2 (STAG2) proteins. Cohesin complex is a ring-shaped structure that surrounds chromosomal DNA and controls its functions, including sister chromatid cohesion, chromatin remodeling, transcriptional regulation, and DNA damage repair (Jann and Tothova, 2021). Nipped-B-like protein (NIPBL) is involved in

cohesin loading to chromatin, translocating cohesin along chromatin fibers, and regulating cohesin after loading (Garcia et al., 2021). Cohesin core subunits and its modulators (including STAG2, RAD21, SMC1A, SMC3 and NIPBL) are recurrently mutated in myeloid malignancies. Cohesin mutations are highly prevalent in ML-DS, where they occur in nearly half of patients (Yoshida et al., 2013). *STAG2* and *RAD21* have the higher mutation frequency than *SMC1*, *SMC3* and *NIPBL*, with approximately 9.1–18.4% and 11.3–22.4% in ML-DS cases respectively (Nikolaev et al., 2013; Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). Each of these mutations results in loss-of-function of the molecules, and in cooperation with *GATA1s* and trisomy 21 each can drive leukemic transformation.

Recent genetic modifications of human iPSC lines derived from DS tissue greatly assisted examination of the cooperativity between *GATA1s* and cohesin mutations in ML-DS. The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system was used to introduce *GATA1s* and *STAG2* mutations into iPSCs in a sequential manner (Barwe et al., 2020; Barwe et al., 2022). *GATA1s* and *STAG2* knockout cooperatively increased the megakaryocytic population and induced the ML-DS immunophenotype (Barwe et al., 2022). In another study, trisomic 21 iPSC line (Chou et al., 2012) was edited to introduce *GATA1s* followed by heterozygous inactivation of *SMC3* ($SMC3^{+/-}$) and then, introduction of a gain-of-function *MPL* mutation (MPLW515K) (Arkoun et al., 2022). It was found that *GATA1s* impaired megakaryocyte differentiation and that $SMC3^{+/-}$ enhanced this effect independent of trisomy 21. MPLW515K further increased the megakaryocyte output in this model, including through the induction of growth factor independence. Low expression of *NFE2* was critical for the induction of megakaryocyte dysplasia by *GATA1s* (Arkoun et al., 2022). These novel iPSC-based models are likely to rapidly advance our understanding of ML-DS pathogenesis, and assist therapy development.

We include a brief description of the relevant cohesin elements to highlight their roles in haematopoiesis. *STAG2*, located on the chromosome X, is the most frequently mutated cohesin gene in human cancer (Viny et al., 2019). *STAG2* deletion in hematopoietic stem and progenitor cells results in abnormal hematopoietic function, increased self-renewal, and impaired differentiation (Viny et al., 2019). *STAG2* loss-of-function decreases cell growth and proliferation, increases cell invasion and metastasis, enhances chemo-resistance, regulates the expression of many immune-related genes, and interplays with *RUNX1* deficiency to perturb chromatin looping (Nie et al., 2019; Ochi et al., 2020). Likewise, *RAD21* loss-of-function confers enhanced hematopoietic self-renewal and impaired cell differentiation (Fisher et al., 2017; Bisaillon et al., 2020). RAD21 is a regulator of gene expression and epigenetic modulation. For example, RAD21 regulates expression of *RUNX1* and methylation of Homeobox a7/Homeobox a9 (Fisher et al., 2017; Bisaillon et al., 2020). NIPBL also regulates *RUNX1* expression, thus its loss-of-function impairs *RUNX1* expression and consequently, hematopoiesis (Mazzola et al., 2020). NIPBL interplaying with nucleophosmin 1 (NPM1) regulates myeloid differentiation through the WNT (Wingless/Integrated) pathway; the disruption of these interactions has been implicated in leukemogenesis (Mazzola et al., 2019a). SMC3 is the cohesin ATPase subunit, with its dosage controlling embryogenesis and hematopoiesis (Wang et al., 2019; Rivas et al., 2021). Homozygous deletion of *SMC3* in mice results in embryonic lethality and the hematopoietic failure (Wang et al., 2019). In comparison, heterozygous *SMC3* deletion leads to developmental defects (e.g., abnormal craniofacial morphology), germinal center hyperplasia with increased B-cell proliferation and increased risk of B-cell lymphoma development (Rivas et al., 2021). SMC1A-R586W mutation is known to interfere with cohesin localization and cohesin-mediated DNA loop interaction in AML cells. This mutation confers wide changes to gene expression and genome organization when engineered into murine embryonic stem cells (Carico et al., 2021). Finally, CCCTC-binding factor (CTCF) is a tumor suppressor and involved in many cellular processes, with approximately 11.3–20.4% of ML-DS cases harboring *CTCF* gene mutations (Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). CTCF interacts with the cohesin complex to control genome architecture and gene expression (Zuin et al., 2014; Wang et al., 2020). CTCF and cohesins are known to assist formation of DNA loops, however depletion of either the cohesin complex or CTCF has differential effects on chromatin organization and gene expression in human HEK293T cells (Zuin et al., 2014). Deletion of cohesin caused a general loss of local chromatin interactions but the topological domains remained intact. In contrast, depletion of CTCF both reduced and increased interdomain interactions and distinct groups of genes became dysregulated. Apart from its interplay with cohesins, CTCF is a highly conserved transcription factor implicated in transcriptional activation and repression, insulation, formation of chromatin barrier, gene imprinting, X-chromosome inactivation, and RNA splicing (Bell et al., 1999; Bell and Felsenfeld, 2000; Xu et al., 2007; Shukla et al., 2011; Wang et al., 2020). CTCF is also involved in maintaining genomic methylation patterns through the control of poly (ADP-ribose) polymerase 1 (PARP1) and the activity of DNA methyltransferase 1 (DNMT1) (Zampieri et al., 2012). *CTCF* haploinsufficiency correlates with altered patterns of DNA methylation and predisposes to cancer in mice (Kemp et al., 2014). CTCF is a critical factor in the control of hematopoiesis and leukemogenesis (Torrano et al., 2005; Kim et al., 2017; Mujahed et al., 2020). In adult mice, conditional *CTCF* deletion causes an acute loss of HSCs, severe bone marrow failure and increased mortality, highlighting CTCF requirement for the maintenance of the HSC pool (Kim et al., 2017). Abnormal CTCF expression reduces growth and enhances

differentiation of the erythroid lineage by down-regulating MYC (Torrano et al., 2005). In AML, CTCF binding was shown to be elevated, compared with normal bone marrow, with increased CTCF binding in promoter regions linked with DNA hypomethylation and increased target gene expression (Mujahed et al., 2020). However, the combination of *CTCF* loss-of-function with *GATA1s* and trisomy 21 is unable to drive leukemic transformation, indicating additional events are required (Wagenblast et al., 2021). Collectively, the cohesin complex and CTCF are involved in ML-DS pathogenesis, but the exact roles of these molecules need further elucidation.

Mutations in signaling pathways

Most mutations affecting signaling pathways occur in genes encoding JAK regulators, *MPL* and *KIT* (CD117) collectively reported in 48% of ML-DS cases (Labuhn et al., 2019). *JAK2* and *JAK3* are more frequently mutated (9.9 and 13.5%) than *JAK1*, *MPL* and *KIT* (4.3, 7.1 and 1.4%) (Labuhn et al., 2019). *JAK1-3* variants are identified in both ML-DS and TAM samples, however, gain-of-function mutations are only detected in ML-DS, highlighting that aberrant activation of JAK-STAT signaling is important for transition to leukemia (Labuhn et al., 2019). The JAK family of tyrosine kinases (JAK1-3 and tyrosine kinase 2, TYK2) are pivotal mediators of growth factor and cytokine signaling, including downstream of thrombopoietin (TPO) and granulocyte-macrophage colony-stimulating factor (GM-CSF) (De Castro et al., 2021; Moser et al., 2021). JAK1, JAK2 and TYK2, are ubiquitously expressed, whilst JAK3 is predominantly expressed in lymphoid and myeloid cells. *JAK3* mutations are more common than of other members of JAK family in ML-DS (Labuhn et al., 2019). Under physiological conditions, JAK-STAT signaling is tightly controlled and involved in a wide range of fundamental biological processes, including cell proliferation, differentiation, apoptosis, inflammation, and blood production (Park et al., 2016; De Castro et al., 2021; Moser et al., 2021). Normal megakaryopoiesis requires TPO-mediated STAT5 activation. Unphosphorylated STAT5 represses megakaryocytic transcriptional program and inhibits megakaryocytic differentiation by competing with ERG for CTCF binding, which can be reversed by TPO-mediated activation of STAT5 (Park et al., 2016). MPL, a receptor of TPO, is also frequently mutated in ML-DS, which contributes to leukemia development. In the presence of trisomy 21 and GATA1s, MPL W515L causes rapid and lethal leukemia in mice (Malinge et al., 2012). Recently, colony stimulating factor 2 receptor subunit beta (CSF2RB) A455D variant was reported in almost 5% of ML-DS children. This variant is mutually exclusive with mutated *JAK1-3*, *MPL* or *RAS* genes, and causes ligand-independent STAT5 activation promoting cytokine-independent cell growth (Labuhn et al., 2019). Upon introduction of the CSF2RB A455D mutant into hematopoietic stem and progenitor cells (HSPCs), megakaryocytic and erythroid proliferation is enhanced, and terminal megakaryocytic maturation is blocked (Labuhn et al., 2019). These alterations are alleviated by the JAK1/2 inhibitor ruxolitinib, emphasizing that aberrant JAK-STAT signaling participates in the CSF2RB A455D-driven leukemogenesis (Labuhn et al., 2019). In addition, CSF2RB binding to FLT3-ITD is found in other AML cell lines and patient cells where CSF2RB deletion decreases STAT5 phosphorylation, inhibits leukemic cell proliferation, and sensitizes cells to FLT3 inhibition (Charlet et al., 2021). These findings demonstrate that CSF2RB is critical for FLT3-ITD-dependent oncogenic signaling and transformation, but its role in ML-DS requires further study.

Mutations in *RAS* (Rat sarcoma virus) gene family members, such as *KRAS*, *NF1*, *NRAS*, and *PTPN11*, are found in 14% of ML-DS samples (Labuhn et al., 2019). *NRAS* and *KRAS* variants are the most common accounting for 4.5–8.2% and 4.3–9.1% of ML-DS cases respectively (Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). Ras belongs to the small GTPase family that binds to guanosine triphosphate (GTP) and hydrolyses it to guanosine diphosphate (GDP), with three distinct isoforms NRas, KRas, and HRas (Padmakumar et al., 2021). Ras is located on the inner surface of the plasma membrane, and acts as a binary molecular switch. Ras can transmit extracellular signals to the nucleus, and cycles between the inactive GDP-bound state and the active GTP-bound state (Zafar et al., 2021). Mutations fix RAS-GTPase proteins in an active GTP-bound state, resulting in constitutive activation of MAPK and PI3K (phosphoinositide 3-kinases) signaling. Consequently, uncontrolled cell proliferation and survival occur in mutated cells. In mouse models, clonal NRAS/KRAS activation increases cell growth, proliferation, and colony formation through a lysine methyltransferase 2A (KMT2A)- polo like kinase 1 (PLK1) axis (Carr et al., 2021). Mutations in *RAS* have a role in TAM progression to ML-DS, but it is not fully understood how these mutations cooperate with trisomy 21, *GATA1s* and other mutations in cohesins or epigenetic modulators.

Mutations in epigenetic regulators

Loss-of-function mutations in epigenetic regulators are emerging as critical contributors to ML-DS progression. Such mutations are reported in approximately 36–45% of ML-DS samples and affect a range of regulators, including additional sex combs-like 1 (*ASXL1*), BCL6 corepressor (*BCOR*), *DNMT1*, *DNMT3A*, embryonic ectoderm development (*EED*), E1A binding protein P300 (*EP300*), *EZH2*, KAT8 regulatory NSL complex subunit 1 (*KANSL1*), lysine demethylase 6A (*KDM6A*), lysine methyltransferase 2C (*KMT2C*), N-acetyltransferase 6 (*NAT6*), *SUZ12*, and tet methylcytosine dioxygenase 2 (*TET2*)

TABLE 3 Chromosomal abnormalities reported in myeloid leukemia associated with Down syndrome.

Class	Cytogenetic alteration	Frequency of alterations in various studies n (%)	References
Whole chromosome gain	Trisomy 2	1/141 (0.7)	Labuhn et al. (2019)
	Trisomy 8	9/141 (6.4); 1/7 (14.3); 4/24 (16.7)	Hama et al. (2008); Malinge et al. (2008); Labuhn et l. (2019)
	Trisomy 11	1/141 (0.7); 2/24 (8.3)	Hama et al. (2008); Labuhn et al. (2019)
	Trisomy 13	3/141 (2.1)	Labuhn et al. (2019)
	Trisomy 14	3/141 (2.1)	Labuhn et al. (2019)
	Tetrasomy 14	1/7 (14.3)	Nikolaev et al. (2013)
	Trisomy 19	2/141 (1.4); 1/24 (4.2)	Hama et al. (2008); Labuhn et al. (2019)
	Tetrasomy 21	9/141 (6.4); 1/7 (14.3); 1/24 (4.2); 1/7 (14.3)	Hama et al. (2008); Malinge et al. (2008); Nikolaev et al. (2013); Labuhn et al. (2019)
	Trisomy 22	1/141 (0.7); 1/24 (4.2)	Hama et al. (2008); Labuhn et al. (2019)
Chromosomal arm gain	add(1q)	4/141 (2.8)	Labuhn et al. (2019)
	add(5p)	1/24 (4.2)	Hama et al. (2008)
	add(5q)	1/24 (4.2)	Hama et al. (2008)
	add(6p)	1/141 (0.7)	Labuhn et al. (2019)
	add(6q)	1/141 (0.7)	Labuhn et al. (2019)
	add(7p)	2/24 (8.3)	Hama et al. (2008)
	add(7q)	2/141 (1.4)	Labuhn et al. (2019)
	add(8p)	1/141 (0.7)	Labuhn et al. (2019)
	add(11q)	1/141 (0.7)	Labuhn et al. (2019)
	add(16q)	2/141 (1.4)	Labuhn et al. (2019)
	add(19p)	1/141 (0.7); 1/24 (4.2)	Hama et al. (2008); Labuhn et al. (2019)
	add(22q)	1/24 (4.2)	Hama et al. (2008)
Whole chromosome loss	−1	1/24 (4.2)	Hama et al. (2008)
	−3	1/24 (4.2)	Hama et al. (2008)
	−4	1/7 (14.3)	Malinge et al. (2008)
	−5	1/24 (4.2)	Hama et al. (2008)
	−7	5/24 (20.8)	Hama et al. (2008)
	−9	1/24 (4.2); 1/7 (14.3)	Hama et al. (2008); Malinge et al. (2008)
	−16	1/7 (14.3)	Malinge et al. (2008)
	−18	1/24 (4.2)	Hama et al. (2008)
Chromosomal arm loss	del(5p)	1/141 (0.7)	Labuhn et al. (2019)
	del(5q)	3/141 (2.1); 1/7 (14.3)	Nikolaev et al. (2013); Labuhn et al. (2019)
	del(6q)	2/141 (1.4); 1/7 (14.3); 1/24 (4.2)	Hama et al. (2008); Nikolaev et al. (2013); Labuhn et al. (2019)
	del(7p)	5/141 (3.5);	Labuhn et al. (2019)
	del(7q)	2/141 (1.4); 1/24 (4.2)	Hama et al. (2008); Labuhn et al. (2019)
	del(8q)	1/7 (14.3)	Nikolaev et al. (2013)
	del(11p)	2/141 (1.4); 1/24 (4.2)	Hama et al. (2008); Labuhn et al. (2019)
	del(11q)	1/24 (4.2)	Hama et al. (2008)
	del(13q)	2/141 (1.4)	Labuhn et al. (2019)
	del(16q)	6/141 (4.3)	Labuhn et al. (2019)
	del(17p)	3/141 (2.1)	Labuhn et al. (2019)
	del(17q)	3/141 (2.1)	Labuhn et al. (2019)
	del(20q)	1/24 (4.2)	Hama et al. (2008)
	del(22q)	1/141 (0.7)	Labuhn et al. (2019)

(Continued on following page)

TABLE 3 (*Continued*) Chromosomal abnormalities reported in myeloid leukemia associated with Down syndrome.

Class	Cytogenetic alteration	Frequency of alterations in various studies n (%)	References
Other changes	+der(1)t(1; ?)	1/24 (4.2)	Hama et al. (2008)
	der(3)t(3;3) (p25;p10)	1/24 (4.2)	Hama et al. (2008)
	+der(5)t(5;7)	1/24 (4.2)	Hama et al. (2008)
	der(7)t(1;7) (q23;q36)	1/24 (4.2)	Hama et al. (2008)
	der(14) t(1;14) (q24~25;p11)	1/141 (0.7)	Labuhn et al. (2019)
	der(17)t(1;17) (q25;q25)	1/24 (4.2)	Hama et al. (2008)
	der(21) (qter- > q22.1::p11.2- > qter)	1/141 (0.7)	Labuhn et al. (2019)
	der(X)t(X;1) (q28;q25)	1/24 (4.2)	Hama et al. (2008)
	inv (9) (p11;q12)	1/7 (14.3)	Malinge et al. (2008)
	isochromosome (7q)	1/7 (14.3); 1/24 (4.2)	Hama et al. (2008); Nikolaev et al. (2013)
	random aberrations	2/7 (28.6)	Malinge et al. (2008)
	t(3;17) (q25;q25)	1/141 (0.7)	Hama et al. (2008)
	t(5;12) (p15;q21)	1/24 (4.2)	Hama et al. (2008)

(Nikolaev et al., 2013; Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). Mutations in *KANSL1*, *EZH2* and *SUZ12* were seen at the highest frequency, in 6.1–12.1%, 2.3–32.7% and 2–6.4% of ML-DS cases respectively (Nikolaev et al., 2013; Yoshida et al., 2013; Labuhn et al., 2019; Panferova et al., 2021). *KANSL1* is essential for the activity of the histone acetylation complex, which takes part in the acetylation of histone H4 lysine 16 and eventually leads to transcriptional activation. Loss-of-function mutations in *KANSL1* are detected in both ML-DS and non-DS-AMKL (Yoshida et al., 2013; Labuhn et al., 2019). *KANSL1* mutations combined with trisomy 21 and *GATA1s* drive leukemic engraftment in mice (Wagenblast et al., 2021). EZH2 forms polycomb repressive complex 2 (PRC2) together with SUZ12, EED and RB binding protein 4 (RBBP4). PRC2 is mainly responsible for the methylation of lysine 27 in the tail of histone H3 family proteins (H3K27me3), which subsequently silences its target gene expression. Thus, EZH2 is a transcriptional repressor with methyltransferase activity, whereas SUZ12 is essential for the structural integrity of PRC2 and the facilitation of chromatin binding (Chen et al., 2018; Zeisig and So, 2021). EZH2 is unable to perform this enzymatic function alone, and the interplay with EED and SUZ12 enables PRC2 function (Chen et al., 2018; Zeisig and So, 2021). In megakaryopoiesis, EZH2 inhibition accelerates megakaryocytic differentiation and blocks megakaryocytic proliferation (Mazzi et al., 2021). EZH2 and SUZ12 act as tumor suppressors; mutations in either gene lead to loss-of-function of PRC2 core subunits and a deficit of H3K27me3 (Ntziachristos et al., 2012). Murine ML-DS leukemia models and ML-DS patients show loss-of-function mutations in EZH2 and other PRC2 members, supporting the PRC2 role in transition from TAM to ML-DS (Labuhn et al., 2019). Although the importance of mutated epigenetic modifiers in ML-DS has been recognized, their pathologic functions and clinical impact remain unclear.

Chromosomal abnormalities

Beyond trisomy 21, additional cytogenetic changes are observed in the majority of children with ML-DS, but rarely in TAM (Malinge et al., 2008; Labuhn et al., 2019). Therefore, these changes could play a role in the development of ML-DS. Cytogenetic changes reported in ML-DS include gains and losses of whole chromosomes or their arms, or chromosomal rearrangements (Table 3). Common chromosomal gains are trisomies: +2, +8, +11, +13, +14, +19, +22, or tetrasomies: +14, +21. Chromosomal losses include monosomies: −1, −3, −4, −5, −7, −9, −16, and −18. Other aberrations include: add(1q), add(6p), add (6q), add (7q), add(8p), add(11q), add (16q), add (19p); or deletions: del(5p), del(5q), del(6q), del(7p), del(7q), del(11p), del(13q), del(16q), del(17p), del(17q), and del(22q). The most common structural abnormalities are del(7p)/del(7q)/−7, del(16q), trisomy 8, and tetrasomy 21. However, none of these changes offer clear insights into the molecular pathogenesis of ML-DS, and their prognostic impact is also largely unknown (Forestier et al., 2008; De Souza et al., 2017; Labuhn et al., 2019). One recent study points out that +8 can be associated with inferior event-free survival in ML-DS (Uffmann et al., 2017). More work is required to elucidate the pathogenetic role and clinical impact of chromosomal abnormalities in ML-DS.

Co-occurrence patterns of additional somatic mutations other than GATA1s

Transformation of TAM to ML-DS often arises on the background of activating signaling mutations interacting with deregulated epigenetic modifiers. For instance, there is a

significant co-occurrence of variants in genes encoding tyrosine kinases and RAS proteins with variants in epigenetic regulators or cohesins both in ML-DS mouse models and patient samples (Labuhn et al., 2019). *CTCF* and *EZH2* mutations alone are insufficient to drive ML-DS in the presence of trisomy 21 and *GATA1s*, implying other somatic mutations are required (Wagenblast et al., 2021). The frequent co-occurrence of variants in *EZH2* and *CB1* is identified in a murine model of ML-DS, while *NF1* mutations appear mutually exclusive with *CB1*, *EZH2*, and *CTCF* variants (Labuhn et al., 2019). Co-occurrence of additional mutations is important for leukemic progression, but their patterns, functional effects, and clinical significance need further investigation.

Mutational landscape of relapsed myeloid leukemia associated with Down syndrome

ML-DS usually has a low incidence of relapse, seen in approximately 5–6% of patients in developed countries, mostly because the initial disease is very sensitive to chemotherapy (Uffmann et al., 2017). However, when relapse occurs in ML-DS patients, the prognosis is less favorable. Little is known about the molecular underpinnings leading to relapse and current treatment options are less effective in relapsed patients. One study showed that in a cohort of 170 pediatric patients with ML-DS, five of 7 relapsed cases harbored trisomy 8, while the other two carried isochromosome 7 and additional material on chromosome 16 respectively (Uffmann et al., 2017). As for somatic mutations, the sequencing data from one paired sample (diagnostic and relapsed) demonstrated the presence of EZH2 F562S, JAK2 V617F and MTNR1B R316H in the relapsed sample but not at the time of diagnosis, while SMC1A R711Q, MPL W515S, JAK2 F694S and EZH2 H206fs were detected at both time-points (Labuhn et al., 2019). A lot more work will need to be done in this area in the future.

Novel therapeutic targets

As the genomic, epigenomic and transcriptomic changes are uncovered in TAM and ML-DS, new molecular targets for prevention and treatment are being proposed. Mutations in signaling effectors are one of the most frequent events in ML-DS associated with the overactivation of pathways such as JAK-STAT, RAS/MEK/ERK and PI3K/PKB (Labuhn et al., 2019). Inhibition of these pathways may help treat ML-DS. FDA-approved JAK1/2 inhibitors, ruxolitinib and momelotinib (Sureau et al., 2021), could be considered for patients with activating JAK-STAT mutations. Similarly, drugs targeting RAS and PI3K/PKB signaling could be trialed in children with mutations in these pathways (De Castro et al., 2021). CD117/KIT expression is a marker of *GATA1s*-induced pre-leukemia- and *GATA1s/STAG2*-knock-out-induced leukemia-initiating cells. The maintenance and expansion of those cells rely on KIT signaling; thus, KIT inhibitors have emerged as potential therapeutic targets (Wagenblast et al., 2021). Further, mutations in cohesin subunits and cohesin regulators are crucial for ML-DS pathogenesis (Labuhn et al., 2019), thus targeting cohesin-mutant cells has been suggested to be a new therapeutic strategy. There are three distinct approaches through which cohesin mutated cells can be targeted: 1) direct modulation of cohesin subunits and its regulators; 2) targeting cohesin-induced deregulated signaling; and 3) targeting altered DNA damage repair mechanisms (Antony et al., 2021). STAG1 inhibition may be a suitable therapy for patients with *STAG2* mutations because it is synthetically lethal with *STAG2* variants. ML-DS displays frequent gains of DNA methylation, thus epigenetic therapies may be useful. In support, lysine-specific demethylase inhibitor T-3775440 inhibits growth of patient-derived blasts *ex vivo* (Labuhn et al., 2019). Finally, three HSA21 miRNAs (miR-99a, miR-125b and miR-155) are overexpressed in blast cells from ML-DS, and their blockage inhibits *GATA1s*-induced pre-leukemia development (Wagenblast et al., 2021). Hence, miRNAs could also become potential therapeutic targets in the future.

Conclusions and future directions

ML-DS has three major molecular features: trisomy 21, *GATA1s* mutations, and tertiary alterations (Figure 1). Trisomy 21 drives megakaryocytic expansion through the increased gene dosage effect, but trisomy 21 may not be required for progression to ML-DS. *GATA1s* mutations are acquired during fetal liver hematopoiesis in susceptible HSCs with high proliferative potential, which leads to abnormal megakaryocytic proliferation and impaired erythroid differentiation. *GATA1s* effects begin in an immunophenotypically distinct population of fetal erythro-megakaryocytic cells. The development of ML-DS requires acquisition and selection of clones with additional somatic mutations and chromosomal structural abnormalities. Substantial progress has been made over the last 20 years in the molecular characterization of ML-DS, but some important questions remain unanswered. How do trisomy 21, *GATA1s*, additional somatic mutations and chromosomal alterations cooperate to drive ML-DS? In the context of trisomy 21 and *GATA1s*, what is the relevant co-occurrence pattern of somatic mutations and cytogenetic changes? What is the biological role and the clinical impact of such changes? Recent application of CRISPR/Cas9 technology in iPSC-based models of ML-DS started to provide some essential answers to these questions.

From the clinical standpoint, new therapies are needed for children with refractory and relapsed disease, in particular as

high-dose chemotherapy causes unacceptable toxicity in DS children. To test emerging therapeutic targets, we need to advance pre-clinical disease models of ML-DS. Chromosome 21-encoded proteins and miRNAs are important players in ML-DS. Is it possible to target these molecules alone or do we need to simultaneously target secondary and tertiary genetic changes to control leukemia growth? The work needs to continue to better elucidate disease mechanisms and to develop more effective therapies.

Author contributions

JL drafted the manuscript. MK-Z provided supervision and guidance, helped write and revise the manuscript. Both authors approved the submitted version.

Funding

Anne Norman and Victoria Nicholls Leukaemia and Lymphoma Research Fund.

Acknowledgments

Special thanks to Taryn Green for her careful proof-reading and technical support.

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

Ahlfors, H., Anyanwu, N., Pakanavicius, E., Dinischiotu, N., Lana-Elola, E., Watson-Scales, S., et al. (2019). Gene expression dysregulation domains are not a specific feature of Down syndrome. *Nat. Commun.* 10 (1), 2489. doi:10.1038/s41467-019-10129-9

Aksu, T., Gumruk, F., and Unal, S. (2020). Comment on: Clinical, cytogenetic, and molecular analyses of 17 neonates with transient abnormal myelopoiesis and nonconstitutional trisomy 21. *Pediatr. Blood Cancer* 67 (11), e28289. doi:10.1002/pbc.28289

Alejo-Valle, O., Weigert, K., Bhayadia, R., Ng, M., Issa, H., Beyer, C., et al. (2021). The megakaryocytic transcription factor ARID3A suppresses leukemia pathogenesis. *Blood* 139, 651–665. doi:10.1182/blood.2021012231

Alford, K. A., Reinhardt, K., Garnett, C., Norton, A., Böhmer, K., von Neuhoff, C., et al. (2011). Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. *Blood* 118 (8), 2222–2238. doi:10.1182/blood-2011-03-342774

Antonarakis, S. E. (2017). Down syndrome and the complexity of genome dosage imbalance. *Nat. Rev. Genet.* 18 (3), 147–163. doi:10.1038/nrg.2016.154

Antony, J., Chin, C. V., and Horsfield, J. A. (2021). Cohesin mutations in cancer: Emerging therapeutic targets. *Int. J. Mol. Sci.* 22 (13), 6788. doi:10.3390/ijms22136788

Apollonsky, N., Shende, A., Ouansafi, I., Brody, J., Atlas, M., Aygun, B., et al. (2008). Transient myeloproliferative disorder in neonates with and without down syndrome: A tale of 2 syndromes. *J. Pediatr. Hematol. Oncol.* 30 (11), 860–864. doi:10.1097/MPH.0b013e31818a953e

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., et al. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127 (20), 2391–2405. doi:10.1182/blood-2016-03-643544

Arkoun, B., Robert, E., Boudia, F., Mazzi, S., Dufour, V., Siret, A., et al. (2022). Stepwise GATA1 and SMC3 mutations alter megakaryocyte differentiation in a Down syndrome leukemia model. *J. Clin. Invest.* 1, e156290. doi:10.1172/JCI156290

Banno, K., Omori, S., Hirata, K., Nawa, N., Nakagawa, N., Nishimura, K., et al. (2016). Systematic cellular disease models reveal synergistic interaction of trisomy 21 and GATA1 mutations in hematopoietic abnormalities. *Cell Rep.* 15 (6), 1228–1241. doi:10.1016/j.celrep.2016.04.031

Barwe, S. P., Sebastian, A., Sidhu, I., Kolb, E. A., and Gopalakrishnapillai, A. (2022). Modeling down syndrome myeloid leukemia by sequential introduction of GATA1 and STAG2 mutations in induced pluripotent stem cells with trisomy 21. *Cells* 11 (4), 628. doi:10.3390/cells11040628

Barwe, S. P., Sidhu, I., Kolb, E. A., and Gopalakrishnapillai, A. (2020). Modeling transient abnormal myelopoiesis using induced pluripotent stem cells and CRISPR/Cas9 technology. *Mol. Ther. Methods Clin. Dev.* 19, 201–209. doi:10.1016/j.omtm.2020.09.007

Bazinet, A., Heath, J., Chong, A. S., Simo-Cheyou, E. R., Worme, S., Rivera Polo, B., et al. (2021). Common clonal origin of chronic myelomonocytic leukemia and B-cell acute lymphoblastic leukemia in a patient with a germline CHEK2 variant. *Cold Spring Harb. Mol. Case Stud.* 7 (3), a006090. doi:10.1101/mcs.a006090

Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405 (6785), 482–485. doi:10.1038/35013100

Bell, A. C., West, A. G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98 (3), 387–396. doi:10.1016/s0092-8674(00)81967-4

Belmonte, R. L., Engbretson, I. L., Kim, J. H., Cajias, I., Ahn, E. E., Stachura, D. L., et al. (2021). Son is necessary for proper vertebrate blood development. *PLoS One* 16 (2), e0247489. doi:10.1371/journal.pone.0247489

Benetatos, L., Benetatou, A., and Vartholomatos, G. (2020). Enhancers and MYC interplay in hematopoiesis. *J. Mol. Med.* 98 (4), 471–481. doi:10.1007/s00109-020-01891-1

Bhatnagar, N., Nizery, L., Tunstall, O., Vyas, P., and Roberts, I. (2016). Transient abnormal myelopoiesis and AML in down syndrome: An update. *Curr. Hematol. Malig. Rep.* 11 (5), 333–341. doi:10.1007/s11899-016-0338-x

Bisaillon, R., Moison, C., Thiollier, C., Krosl, J., Bordeleau, M. E., Lehnertz, B., et al. (2020). Genetic characterization of ABT-199 sensitivity in human AML. *Leukemia* 34 (1), 63–74. doi:10.1038/s41375-019-0485-x

Bombery, M., and Vergilio, J. A. (2014). Transient abnormal myelopoiesis in neonates: GATA get the diagnosis. *Arch. Pathol. Lab. Med.* 138 (10), 1302–1306. doi:10.5858/arpa.2014-0304-CC

Bourquin, J. P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., et al. (2006). Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proc. Natl. Acad. Sci. U. S. A.* 103 (9), 3339–3344. doi:10.1073/pnas.0511150103

Brás, A., Rodrigues, A. S., Gomes, B., and Rueff, J. (2018). Down syndrome and microRNAs. *Biomed. Rep.* 8 (1), 11–16. doi:10.3892/br.2017.1019

Cabal-Hierro, L., van Galen, P., Prado, M. A., Higby, K. J., Togami, K., Mowery, C. T., et al. (2020). Chromatin accessibility promotes hematopoietic and leukemia stem cell activity. *Nat. Commun.* 11 (1), 1406. doi:10.1038/s41467-020-15221-z

Carico, Z. M., Stefan, H. C., Justice, M., Yimit, A., and Dowen, J. M. (2021). A cohesin cancer mutation reveals a role for the hinge domain in genome organization and gene expression. *PLoS Genet.* 17 (3), e1009435. doi:10.1371/journal.pgen.1009435

Carmichael, C. L., Metcalf, D., Henley, K. J., Kruse, E. A., Di Rago, L., Mifsud, S., et al. (2012). Hematopoietic overexpression of the transcription factor Erg induces lymphoid and erythro-megakaryocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 109 (38), 15437–15442. doi:10.1073/pnas.1213454109

Carr, R. M., Vorobyev, D., Lasho, T., Marks, D. L., Tolosa, E. J., Vedder, A., et al. (2021). RAS mutations drive proliferative chronic myelomonocytic leukemia via a KMT2A-PLK1 axis. *Nat. Commun.* 12 (1), 2901. doi:10.1038/s41467-021-23186-w

Charlet, A., Kappenstein, M., Keye, P., Kläsener, K., Endres, C., Poggio, T., et al. (2021). The IL-3, IL-5, and GM-CSF common receptor beta chain mediates oncogenic activity of FLT3-ITD-positive AML. *Leukemia* 36, 701–711. doi:10.1038/s41375-021-01462-4

Chattopadhyaya, S., and Ghosal, S. (2022). DNA methylation: A saga of genome maintenance in hematological perspective. *Hum. Cell* 35 (2), 448–461. doi:10.1007/s13577-022-00674-9

Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique structural platforms of Suz12 dictate distinct classes of PRC2 for chromatin binding. *Mol. Cell* 69 (5), 840–852. e845. doi:10.1016/j.molcel.2018.01.039

Chlon, T. M., McNulty, M., Goldenson, B., Rosinski, A., and Crispino, J. D. (2015). Global transcriptome and chromatin occupancy analysis reveal the short isoform of GATA1 is deficient for erythroid specification and gene expression. *Haematologica* 100 (5), 575–584. doi:10.3324/haematol.2014.112714

Chou, S. T., Byrska-Bishop, M., Tober, J. M., Yao, Y., Vandorn, D., Opalinska, J. B., et al. (2012). Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 109 (43), 17573–17578. doi:10.1073/pnas.1211175109

De Castro, C. P. M., Cadefau, M., and Cuartero, S. (2021). The mutational landscape of myeloid leukaemia in down syndrome. *Cancers (Basel)* 13 (16), 4144. doi:10.3390/cancers13164144

De Rooij, J. D., Branstetter, C., Ma, J., Li, Y., Walsh, M. P., Cheng, J., et al. (2017). Pediatric non-Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nat. Genet.* 49 (3), 451–456. doi:10.1038/ng.3772

De Souza, D. C., de Figueiredo, A. F., Ney Garcia, D. R., da Costa, E. S., Othman, M. A. K., Liehr, T., et al. (2017). A unique set of complex chromosomal abnormalities in an infant with myeloid leukemia associated with Down syndrome. *Mol. Cytogenet.* 10, 35. doi:10.1186/s13039-017-0335-3

El Hussein, S., DiNardo, C. D., Takahashi, K., Khoury, J. D., Fang, H., Furudate, K., et al. (2022). Acquired WT1 mutations contribute to relapse of NPM1-mutated acute myeloid leukemia following allogeneic hematopoietic stem cell transplant. *Bone Marrow Transpl.* 57, 370–376. doi:10.1038/s41409-021-01538-w

Elagib, K. E., Racke, F. K., Mogass, M., Khetawat, R., Delehanty, L. L., Goldfarb, A. N., et al. (2003). RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* 101 (11), 4333–4341. doi:10.1182/blood-2002-09-2708

Elton, T. S., Sansom, S. E., and Martin, M. M. (2010). Trisomy-21 gene dosage over-expression of miRNAs results in the haploinsufficiency of specific target proteins. *RNA Biol.* 7 (5), 540–547. doi:10.4161/rna.7.5.12685

Emmrich, S., Rasche, M., Schöning, J., Reimer, C., Keihani, S., Maroz, A., et al. (2014). miR-99a/100~125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGFβ and Wnt signaling. *Genes Dev.* 28 (8), 858–874. doi:10.1101/gad.233791.113

Fasouli, E. S., and Katsantoni, E. (2021). JAK-STAT in early hematopoiesis and leukemia. *Front. Cell Dev. Biol.* 9, 669363. doi:10.3389/fcell.2021.669363

Fisher, J. B., Peterson, J., Reimer, M., Stelloh, C., Pulakanti, K., Gerbec, Z. J., et al. (2017). The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. *Leukemia* 31 (3), 712–719. doi:10.1038/leu.2016.240

Flasinski, M., Scheibke, K., Zimmermann, M., Creutzig, U., Reinhardt, K., Verwer, F., et al. (2018). Low-dose cytarabine to prevent myeloid leukemia in children with down syndrome: TMD prevention 2007 study. *Blood Adv.* 2 (13), 1532–1540. doi:10.1182/bloodadvances.2018018945

Fontana, D., Mauri, M., Renso, R., Docci, M., Crespiatico, I., Røst, L. M., et al. (2020). ETNK1 mutations induce a mutator phenotype that can be reverted with phosphoethanolamine. *Nat. Commun.* 11 (1), 5938. doi:10.1038/s41467-020-19721-w

Forestier, E., Izraeli, S., Beverloo, B., Haas, O., Pession, A., Michalová, K., et al. (2008). Cytogenetic features of acute lymphoblastic and myeloid leukemias in pediatric patients with down syndrome: An iBFM-SG study. *Blood* 111 (3), 1575–1583. doi:10.1182/blood-2007-09-114231

Fujiwara, T. (2017). GATA transcription factors: Basic principles and related human disorders. *Tohoku J. Exp. Med.* 242 (2), 83–91. doi:10.1620/tjem.242.83

Gamis, A. S., Woods, W. G., Alonzo, T. A., Buxton, A., Lange, B., Barnard, D. R., et al. (2003). Increased age at diagnosis has a significantly negative effect on outcome in children with down syndrome and acute myeloid leukemia: A report from the Children's cancer group study 2891. *J. Clin. Oncol.* 21 (18), 3415–3422. doi:10.1200/JCO.2003.08.060

Garcia, P., Fernandez-Hernandez, R., Cuadrado, A., Coca, I., Gomez, A., Maqueda, M., et al. (2021). Disruption of NIPBL/Scc2 in Cornelia de Lange Syndrome provokes cohesin genome-wide redistribution with an impact in the transcriptome. *Nat. Commun.* 12 (1), 4551. doi:10.1038/s41467-021-24808-z

Garnett, C., Cruz Hernandez, D., and Vyas, P. (2020). GATA1 and cooperating mutations in myeloid leukaemia of Down syndrome. *IUBMB Life* 72 (1), 119–130. doi:10.1002/iub.2197

Ge, Y., LaFiura, K. M., Dombkowski, A. A., Chen, Q., Payton, S. G., Buck, S. A., et al. (2008). The role of the proto-oncogene ETS2 in acute megakaryocytic leukemia biology and therapy. *Leukemia* 22 (3), 521–529. doi:10.1038/sj.leu.2405066

George, B., Kantarjian, H., Baran, N., Krocker, J. D., and Rios, A. (2021). TP53 in acute myeloid leukemia: Molecular aspects and patterns of mutation. *Int. J. Mol. Sci.* 22 (19), 10782. doi:10.3390/ijms221910782

Gilliland, D. G., and Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. *Blood* 100 (5), 1532–1542. doi:10.1182/blood-2002-02-0492

Gonzales, F., Barthélémy, A., Peyrouze, P., Fenwarth, L., Preudhomme, C., Duployez, N., et al. (2021). Targeting RUNX1 in acute myeloid leukemia: Preclinical innovations and therapeutic implications. *Expert Opin. Ther. Targets* 25 (4), 299–309. doi:10.1080/14728222.2021.1915991

Grafone, T., Palmisano, M., Nicci, C., and Storti, S. (2012). An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: Biology and treatment. *Oncol. Rev.* 6 (1), e8. doi:10.4081/oncol.2012.e8

Grimm, J., Heckl, D., and Klusmann, J. H. (2021). Molecular mechanisms of the genetic predisposition to acute megakaryoblastic leukemia in infants with down syndrome. *Front. Oncol.* 11, 636633. doi:10.3389/fonc.2021.636633

Gu, Z., Liu, Y., Cai, F., Patrick, M., Zmajkovic, J., Cao, H., et al. (2019). Loss of EZH2 reprograms BCAA metabolism to drive leukemic transformation. *Cancer Discov.* 9 (9), 1228–1247. doi:10.1158/2159-8290.CD-19-0152

Hama, A., Yagasaki, H., Takahashi, Y., Nishio, N., Muramatsu, H., Yoshida, N., et al. (2008). Acute megakaryoblastic leukaemia (AMKL) in children: A comparison of AMKL with and without down syndrome. *Br. J. Haematol.* 140 (5), 552–561. doi:10.1111/j.1365-2141.2007.06971.x

Hamarsheh, S., Osswald, L., Saller, B. S., Unger, S., De Feo, D., Vinnakota, J. M., et al. (2020). Oncogenic Kras(G12D) causes myeloproliferation via NLRP3 inflammasome activation. *Nat. Commun.* 11 (1), 1659. doi:10.1038/s41467-020-15497-1

Hasle, H., Abrahamsson, J., Arola, M., Karow, A., O'Marcaigh, A., Reinhardt, D., et al. (2008). Myeloid leukemia in children 4 years or older with Down syndrome often lacks GATA1 mutation and cytogenetics and risk of relapse are more akin to sporadic AML. *Leukemia* 22 (7), 1428–1430. doi:10.1038/sj.leu.2405060

Hasle, H., Kline, R. M., Kjeldsen, E., Nik-Abdul-Rashid, N. F., Bhojwani, D., Verboon, J. M., et al. (2022). Germline GATA1s-generating mutations predispose to leukemia with acquired trisomy 21 and Down syndrome-like phenotype. *Blood* 139 (21), 3159–3165. doi:10.1182/blood.2021011463

Haupt, S., Mitchell, C., Corneille, V., Shortt, J., Fox, S., Pandolfi, P. P., et al. (2013). Loss of PML cooperates with mutant p53 to drive more aggressive cancers in a gender-dependent manner. *Cell Cycle* 12 (11), 1722–1731. doi:10.4161/cc.24805

Hitzler, J. K., Cheung, J., Li, Y., Scherer, S. W., and Zipursky, A. (2003). GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome. *Blood* 101 (11), 4301–4304. doi:10.1182/blood-2003-01-0013

Hong, Q., Li, Y., Chen, X., Ye, H., Tang, L., Zhou, A., et al. (2016). CDKN2B, SLC19A3 and DLEC1 promoter methylation alterations in the bone marrow of patients with acute myeloid leukemia during chemotherapy. *Exp. Ther. Med.* 11 (5), 1901–1907. doi:10.3892/etm.2016.3092

Ikeda, K., Ueda, T., Yamasaki, N., Nakata, Y., Sera, Y., Nagamachi, A., et al. (2016). Maintenance of the functional integrity of mouse hematopoiesis by EED and promotion of leukemogenesis by EED haploinsufficiency. *Sci. Rep.* 6, 29454. doi:10.1038/srep29454

Izzo, F., Lee, S. C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* 52 (4), 378–387. doi:10.1038/s41588-020-0595-4

Jann, J. C., and Tothova, Z. (2021). Cohesin mutations in myeloid malignancies. *Blood* 138, 649–661. doi:10.1182/blood.2019004259

Joshi, S. K., Qian, K., Bisson, W. H., Watanabe-Smith, K., Huang, A., Bottomly, D., et al. (2020). Discovery and characterization of targetable NTRK point mutations in hematologic neoplasms. *Blood* 135 (24), 2159–2170. doi:10.1182/blood.2019003691

Juban, G., Sakakini, N., Chagraoui, H., Cruz Hernandez, D., Cheng, Q., Soady, K., et al. (2021). Oncogenic Gata1 causes stage-specific megakaryocyte differentiation delay. *Haematologica* 106 (4), 1106–1119. doi:10.3324/haematol.2019.244541

Kanezaki, R., Toki, T., Terui, K., Xu, G., Wang, R., Shimada, A., et al. (2010). Down syndrome and GATA1 mutations in transient abnormal myeloproliferative disorder: Mutation classes correlate with progression to myeloid leukemia. *Blood* 116 (22), 4631–4638. doi:10.1182/blood-2010-05-282426

Kawara, H., Akahori, R., Wakasugi, M., Sancar, A., and Matsunaga, T. (2019). DCAF7 is required for maintaining the cellular levels of ERCC1-XPF and nucleotide excision repair. *Biochem. Biophys. Res. Commun.* 519 (1), 204–210. doi:10.1016/j.bbrc.2019.08.147

Kelly, M. J., So, J., Rogers, A. J., Gregory, G., Li, J., Zethoven, M., et al. (2019). Bcor loss perturbs myeloid differentiation and promotes leukaemogenesis. *Nat. Commun.* 10 (1), 1347. doi:10.1038/s41467-019-09250-6

Kemp, C. J., Moore, J. M., Moser, R., Bernard, B., Teater, M., Smith, L. E., et al. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* 7 (4), 1020–1029. doi:10.1016/j.celrep.2014.04.004

Kim, T. G., Kim, S., Jung, S., Kim, M., Yang, B., Lee, M. G., et al. (2017). CCCTC-binding factor is essential to the maintenance and quiescence of hematopoietic stem cells in mice. *Exp. Mol. Med.* 49 (8), e371. doi:10.1038/emm.2017.124

Kiyoi, H., Yamaji, S., Kojima, S., and Naoe, T. (2007). JAK3 mutations occur in acute megakaryoblastic leukemia both in Down syndrome children and non-Down syndrome adults. *Leukemia* 21 (3), 574–576. doi:10.1038/sj.leu.2404527

Klusmann, J. H., Creutzig, U., Zimmermann, M., Dworzak, M., Jorch, N., Langebrake, C., et al. (2008). Treatment and prognostic impact of transient leukemia in neonates with Down syndrome. *Blood* 111 (6), 2991–2998. doi:10.1182/blood-2007-10-118810

Klusmann, J. H., Li, Z., Böhmer, K., Maroz, A., Koch, M. L., Emmrich, S., et al. (2010). miR-125b-2 is a potential oncomiR on human chromosome 21 in megakaryoblastic leukemia. *Genes Dev.* 24 (5), 478–490. doi:10.1101/gad.1856210

Klusmann, J. H., Reinhardt, D., Hasle, H., Kaspers, G. J., Creutzig, U., Hahlen, K., et al. (2007). Janus kinase mutations in the development of acute megakaryoblastic leukemia in children with and without Down's syndrome. *Leukemia* 21 (7), 1584–1587. doi:10.1038/sj.leu.2404694

Labuhn, M., Perkins, K., Matzk, S., Varghese, L., Garnett, C., Papaemmanuil, E., et al. (2019). Mechanisms of progression of myeloid preleukemia to transformed myeloid leukemia in children with down syndrome. *Cancer Cell* 36 (2), 123–138. e110. doi:10.1016/j.ccell.2019.06.007

Lange, B. J., Kobrinsky, N., Barnard, D. R., Arthur, D. C., Buckley, J. D., Howells, W. B., et al. (1998). Distinctive demography, biology, and outcome of acute myeloid leukemia and myelodysplastic syndrome in children with Down syndrome: Children's Cancer Group Studies 2861 and 2891. *Blood* 91 (2), 608–615.

Langebrake, C., Klusmann, J. H., Wortmann, K., Kolar, M., Puhlmann, U., Reinhardt, D., et al. (2006). Concomitant aberrant overexpression of RUNX1 and NCAM in regenerating bone marrow of myeloid leukemia of Down's syndrome. *Haematologica* 91 (11), 1473–1480.

Laurent, A. P., Kotecha, R. S., and Malinge, S. (2020). Gain of chromosome 21 in hematological malignancies: Lessons from studying leukemia in children with down syndrome. *Leukemia* 34 (8), 1984–1999. doi:10.1038/s41375-020-0854-5

Letourneau, A., Santoni, F. A., Bonilla, X., Sailani, M. R., Gonzalez, D., Kind, J., et al. (2014). Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* 508 (7496), 345–350. doi:10.1038/nature13200

Lindberg, M. F., and Meijer, L. (2021). Dual-specificity, tyrosine phosphorylation-regulated kinases (DYRKs) and cdc2-like kinases (CLKs) in human disease, an overview. *Int. J. Mol. Sci.* 22 (11), 6047. doi:10.3390/ijms22116047

Liu, B., Filippi, S., Roy, A., and Roberts, I. (2015). Stem and progenitor cell dysfunction in human trisomies. *EMBO Rep.* 16 (1), 44–62. doi:10.15252/embr.201439583

Lorenzini, P. A., Chew, R. S. E., Tan, C. W., Yong, J. Y., Zhang, F., Zheng, J., et al. (2019). Human PRPF40B regulates hundreds of alternative splicing targets and represses a hypoxia expression signature. *RNA* 25 (8), 905–920. doi:10.1261/rna.069534.118

Loscocco, G. G., Guglielmelli, P., and Vannucchi, A. M. (2020). Impact of mutational profile on the management of myeloproliferative neoplasms: A short review of the emerging data. *Onco. Targets. Ther.* 13, 12367–12382. doi:10.2147/OTT.S287944

Lu, H. Y., Sharma, M., Sharma, A. A., Lacson, A., Szpurko, A., Luider, J., et al. (2021). Mechanistic understanding of the combined immunodeficiency in complete human CARD11 deficiency. *J. Allergy Clin. Immunol.* 148 (6), 1559–1574. e13.e1513. doi:10.1016/j.jaci.2021.04.006

Magrin, L., Fanale, D., Brando, C., Fiorino, A., Corsini, L. R., Sciacchitano, R., et al. (2021). POLE, POLD1, and NTHL1: The last but not the least hereditary cancer-predisposing genes. *Oncogene* 40 (40), 5893–5901. doi:10.1038/s41388-021-01984-2

Majewski, I. J., Blewitt, M. E., de Graaf, C. A., McManus, E. J., Bahlo, M., Hilton, A. A., et al. (2008). Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol.* 6 (4), e93. doi:10.1371/journal.pbio.0060093

Malinge, S., Bliss-Moreau, M., Kirsammer, G., Diebold, L., Chlon, T., Gurbuxani, S., et al. (2012). Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *J. Clin. Invest.* 122 (3), 948–962. doi:10.1172/JCI60455

Malinge, S., Chlon, T., Doré, L. C., Ketterling, R. P., Tallman, M. S., Paietta, E., et al. (2013). Development of acute megakaryoblastic leukemia in Down syndrome is associated with sequential epigenetic changes. *Blood* 122 (14), e33–43. doi:10.1182/blood-2013-05-503011

Malinge, S., Ragu, C., Della-Valle, V., Pisani, D., Constantinescu, S. N., Perez, C., et al. (2008). Activating mutations in human acute megakaryoblastic leukemia. *Blood* 112 (10), 4220–4226. doi:10.1182/blood-2008-01-136366

Man, N., Mas, G., Karl, D. L., Sun, J., Liu, F., Yang, Q., et al. (2021). p300 suppresses the transition of myelodysplastic syndromes to acute myeloid leukemia. *JCI Insight* 6 (19), e138478. doi:10.1172/jci.insight.138478

Manukjan, G., Ripperger, T., Santer, L., von Neuhoff, N., Ganser, A., Schambach, A., et al. (2015). Expression of the ETS transcription factor GABPα is positively correlated to the BCR-ABL1/ABL1 ratio in CML patients and affects imatinib sensitivity *in vitro*. *Exp. Hematol.* 43 (10), 880–890. doi:10.1016/j.exphem.2015.05.011

Maslah, N., Cassinat, B., Verger, E., Kiladjian, J. J., and Velazquez, L. (2017). The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. *Leukemia* 31 (8), 1661–1670. doi:10.1038/leu.2017.139

Massey, G. V., Zipursky, A., Chang, M. N., Doyle, J. J., Nasim, S., Taub, J. W., et al. (2006). A prospective study of the natural history of transient leukemia (TL) in neonates with Down syndrome (DS): Children's Oncology Group (COG) study POG-9481. *Blood* 107 (12), 4606–4613. doi:10.1182/blood-2005-06-2448

Matsuo, S., Nishinaka-Arai, Y., Kazuki, Y., Oshimura, M., Nakahata, T., Niwa, A., et al. (2021). Pluripotent stem cell model of early hematopoiesis in Down syndrome reveals quantitative effects of short-form GATA1 protein on lineage specification. *PLoS One* 16 (3), e0247595. doi:10.1371/journal.pone.0247595

Maurya, S., Yang, W., Tamai, M., Zhang, Q., Erdmann-Gilmore, P., Bystry, A., et al. (2021). Loss of KMT2C reprograms the epigenomic landscape in hPSCs resulting in NODAL overexpression and a failure of hemogenic endothelium specification. *Epigenetics* 17, 220–238. doi:10.1080/15592294.2021.1954780

Maxson, J. E., and Tyner, J. W. (2017). Genomics of chronic neutrophilic leukemia. *Blood* 129 (6), 715–722. doi:10.1182/blood-2016-10-695981

Mazzi, S., Dessen, P., Vieira, M., Dufour, V., Cambot, M., El Khoury, M., et al. (2021). Dual role of EZH2 in megakaryocyte differentiation. *Blood* 138 (17), 1603–1614. doi:10.1182/blood.2019004638

Mazzola, M., Deflorian, G., Pezzotta, A., Ferrari, L., Fazio, G., Bresciani, E., et al. (2019a). Nipbl: A new player in myeloid cell differentiation. *Haematologica* 104 (7), 1332–1341. doi:10.3324/haematol.2018.200899

Mazzola, M., Deflorian, G., Pezzotta, A., Ferrari, L., Fazio, G., Bresciani, E., et al. (2019b). Nipbl: A new player in myeloid cell differentiation. *Haematologica* 104 (7), 1332–1341. doi:10.3324/haematol.2018.200899

Mazzola, M., Pezzotta, A., Fazio, G., Rigamonti, A., Bresciani, E., Gaudenzi, G., et al. (2020). Dysregulation of NIPBL leads to impaired RUNX1 expression and haematopoietic defects. *J. Cell. Mol. Med.* 24 (11), 6272–6282. doi:10.1111/jcmm.15269

Moles, R., Bai, X. T., Chaib-Mezrag, H., and Nicot, C. (2016). WRN-targeted therapy using inhibitors NSC 19630 and NSC 617145 induce apoptosis in HTLV-1-transformed adult T-cell leukemia cells. *J. Hematol. Oncol.* 9 (1), 121. doi:10.1186/s13045-016-0352-4

Moser, B., Edtmayer, S., Witalisz-Siepracka, A., and Stoiber, D. (2021). The ups and downs of STAT inhibition in acute myeloid leukemia. *Biomedicines* 9 (8), 1051. doi:10.3390/biomedicines9081051

Muffels, I. J. J., Wiame, E., Fuchs, S. A., Massink, M. P. G., Rehmann, H., Musch, J. L. I., et al. (2021). NAA80 bi-allelic missense variants result in high-frequency hearing loss, muscle weakness and developmental delay. *Brain Commun.* 3 (4), fcab256. doi:10.1093/braincomms/fcab256

Mujahed, H., Miliara, S., Neddermeyer, A., Bengtzén, S., Nilsson, C., Deneberg, S., et al. (2020). AML displays increased CTCF occupancy associated with aberrant

gene expression and transcription factor binding. *Blood* 136 (3), 339–352. doi:10.1182/blood.2019002326

Muskens, I. S., Li, S., Jackson, T., Elliot, N., Hansen, H. M., Myint, S. S., et al. (2021). The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis. *Nat. Commun.* 12 (1), 821. doi:10.1038/s41467-021-21064-z

Nagase, R., Inoue, D., Pastore, A., Fujino, T., Hou, H. A., Yamasaki, N., et al. (2018). Expression of mutant Asxl1 perturbs hematopoiesis and promotes susceptibility to leukemic transformation. *J. Exp. Med.* 215 (6), 1729–1747. doi:10.1084/jem.20171151

Nakamura-Ishizu, A., and Suda, T. (2020). Multifaceted roles of thrombopoietin in hematopoietic stem cell regulation. *Ann. N. Y. Acad. Sci.* 1466 (1), 51–58. doi:10.1111/nyas.14169

Nie, Z., Gao, W., Zhang, Y., Hou, Y., Liu, J., Li, Z., et al. (2019). STAG2 loss-of-function mutation induces PD-L1 expression in U2OS cells. *Ann. Transl. Med.* 7 (7), 127. doi:10.21037/atm.2019.02.23

Nikolaev, S. I., Santoni, F., Vannier, A., Falconnet, E., Giarin, E., Basso, G., et al. (2013). Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down syndrome. *Blood* 122 (4), 554–561. doi:10.1182/blood-2013-03-491936

Nishinaka-Arai, Y., Niwa, A., Matsuo, S., Kazuki, Y., Yakura, Y., Hiroma, T., et al. (2021). Down syndrome-related transient abnormal myelopoiesis is attributed to a specific erythro-megakaryocytic subpopulation with GATA1 mutation. *Haematologica* 106 (2), 635–640. doi:10.3324/haematol.2019.242693

Ntziachristos, P., Tsirigos, A., Van Vlierberghe, P., Nedjic, J., Trimarchi, T., Flaherty, M. S., et al. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* 18 (2), 298–301. doi:10.1038/nm.2651

Ochi, Y., Kon, A., Sakata, T., Nakagawa, M. M., Nakazawa, N., Kakuta, M., et al. (2020). Combined cohesin-RUNX1 deficiency synergistically perturbs chromatin looping and causes myelodysplastic syndromes. *Cancer Discov.* 10 (6), 836–853. doi:10.1158/2159-8290.CD-19-0982

Ono, R., Hasegawa, D., Hirabayashi, S., Kamiya, T., Yoshida, K., Yonekawa, S., et al. (2015). Acute megakaryoblastic leukemia with acquired trisomy 21 and GATA1 mutations in phenotypically normal children. *Eur. J. Pediatr.* 174 (4), 525–531. doi:10.1007/s00431-014-2430-3

Padmakumar, D., Chandraprabha, V. R., Gopinath, P., Vimala Devi, A. R. T., Anitha, G. R. J., Sreelatha, M. M., et al. (2021). A concise review on the molecular genetics of acute myeloid leukemia. *Leuk. Res.* 111, 106727. doi:10.1016/j.leukres.2021.106727

Pandey, R., Saxena, M., and Kapur, R. (2017). Role of SHP2 in hematopoiesis and leukemogenesis. *Curr. Opin. Hematol.* 24 (4), 307–313. doi:10.1097/MOH.0000000000000345

Panferova, A., Gaskova, M., Nikitin, E., Baryshev, P., Timofeeva, N., Kazakova, A., et al. (2021). GATA1 mutation analysis and molecular landscape characterization in acute myeloid leukemia with trisomy 21 in pediatric patients. *Int. J. Lab. Hematol.* 43 (4), 713–723. doi:10.1111/ijlh.13451

Park, H. J., Li, J., Hannah, R., Biddie, S., Leal-Cervantes, A. I., Kirschner, K., et al. (2016). Cytokine-induced megakaryocytic differentiation is regulated by genome-wide loss of a uSTAT transcriptional program. *EMBO J.* 35 (6), 580–594. doi:10.15252/embj.201592383

Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., et al. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* 349 (6306), 257–260. doi:10.1038/349257a0

Pine, S. R., Guo, Q., Yin, C., Jayabose, S., Druschel, C. M., Sandoval, C., et al. (2007). Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood* 110 (6), 2128–2131. doi:10.1182/blood-2007-01-069542

Polachek, W. S., Moshrif, H. F., Franti, M., Coen, D. M., Sreenu, V. B., Strang, B. L., et al. (2016). High-throughput small interfering RNA screening identifies phosphatidylinositol 3-kinase class II alpha as important for production of human cytomegalovirus virions. *J. Virol.* 90 (18), 8360–8371. doi:10.1128/JVI.01134-16

Rainis, L., Toki, T., Pimanda, J. E., Rosenthal, E., Machol, K., Strehl, S., et al. (2005). The proto-oncogene ERG in megakaryoblastic leukemias. *Cancer Res.* 65 (17), 7596–7602. doi:10.1158/0008-5472.CAN-05-0147

Reymond, A., Friedli, M., Henrichsen, C. N., Chapot, F., Deutsch, S., Ucla, C., et al. (2001). From PREDs and open reading frames to cDNA isolation: Revisiting the human chromosome 21 transcription map. *Genomics* 78 (1-2), 46–54. doi:10.1006/geno.2001.6640

Rivas, M. A., Meydan, C., Chin, C. R., Challman, M. F., Kim, D., Bhinder, B., et al. (2021). Smc3 dosage regulates B cell transit through germinal centers and restricts their malignant transformation. *Nat. Immunol.* 22 (2), 240–253. doi:10.1038/s41590-020-00827-8

Roberts, I., Alford, K., Hall, G., Juban, G., Richmond, H., Norton, A., et al. (2013). GATA1-mutant clones are frequent and often unsuspected in babies with down syndrome: Identification of a population at risk of leukemia. *Blood* 122 (24), 3908–3917. doi:10.1182/blood-2013-07-515148

Saida, S., Watanabe, K., Sato-Otsubo, A., Terui, K., Yoshida, K., Okuno, Y., et al. (2013). Clonal selection in xenografted TAM recapitulates the evolutionary process of myeloid leukemia in Down syndrome. *Blood* 121 (21), 4377–4387. doi:10.1182/blood-2012-12-474387

Salek-Ardakani, S., Smooha, G., de Boer, J., Sebire, N. J., Morrow, M., Rainis, L., et al. (2009). ERG is a megakaryocytic oncogene. *Cancer Res.* 69 (11), 4665–4673. doi:10.1158/0008-5472.CAN-09-0075

Sas, V., Pasca, S., Jurj, A., Pop, L., Muramatsu, H., Ono, H., et al. (2020). MicroRNA-155-5p plays a critical role in transient leukemia of down syndrome by targeting tumor necrosis factor receptor superfamily members. *Cell. Physiol. biochem.* 54 (5), 994–1012. doi:10.33594/000000283

Sasine, J. P., Himburg, H. A., Termini, C. M., Roos, M., Tran, E., Zhao, L., et al. (2018). Wild-type Kras expands and exhausts hematopoietic stem cells. *JCI Insight* 3 (11), 98197. doi:10.1172/jci.insight.98197

Shaham, L., Vendramini, E., Ge, Y., Goren, Y., Birger, Y., Tijssen, M. R., et al. (2015). MicroRNA-486-5p is an erythroid oncomiR of the myeloid leukemias of Down syndrome. *Blood* 125 (8), 1292–1301. doi:10.1182/blood-2014-06-581892

Shi, X., Yang, Y., Shang, S., Wu, S., Zhang, W., Peng, L., et al. (2019). Cooperation of Dnmt3a R878H with nras G12D promotes leukemogenesis in knock-in mice: A pilot study. *BMC Cancer* 19 (1), 1072. doi:10.1186/s12885-019-6207-y

Shimizu, R., Kobayashi, E., Engel, J. D., and Yamamoto, M. (2009). Induction of hyperproliferative fetal megakaryopoiesis by an N-terminally truncated GATA1 mutant. *Genes cells.* 14 (9), 1119–1131. doi:10.1111/j.1365-2443.2009.01338.x

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479 (7371), 74–79. doi:10.1038/nature10442

Si, X., Zhang, X., Hao, X., Li, Y., Chen, Z., Ding, Y., et al. (2016). Upregulation of miR-99a is associated with poor prognosis of acute myeloid leukemia and promotes myeloid leukemia cell expansion. *Oncotarget* 7 (47), 78095–78109. doi:10.18632/oncotarget.12947

Singh, A., Mandal, A., Guru, V., Srinivasan, S., and Seth, R. (2017). Transient abnormal myelopoiesis: A varied spectrum of clinical presentation. *J. Hematol.* 6 (1), 25–28. doi:10.14740/jh306w

Singh, A., Mencia-Trinchant, N., Griffiths, E. A., Altahan, A., Swaminathan, M., Gupta, M., et al. (2022). Mutant PPM1D- and TP53-driven hematopoiesis populates the hematopoietic compartment in response to peptide receptor radionuclide therapy. *JCO Precis. Oncol.* 6, e2100309. doi:10.1200/PO.21.00309

Song, L., Jiang, W., Liu, W., Ji, J. H., Shi, T. F., Zhang, J., et al. (2016). Protein tyrosine phosphatases receptor type D is a potential tumour suppressor gene inactivated by deoxyribonucleic acid methylation in paediatric acute myeloid leukaemia. *Acta Paediatr.* 105 (3), e132–141. doi:10.1111/apa.13284

Soukup, A. A., and Bresnick, E. H. (2020). GATA2 +9.5 enhancer: From principles of hematopoiesis to genetic diagnosis in precision medicine. *Curr. Opin. Hematol.* 27 (3), 163–171. doi:10.1097/MOH.0000000000000576

Stankiewicz, M. J., and Crispino, J. D. (2013). AKT collaborates with ERG and Gata1s to dysregulate megakaryopoiesis and promote AMKL. *Leukemia* 27 (6), 1339–1347. doi:10.1038/leu.2013.33

Stankiewicz, M. J., and Crispino, J. D. (2009). ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. *Blood* 113 (14), 3337–3347. doi:10.1182/blood-2008-08-174813

Stankov, K., Popovic, S., and Mikov, M. (2014). C-KIT signaling in cancer treatment. *Curr. Pharm. Des.* 20 (17), 2849–2880. doi:10.2174/13816128113199990593

Sureau, L., Orvain, C., Ianotto, J. C., Ugo, V., Kiladjian, J. J., Luque Paz, D., et al. (2021). Efficacy and tolerability of janus kinase inhibitors in myelofibrosis: A systematic review and network meta-analysis. *Blood Cancer J.* 11 (7), 135. doi:10.1038/s41408-021-00526-z

Sweeney, T. R., Dhote, V., Guca, E., Hellen, C. U. T., Hashem, Y., Pestova, T. V., et al. (2021). Functional role and ribosomal position of the unique N-terminal region of DHX29, a factor required for initiation on structured mammalian mRNAs. *Nucleic Acids Res.* 49 (22), 12955–12969. doi:10.1093/nar/gkab1192

Thoms, J. A. I., Truong, P., Subramanian, S., Knezevic, K., Harvey, G., Huang, Y., et al. (2021). Disruption of a GATA2-TAL1-ERG regulatory circuit promotes erythroid transition in healthy and leukemic stem cells. *Blood* 138 (16), 1441–1455. doi:10.1182/blood.2020009707

Tian, L., Chavez, M., Chang, G. S., Helton, N. M., Katerndahl, C. D. S., Miller, C. A., et al. (2021). Kdm6a deficiency restricted to mouse hematopoietic cells causes an age- and sex-dependent myelodysplastic syndrome-like phenotype. *PLoS One* 16 (11), e0255706. doi:10.1371/journal.pone.0255706

Todisco, G., Creignou, M., Gallì, A., Guglielmelli, P., Rumi, E., Roncador, M., et al. (2021). Co-mutation pattern, clonal hierarchy, and clone size concur to determine disease phenotype of SRSF2^{P95}-mutated neoplasms. *Leukemia* 35 (8), 2371–2381. doi:10.1038/s41375-020-01106-z

Toki, T., Katsuoka, F., Kanezaki, R., Xu, G., Kurotaki, H., Sun, J., et al. (2005). Transgenic expression of BACH1 transcription factor results in megakaryocytic impairment. *Blood* 105 (8), 3100–3108. doi:10.1182/blood-2004-07-2826

Torrano, V., Chernukhin, I., Docquier, F., D'Arcy, V., León, J., Klenova, E., et al. (2005). CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J. Biol. Chem.* 280 (30), 28152–28161. doi:10.1074/jbc.M501481200

Tsai, M. H., Hou, J. W., Yang, C. P., Yang, P. H., Chu, S. M., Hsu, J. F., et al. (2011). Transient myeloproliferative disorder and GATA1 mutation in neonates with and without Down syndrome. *Indian J. Pediatr.* 78 (7), 826–832. doi:10.1007/s12098-010-0312-x

Uffmann, M., Rasche, M., Zimmermann, M., von Neuhoff, C., Creutzig, U., Dworzak, M., et al. (2017). Therapy reduction in patients with down syndrome and myeloid leukemia: The international ML-DS 2006 trial. *Blood* 129 (25), 3314–3321. doi:10.1182/blood-2017-01-765057

Van Der Werf, I., Wojtuszkiewicz, A., Yao, H., Sciarrillo, R., Meggendorfer, M., Hutter, S., et al. (2021). SF3B1 as therapeutic target in FLT3/ITD positive acute myeloid leukemia. *Leukemia* 35 (9), 2698–2702. doi:10.1038/s41375-021-01273-7

Vara, N., Liu, Y., Yan, Y., Lensing, S. Y., Colorado, N., Robinson, D., et al. (2020). Sustained fetal hematopoiesis causes juvenile death from leukemia: Evidence from a dual-age-specific mouse model. *Blood Adv.* 4 (15), 3728–3740. doi:10.1182/bloodadvances.2020002326

Viny, A. D., Bowman, R. L., Liu, Y., Lavallée, V. P., Eisman, S. E., Xiao, W., et al. (2019). Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell* 25 (5), 682–696. e688. doi:10.1016/j.stem.2019.08.003

Volk, A., Liang, K., Suraneni, P., Li, X., Zhao, J., Bulic, M., et al. (2018). A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis. *Cancer Cell* 34 (5), 707–723. e707. doi:10.1016/j.ccell.2018.10.004

Vukadin, L., Kim, J. H., Park, E. Y., Stone, J. K., Ungerleider, N., Baddoo, M. C., et al. (2021). SON inhibits megakaryocytic differentiation via repressing RUNX1 and the megakaryocytic gene expression program in acute megakaryoblastic leukemia. *Cancer Gene Ther.* 28 (9), 1000–1015. doi:10.1038/s41417-020-00262-9

Wagenblast, E., Araújo, J., Gan, O. I., Cutting, S. K., Murison, A., Krivdova, G., et al. (2021). Mapping the cellular origin and early evolution of leukemia in Down syndrome. *Science* 373 (6551), eabf6202. doi:10.1126/science.abf6202

Walters, D. K., Mercher, T., Gu, T. L., O'Hare, T., Tyner, J. W., Loriaux, M., et al. (2006). Activating alleles of JAK3 in acute megakaryoblastic leukemia. *Cancer Cell* 10 (1), 65–75. doi:10.1016/j.ccr.2006.06.002

Wang, A. J., Han, Y., Jia, N., Chen, P., and Minden, M. D. (2020). NPM1c impedes CTCF functions through cytoplasmic mislocalization in acute myeloid leukemia. *Leukemia* 34 (5), 1278–1290. doi:10.1038/s41375-019-0681-8

Wang, T., Glover, B., Hadwiger, G., Miller, C. A., di Martino, O., Welch, J. S., et al. (2019). Smc3 is required for mouse embryonic and adult hematopoiesis. *Exp. Hematol.* 70, 70–84. e76. doi:10.1016/j.exphem.2018.11.008

Wechsler, J., Greene, M., McDevitt, M. A., Anastasi, J., Karp, J. E., Le Beau, M. M., et al. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* 32 (1), 148–152. doi:10.1038/ng955

Wu, Y., Zhu, H., and Wu, H. (2020). PTEN in regulating hematopoiesis and leukemogenesis. *Cold Spring Harb. Perspect. Med.* 10 (10), a036244. doi:10.1101/cshperspect.a036244

Xu, G., Kato, K., Toki, T., Takahashi, Y., Terui, K., Ito, E., et al. (2006). Development of acute megakaryoblastic leukemia from a minor clone in a Down syndrome patient with clinically overt transient myeloproliferative disorder. *J. Pediatr. Hematol. Oncol.* 28 (10), 696–698. doi:10.1097/01.mph.0000212997.02554.f6

Xu, N., Donohoe, M. E., Silva, S. S., and Lee, J. T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat. Genet.* 39 (11), 1390–1396. doi:10.1038/ng.2007.5

Yan, F., Li, J., Milosevic, J., Petroni, R., Liu, S., Shi, Z., et al. (2021). KAT6A and ENL form an epigenetic transcriptional control module to drive critical leukemogenic gene expression programs. *Cancer Discov.* 12, 792–811. doi:10.1158/2159-8290.CD-20-1459

Yanagida, M., Osato, M., Yamashita, N., Liqun, H., Jacob, B., Wu, F., et al. (2005). Increased dosage of Runx1/AML1 acts as a positive modulator of myeloid leukemogenesis in BXH2 mice. *Oncogene* 24 (28), 4477–4485. doi:10.1038/sj.onc.1208675

Yang, Z. F., Zhang, H., Ma, L., Peng, C., Chen, Y., Wang, J., et al. (2013). GABP transcription factor is required for development of chronic myelogenous leukemia via its control of PRKD2. *Proc. Natl. Acad. Sci. U. S. A.* 110 (6), 2312–2317. doi:10.1073/pnas.1212904110

Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., et al. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* 45 (11), 1293–1299. doi:10.1038/ng.2759

Yousefelahiyeh, M., Xu, J., Alvarado, E., Yu, Y., Salven, D., Nissen, R. M., et al. (2018). DCAF7/WDR68 is required for normal levels of DYRK1A and DYRK1B. *PLoS One* 13 (11), e0207779. doi:10.1371/journal.pone.0207779

Yuzawa, K., Terui, K., Toki, T., Kanezaki, R., Kobayashi, A., Sato, T., et al. (2020). Clinical, cytogenetic, and molecular analyses of 17 neonates with transient abnormal myelopoiesis and nonconstitutional trisomy 21. *Pediatr. Blood Cancer* 67 (4), e28188. doi:10.1002/pbc.28188

Zafar, N., Ghias, K., and Fadoo, Z. (2021). Genetic aberrations involved in relapse of pediatric acute myeloid leukemia: A literature review. *Asia. Pac. J. Clin. Oncol.* 17 (5), e135–e141. doi:10.1111/ajco.13367

Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M. G., Reale, A., et al. (2012). ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. *Biochem. J.* 441 (2), 645–652. doi:10.1042/BJ20111417

Zaslavsky, A., Chou, S. T., Schadler, K., Lieberman, A., Pimkin, M., Kim, Y. J., et al. (2013). The calcineurin-NFAT pathway negatively regulates megakaryopoiesis. *Blood* 121 (16), 3205–3215. doi:10.1182/blood-2012-04-421172

Zeisig, B. B., and So, C. W. E. (2021). Therapeutic opportunities of targeting canonical and noncanonical PcG/TrxG functions in acute myeloid leukemia. *Annu. Rev. Genomics Hum. Genet.* 22, 103–125. doi:10.1146/annurev-genom-111120-102443

Zhang, L., Li, X., Ke, Z., Huang, L., Liang, Y., Wu, J., et al. (2013). MiR-99a may serve as a potential oncogene in pediatric myeloid leukemia. *Cancer Cell Int.* 13 (1), 110. doi:10.1186/1475-2867-13-110

Zhang, W., Berthelet, J., Michail, C., Bui, L. C., Gou, P., Liu, R., et al. (2021). Human CREBBP acetyltransferase is impaired by etoposide quinone, an oxidative and leukemogenic metabolite of the anticancer drug etoposide through modification of redox-sensitive zinc-finger cysteine residues. *Free Radic. Biol. Med.* 162, 27–37. doi:10.1016/j.freeradbiomed.2020.11.027

Zhang, Y., Taylor, B. R., Shannon, K., and Clapp, D. W. (2001). Quantitative effects of Nf1 inactivation on *in vivo* hematopoiesis. *J. Clin. Invest.* 108 (5), 709–715. doi:10.1172/JCI12758

Zhou, J., Wu, H., Guo, C., Li, B., Zhou, L. L., Liang, A. B., et al. (2021). A comprehensive genome-wide analysis of long non-coding RNA and mRNA expression profiles of JAK2V617F-positive classical myeloproliferative neoplasms. *Bioengineered* 12 (2), 10564–10586. doi:10.1080/21655979.2021.2000226

Zhuo, M. Q., Pan, Y. X., Wu, K., Xu, Y. H., and Luo, Z. (2017). Characterization and mechanism of phosphoinositide 3-kinases (PI3Ks) members in insulin-induced changes of protein metabolism in yellow catfish *Pelteobagrus fulvidraco*. *Gen. Comp. Endocrinol.* 247, 34–45. doi:10.1016/j.ygcen.2017.04.002

Zimmermannova, O., Doktorova, E., Stuchly, J., Kanderova, V., Kuzilkova, D., Strnad, H., et al. (2017). An activating mutation of GNB1 is associated with resistance to tyrosine kinase inhibitors in ETV6-ABL1-positive leukemia. *Oncogene* 36 (43), 5985–5994. doi:10.1038/onc.2017.210

Zuin, J., Dixon, J. R., van der Reijden, M. I., Ye, Z., Kolovos, P., Brouwer, R. W., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 111 (3), 996–1001. doi:10.1073/pnas.1317788111

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 September 2022
DOI 10.3389/fgene.2022.946854

OPEN ACCESS

EDITED BY
Yan-Ming Xu,
Shantou University, China

REVIEWED BY
Laura Southgate,
University of London, United Kingdom
Yongqiang Deng,
Southern Medical University, China
Shangze Li,
Chongqing University, China

*CORRESPONDENCE
Hong Tian,
tianhong1110@hust.edu.cn
Chuanzhou Li,
chuanzhouli@hust.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 18 May 2022
ACCEPTED 22 August 2022
PUBLISHED 13 September 2022

CITATION
Tian H, Chu F, Li Y, Xu M, Li W and Li C (2022), Synergistic effects of rare variants of ARHGAP31 and FBLN1 in vitro in terminal transverse limb defects. *Front. Genet.* 13:946854. doi: 10.3389/fgene.2022.946854

Synergistic effects of rare variants of ARHGAP31 and FBLN1 *in vitro* in terminal transverse limb defects

Hong Tian[1*†], Fan Chu[1†], Yingjie Li[1], Mengmeng Xu[1], Wenjiao Li[2] and Chuanzhou Li[1*]

[1]Department of Medical Genetics, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Clinical Laboratory, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China

Background: Aplasia cutis congenita (ACC) and terminal transverse limb defects (TTLDs) are the most common features of Adams-Oliver syndrome (AOS). *ARHGAP31* is one of the causative genes for autosomal dominant forms of AOS, meanwhile its variants may only cause isolated TTLD. Here, we report a proband presented with apparent TTLD but not ACC.

Methods: Whole exome sequencing (WES) and Sanger sequencing were applied to identify causative genes. Expression vectors were constructed for transfections in mammalian cell cultures followed by biochemical and functional analysis including immunoblotting, immunofluorescence staining, and cell counting kit-8 assay.

Results: WES and Sanger sequencing suggested that the proband inherited rare *ARHGAP31* variant [c.2623G > A (*p*.Glu875Lys)] and a rare *FBLN1* variant [c.1649G > A (*p*.Arg550His)] from one of her asymptomatic parents, respectively. Given *FBLN1* variation has also been linked to syndactyly, we suspected that the two genes together contributed to the TTLD phenotype and explored their possible roles *in vitro*. Mutant FBLN1 showed reduced expression resulted from impaired protein stability, whereas ARHGAP31 protein expression was unaltered by mutation. Functional assays showed that only in the co-transfected group of two mutants cell viability was decreased, cell proliferation was impaired, and apoptosis was activated. Cdc42 activity was declined by both ARHGAP31 mutation and FBLN1 mutation alone, and the two together. Furthermore, the MAPK/ERK pathway was only activated by two mutants co-transfected group compared with two wild-type transfections.

Conclusion: We report a case carrying two rare variants of limb defects associated genes, *ARHGAP31* and *FBLN1*, and provide *in vitro* evidence that synergistic disruption of cellular functions attributed by the two mutants may potentiate the penetrance of clinical manifestations, expanding our knowledge of clinical complexity of causal gene interactions in TTLD and other genetic disorders.

KEYWORDS

***ARHGAP31*, *FBLN1*, AOS, Cdc42, syndactyly, Limb defects**

Introduction

Adams-Oliver syndrome [AOS (MIM#100300)], also known as AOS1, is a rare inherited disease characterized by aplasia cutis congenita (ACC) and terminal transverse limb defects (TTLDs) as the most common clinical features (Kuster et al., 1988). ACC is mostly present on the scalp, whereas some patients show extra skull defects in various degrees (Adams and Oliver, 1945). TTLD could present different degrees of injury at one limb end to four limbs end. In addition to ACC and TTLD, other common anomalies include central nervous system anomalies, cutis marmorata telangiectasia congenita, and congenital heart defects (Hassed et al., 2017). AOS patients may show severe-to-mild symptoms, but carriers with same gene variant within one family may show phenotypes ranging from no obvious clinical manifestations to severe multi-system abnormalities (Hassed et al., 2017). Clinically, in cases with a known family history, presence of either ACC or TTLD has been considered sufficient to warrant the diagnosis of AOS (Snape et al., 2009).

ARHGAP31 is one of the causal genes identified in AOS (Hassed et al., 2017; Meester et al., 2018). Previously, Southgate has found two *ARHGAP31* variants in two independent families with patients present with ACC and TTLD, which are the most common clinical features of AOS (Southgate et al., 2011). Later in 2014, an *ARHGAP31* variant in a big family was reported by Isrie et al. (2014), with the only clinical feature of isolated TTLD, suggesting the possibility of segregated phenotype caused by a single gene variant. Along with variants in other causative genes such as *DLL4* and *NOTCH1*, *ARHGAP31* variants have consistently been confirmed in a large cohort of AOS by Meester et al. (2018). As a GTPase-activating protein (GAP) encoded by *ARHGAP31* gene, ARHGAP31 inactivates the Rho GTPases family members Cdc42/Rac1, which master central roles in cell division, survival and migration, and the maintenance of the actin cytoskeleton (Melendez et al., 2011). Truncated forms of ARHGAP31 disrupt the actin cytoskeleton structures through inactivation of Cdc42/Rac1 (Southgate et al., 2011).

Fibulin-1, encoded by *FBLN1*, is an extracellular matrix protein that has an important role in the structure of elastic fibers and basement membranes of various tissues (De Vega et al., 2009). A missense mutation in fibulin-1 in a consanguineous family showed a novel syndrome of syndactyly and other abnormalities in the central nervous system (CNS) (Bohlega et al., 2014), indicative of the crucial role of fibulin-1 in development of the CNS and various connective tissues. Nonetheless, no involvement of FBLN1 in AOS has been reported so far.

Here, we report a proband with similar condition to the cases of arrested development described by Isrie et al. (2014). Using whole exome sequencing and Sanger sequencing, two rare variations were identified. The proband carries an *ARHGAP31* variant (c.2623G > A (p.Glu875Lys) concomitant with *FBLN1* variant (c.1649G > A (*p*.Arg550His). Each of her parents carrying one of the former mutants was phenotypically normal. Since aforementioned *FBLN1* variation has also been shown to be linked with syndactyly (Debeer et al., 2002), we therefore suspected that the two genes together contributed to the TTLD phenotype and explored their possible roles *in vitro*. With exogenous expressions of ARHGAP31 and FBLN1 in mammalian cells, cell viability, target gene expression, cytotoxicity, apoptosis, and related signaling molecules were analyzed, and all the evidence indicate a pro-apoptotic role of FBLN1 in exacerbating the dysfunction of ARHGAP31 caused by variation, ultimately triggering the synergistic effects of penetrance of TTLD phenotype by *ARHGAP31* and *FBLN1* variations.

Materials and methods

Patient information

A 30-year-old female was inspected in our consultation center with abnormal finger and toe development, and her parents had no AOS family history of the defects. She is the only child in the family whose mother naturally conceived and gave birth with normal birth parameters. The left hand fingers and two foot toes with abnormal development were observed at birth. There were no other abnormalities during development. Her parents have no history of teratogenic contact. She came for preconception counseling. This study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. All subjects have signed an informed consent form, and phenotypic photos have been approved for publication in this study.

Whole exome sequencing and bioinformatics analysis

Peripheral blood from all subjects within the family was collected for genomic DNA extraction. WES of genomic DNA of the proband was performed by Shenzhen BGI Clinical Inspection Center Co., Ltd. Briefly, genomic DNA was first extracted for library preparation. A BGIV4 chip was used to capture and enrich gene exons and DNA adjacent to the clipping region. The MGISEQ-2000 sequencing platform was employed for variants detection and analysis. The quality control index of sequencing data is: the average sequencing depth of target area is ≥180×, and the proportion of sites with average depth of target area >20× is >95%.

Sequenced fragments were compared with UCSC hg19 human reference genome by the Burrows–Wheeler alignment (BWA) to eliminate duplication. Database including the 1000 genomes, ExAC, ESP6500, gnomAD, TOPMed, and

BGI local database were referenced. Variants with a frequency of less than or equal to 1% of known causative genes in OMIM was listed in the clinical test report. Genome Analysis Toolkit (GATK) software was used for SNV, INDEL, and genotype detection. ExomeDepth was used to detect at exon level copy number variation. The naming of variation follows the naming standard of Human Genome Variation Society (HGVS). Variant pathogenicity annotation is based on the American College of Medical Genetics and Genomics (ACMG) and the American Society for Molecular Pathology (AMP) sequence interpretation guidelines, and with reference to interpretation of the guidelines from the ClinGen Sequence Variant Interpretation working group and the British Society for Clinical Genome Sciences (ACGSs).

Validation of variations *via* Sanger sequencing

ARHGAP31 and *FBLN1* variants were identified by aforementioned WES. Variations in *ARHGAP31* and *FBLN1* genes were validated by Sanger sequencing. The PCR products were sequenced on an ABI 3130 DNA Analyzer (Applied Biosystems, United States). The sequencing data were visualized in DNA Star software to confirm the variations. We carried out family analysis in order to explore the source of variations in *ARHGAP31* [c.2623G > A (*p*.Glu875Lys)] and *FBLN1* [c.1649G > A (*p*.Arg550His)]. Three DNA samples were extracted respectively from peripheral blood lymphocytes of patient and her parents using a genomic DNA extraction kit (Cwbio Company, CW 2087M). Sanger sequencing were performed using primers as follows: *ARHGAP31* forward primer: 5′-CCAGGCAATCTGTCTCCTCC-3′, *ARHGAP31* reverse primer: 5′-CATGCTGGGCAATGTCTGTC-3′; *FBLN1* forward primer: 5′-CGTCTCCAGATGGGTATGGC-3′, and *FBLN1* reverse primer: 5′-CCCCTTGACTTTCCGAGACC-3′.

Construction of *ARHGAP31*- and FBLN1-expressing vectors

The C-terminus of coding sequences (CDS) for *ARHGAP31* (4335 bp) and *FBLN1* (2052 bp) were added with CDS for FLAG (DYKDDDDK) tag and hemagglutinin (HA, YPYDVPDYA) tag, respectively. *ARHGAP31* (c.2623G > A) and *FBLN1* (c.1649G > A) mutations were directly incorporated in the sequence design by substituting the corresponding nucleotide in wild-type sequences. The above full-length sequences with and without mutations were straightforwardly synthesized by TIANYI HUIYUAN biotechnology (Wuhan, China), using synthetic oligonucleotide primers-based PCR amplification. CDS for ARHGAP31-FLAG was then subcloned into eukaryotic-expressing vector pcDNA3.1 using HindIII/NotI restriction sites, while EcoRI/NotI sites were used for cloning of FBLN1-HA into pcDNA3.1. Successful insertions were verified by digestion using corresponding restriction enzymes followed by Sanger sequencing. To induce site-directed mutagenesis at c.2047C > T to construct the *p*. Gln683X mutant form of *ARHGAP31*, PCR amplification of the entire plasmid (pcDNA3.1-ARHGAP31-wild-type-FLAG) was performed using primer pair against c.2047C > T site (*p*. Gln683X-FP: 5′-CCAATTTAGCCTATTCTCGAGTCG-3′; *p*. Gln683X-RP: 5′-TAGGCTAAATTGGGCTGGTCTTCA-3′). PCR product was digested with DPN1 enzyme to eliminate original plasmid template before transformed into competent E. coli cells. Clones were sent for Sanger sequencing to confirm each single nucleotide including the mutation.

Cell culture and plasmid transfection

The human embryonic kidney HEK293T cell line and human cervical cancer HeLa cells line were obtained from China Center for Type Culture Collection (CCTCC). The cells were routinely maintained in Dulbecco's modified eagle medium at 37°C (5% CO_2) (Gibico). The medium was supplemented with 10% fetal calf serum (Life Technologies) and 1% penicillin–streptomycin (10,000 U/ml, Life Technologies). The cells were cultured in 96-well plates for cell proliferation assays, and in 12-well plates and Petri dishes for immunofluorescence and Western blotting analysis. The cells were plated onto plates for 18–24 h before transfection at a confluence of 60%–70% or onto coverslips as required.

Previously indicated plasmids and control vector (pcDNA3.1) were prepared and quantified before transfections. Prior to transfections, the cells were cultured to 70%–80% confluence. Transfection was conducted using polyethylenimine reagent (PEI, Polyscience). Briefly, for each well in a 12-well/plate, 1 μg DNA was diluted in 200 μl Opti-MEM (Gibco) followed by mixing with 1 μl PEI solution (5 mg/ml). Transfection mixture was incubated at room temperature for 20 min before added into each well. The cells were harvested at 24 h or 48 h post-transfection for immunofluorescence, Western blotting, and Cdc42 activity as indicated.

Cell counting Kit-8 assay

HeLa cell suspension was inoculated at 5,000 cells per well in a 96-well plate, which was later pre-incubated for 24 h in a humidified incubator at 37°C, 5.0% CO_2 with saturated humidity. Prepared expression plasmids were transfected using PEI reagent. For each well, 10 μl transfection mixture including 200 ng of total DNA and 0.5 μg PEI was incubated and added into the cell medium; 24 and 48 h after transfection, 10 μl of CCK-8 solution (Biosharp Life Science, BS350A) was added to

each well of the plate. Bubbles were avoided, and plates were incubated for 1 h before absorbance was measured at 450 nm with a microplate reader.

Quantitative real-time PCR

HeLa cells transfected with FBLN1 plasmids were collected in an RNA isolator (Vazyme Biotech, R401) for total RNA extraction, followed by reverse-transcription using a HiScript II 1st strand cDNA synthesis kit (Vazyme Biotech, R212-01). The resulting cDNA was used for qRT-PCR using a SYBR Green Master mix (Vazyme Biotech, Q221-01) and analyzed using the QUANTAGENE q225 real-time PCR machine (KUBO biotechnology). The following primer pairs were used: Primer Pair1 forward, 5′-GGATCTCTCTCGCCACGG-3′, Primer Pair1 reverse, 5′- TCAAGCGTATGTCTGGGAC-3′ and Primer Pair2 forward, 5′- CTCACCAAGCCTGTCCCC-3′, Primer Pair2 reverse, 5′- TCAAGCGTATGTCTGGGAC-3′. *GAPDH* forward, 5′- TGCACCACCAACTGCTTAGC-3′, *GAPDH* reverse, 5′- GGCATGGACTGTGGTCATGAG-3′. The 2-ΔΔCT method has been used for fold change analysis using *GAPDH* as internal control.

Protein stability analysis

Protein stability was assessed using protein translation inhibitor cycloheximide (CHX, Merck Sigma-Aldrich, C7698) treatment. HeLa cells were transfected with FBLN1 plasmids for 24 h and subjected to CHX treatment at final concentration of 50 μg/ml. The cells were collected and lyzed at indicated time points (0, 2, 4, 6, 8, and 12 h) after treatment, followed by immunoblotting.

Co-immunoprecipitation

Co-immunoprecipitation was performed following standard procedures. In brief, the cells were washed in PBS and lyzed in cold NP-40 buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40) supplemented with a protease/phosphatase inhibitor cocktail (Cell Signaling Technology). The lysate was briefly sonicated and centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was then collected and 1 mg of total protein lysate was incubated with protein A-conjugated agarose beads (Beyotime technology, P2051) and primary antibodies overnight at 4°C. Anti-FLAG antibody (Beyotime biotechnology, AF519, 1: 250 dilution) and anti-HA antibody (Beyotime biotechnology, AF0039, 1:250 dilution) were used. The beads were spun down, washed with PBS buffer, and denatured with 2× Laemmli sample buffer (Bio-Rad), followed by Western blotting for validation.

Western blotting

The cells were collected in 1× radioimmunoprecipitation assay (RIPA) buffer (Cell Signaling Technology) supplemented with a protease/phosphatase inhibitor cocktail (Beyotime biotechnology) or 2× Laemmli sample buffer (Bio-Rad). They were then thoroughly lyzed by sonication and heat denatured at 95°C for 5 min before loading onto gels for analysis. Equal amounts of protein were loaded (10–50 μg depending on the assay) and resolved by SDS-PAGE in Tris-glycine-SDS buffer, followed by transfer onto nitrocellular membranes (millipore). The membranes were blocked with blocking buffer (Beyotime P0256) for 0.5 h and then reacted with primary antibodies in primary antibody dilution buffer (Beyotime P0023A) overnight at 4°C under gentle rocking. Primary antibodies are as follows: anti-FLAG antibody (Beyotime biotechnology, AF519-1, 1: 2,000 dilution), anti-HA antibody (Beyotime biotechnology, AF0039, 1:2,000 dilution), anti-actin antibody (Proteinab Biotech, Cat#10003-M01, 1:5,000 dilution), anti-GAPDH antibody (EMD Millipore, AB2302, 1:5000 dilution), anti-Cdc42 antibody (Cytoskeleton, Cat#ACD03, 1:250 dilution), anti-pERK1/2 antibody (Cell signaling Technology, #4370, 1:2,000 dilution), and anti-ERK1/2 antibody (Cell signaling Technology, #4696, 1: 2,000 dilution). After washing, the corresponding IR Dye 680RD/800CW secondary antibodies (LICOR, 1:10,000 dilution) were added for 1 h at room temperature under constant agitation. After washing with Tris buffered saline wash buffer with Tween 20, the membrane was scanned in the Odyssey Fc Imaging System (LICOR) for detection of an infrared signal. For the Cdc42 activity experiment, the corresponding horseradish peroxidase (HRP)-conjugated anti-rabbit secondary antibodies (Beyotime biotechnology, 1:1000 dilution) was used, and the membrane was probed with enhanced chemiluminescence (ECL) Western blotting substrate (Beyotime biotechnology P0018AM) and visualized using a Tanon scanner system (Tanon5200).

Assessment of Cdc42 activity

Cdc42 activation assay was examined using the Cdc42 Activation Assay Biochem Kit (Cat. # BK034, cytoskeleton), following the manufacturer's instructions. HeLa cells were grown at approximately 70%–80% confluent prior to transfection. The cells were treated with HB-EGF (C600219, Sangon Biotech) for 3 min before lyzed in cell lysis buffer. Then, 10 μl PAK-PBD beads were added to equivalent amount of lysate (800 μg total protein) with gentle rotation for 1 h at 4°C. After three washes, the beads were resuspended and processed for immunoblotting using total Cdc42 rabbit monoclonal antibody (Beyotime biotechnology, AF2794, 1: 2,000 dilution).

Cell proliferation using EdU labeling

EdU cell proliferation image kit (green fluorescence) from Abbkine (Cat#. KTA 2030) was used, following manufacturer's instructions. Briefly, the cells grown on coverslips were transfected with desired plasmids for 24 h and were incubated with 10 μM EdU solution for 2 h. After incubation, the cells were fixed in 4% paraformaldehyde (PFA) for 10 min and were subjected to Click-iT reaction mix including copper reagent, AbFluor 488 azide, and reducing agent for 30 min. The cells were washed and counterstained with DAPI to visualize the nuclei. Fluorescence images were captured at 10× objective by using an inverted fluorescence microscope (ICX41, Sunny optical technology). The images were analyzed with ImageJ.

Live and dead cell double staining

Live and dead cell double staining kit from Abbkine (Cat#. KTA1001) was used, following manufacturer's instructions. Briefly, the cells were transfected as desired for 24 h; 1× assay buffer was prepared and warmed. LiveDye and NucleiDye were diluted at 1/1,000 dilution in assay buffer as staining solution. The cells were washed with PBS twice before incubated with staining solution at dark for 30 min. After staining, the cells were washed with PBS again and subjected for imaging immediately. Fluorescence images were captured at 10× objective by using an inverted fluorescence microscope (ICX41, Sunny optical technology). The images were analyzed with ImageJ.

TUNEL (TdT-mediated dUTP Nick End Labeling) assay

TUNEL apoptosis detection kit was purchased from Yeason Biotechnology (Cat. # 40306ES20). Briefly, the cells grown on coverslips were transfected with desired plasmids for 24 h before fixation at 4% PFA, followed by permeabilization with 0.3% Triton X-100 in PBS. For the positive control, permeabilized cells were with treated DNase I (10 U/ml) solution for 10 min at room temperature and washed with deionized water thoroughly. TdT-labeling mix was prepared with FITC-12-dUTP labeling mix and recombinant TdT enzyme in equilibration buffer. Negative control was performed using labeling mix without TdT enzyme. The cells were washed after labeling at 37°C for 1 h and counterstained with DAPI to visualize the nuclei. Fluorescence images were captured at 10× objective by using an inverted fluorescence microscope (ICX41, Sunny optical technology). The images were analyzed with ImageJ.

Immunofluorescence staining

Cells grown on coverslips were washed gently with phosphate buffered saline (PBS) and fixed in 4% PFA (except that cells were fixed in cold methanol for endogenous ARHGAP31 staining), followed by permeabilization with 0.3% Triton X-100 in PBS. The cells were then blocked for 1 h in 5% goat serum, followed by incubation overnight with primary antibodies at 4°C. Anti-FLAG (Beyotime biotechnology, AF519, 1:500 dilution), anti-GM130 (Sigma-Aldrich, #610822, 1:200 dilution), anti-ARHGAP31 (CdGAP (G-8), Santa Cruz, sc-393839, 1:25 dilution), anti-HA (Beyotime biotechnology, AF0039, 1:200 dilution), and antibodies were used. Secondary antibodies were Alexa Fluor (488, 555)-labeled goat antimouse and goat antirabbit antibodies (Thermo Fisher, 1:500 dilution). DAPI was used to visualize the nuclei. Fluorescence images were captured with a 20× or 63× objective using a Zeiss LSM780 laser scanning confocal microscope.

Statistical analysis

Statistical analysis was performed using Prism 9 software (GraphPad). The biochemical analysis was performed using a minimum of three biological replicates per condition, and randomization of experimental groups was not required. The data distribution was assumed to be normal with similar variance between groups; however, datasets were first assessed with normality tests. Two groups were analyzed with the unpaired Student's t test when the datasets were normally distributed, otherwise non-parametric tests were used with a non-paired Mann–Whitney test. Three or more groups were assessed with one-way ANOVA followed by a post hoc test for multiple comparisons. For statistical analysis of positivity in fluorescent images, Kruskal–Wallis test were used followed by Dunn's multiple comparisons test. Wild-type groups always serve as controls in multiple comparisons. All comparisons were performed whenever possible with all significant changes being indicated on the graphs unless changes are not significant. Changes close to significant are also presented. Values are presented as the mean ± standard error of mean (s.e.m).

Results

Clinical manifestations and identification of variations

The proband is a female who has normal intelligence and no obvious phenotypic abnormalities in cardiovascular and nervous systems. As indicated, the middle and distal phalanges of the left middle finger are missing, and the second finger on the left hand is completely absent; distal limb reduction appeared on the left foot in the first and second left toe; her second, third, fourth, and

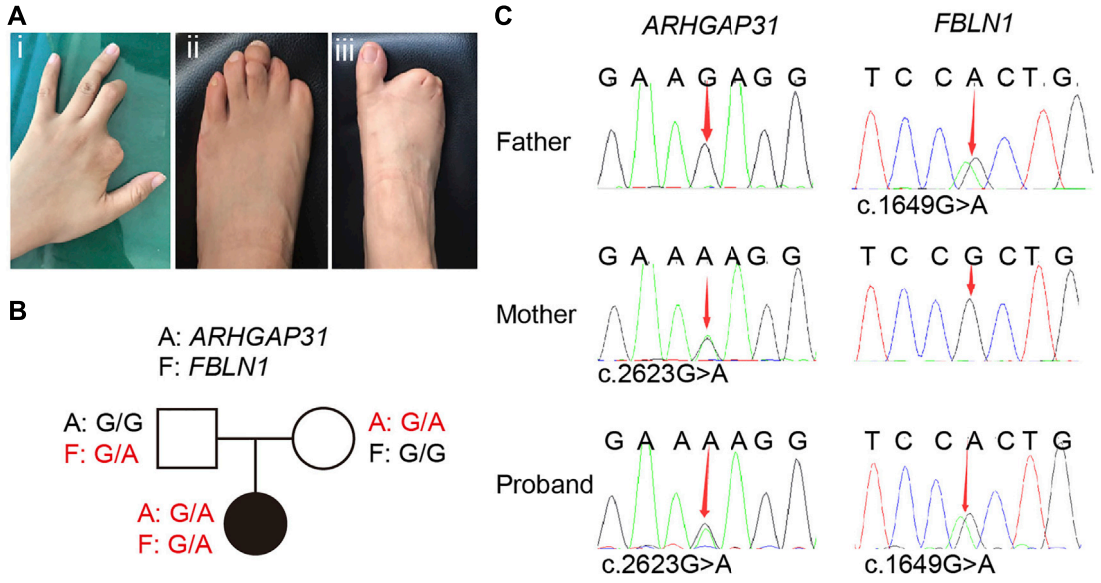


FIGURE 1
Clinical manifestations in the proband and related sequencing information. **(A)** Defects of the distal digits. (i) Second finger is completely absent, the middle and distal phalanges of the left third finger are missing; (ii) distal limb reduction appeared on the left foot in the first and second left toe with hypoplastic toenails on the third and fifth toes; (iii) her second, third, fourth, and fifth toes of the right foot are cutaneous syndactyly. **(B)** Family pedigree of the proband. **(C)** Sanger sequencing of the family members revealed inheritance of gene variants from her parents.

fifth toes in the right foot showed cutaneous syndactyly. In addition, the fourth finger on the left hand and the third toe on the left foot appear to have constriction rings (Figure 1A). No other abnormalities were observed at birth, and there was no other banding on the body elsewhere.

Her parents were assessed by a clinical geneticist and are clinically normal with no family history, we therefore first speculated typical inheritance of two recessive copies on one causative gene, or a *de novo* variation occurred in a pathogenic gene in the proband. Therefore, whole exome sequencing (WES) analysis were first performed in the proband. In general, a total of 25,539.75 Mb of original data were generated, about 98.33% of the target bases were covered with at least 20× per individual, and the mean depth of coverage for all target regions was 201.52. Surprisingly, WES specifically identified two independent variations in two genes, other than classic two copies of variation in one gene. Based on the database and annotation, the allele frequency of *ARHGAP31* (NM_020,754.4): c.2623G > A is absent in gnomAD and is 2/125,568 in TOPMed, while that of FBLN1(NM_006,486.3): c.1649G > A variant is 3/282,786 in gnomAD and is 1/264,690 in TOPMed. Next, the pathogenicity prediction indicates that the *ARHGAP31* variant is deleterious from SIFT (0.03) and benign from PolyPhen (0.211), while the FBLN1 variant is deleterious from SIFT (0.01) and probably damaging from PolyPhen (0.993). We then attempted to assess whether the variations are *de novo* in the family. Sanger sequencing confirmed that the proband carried each copy of the variation from one of the parents (Figure 1B), and the variations in *ARHGAP31* and *FBLN1* cause Glu875 to Lys (E875K) and Arg550 to His (R550H) missense mutations, respectively (Figure 1C). Since *ARHGAP31* is one of the causative genes identified for AOS, and *FBLN1* variation has been shown to cause syndactyly, we highly suspected the two variations together contributes to the TTLD phenotype in the proband.

Exogenous expression of *ARHGAP31* variant in cultured cells

Next, we attempted to investigate gene functions by overexpressing the encoding genes in HEK293T cells. As illustrated in Figure 2A, open reading frame sequences were fused with FLAG-tag. Clinically identified variations and corresponding mutated amino acid were indicated. Protein expression was validated by immunoblotting with anti-FLAG antibodies. Exogenous WT and mutant ARHGAP31 expression showed similar band patterns, with full-length ARHGAP31 at size around 200–250 kDa and a group of weaker cleaved bands below (Figure 2B). Size of full-length ARHGAP31 was validated with antibody against endogenous expression (data not shown). Expression levels were not altered between groups (Figure 2C).

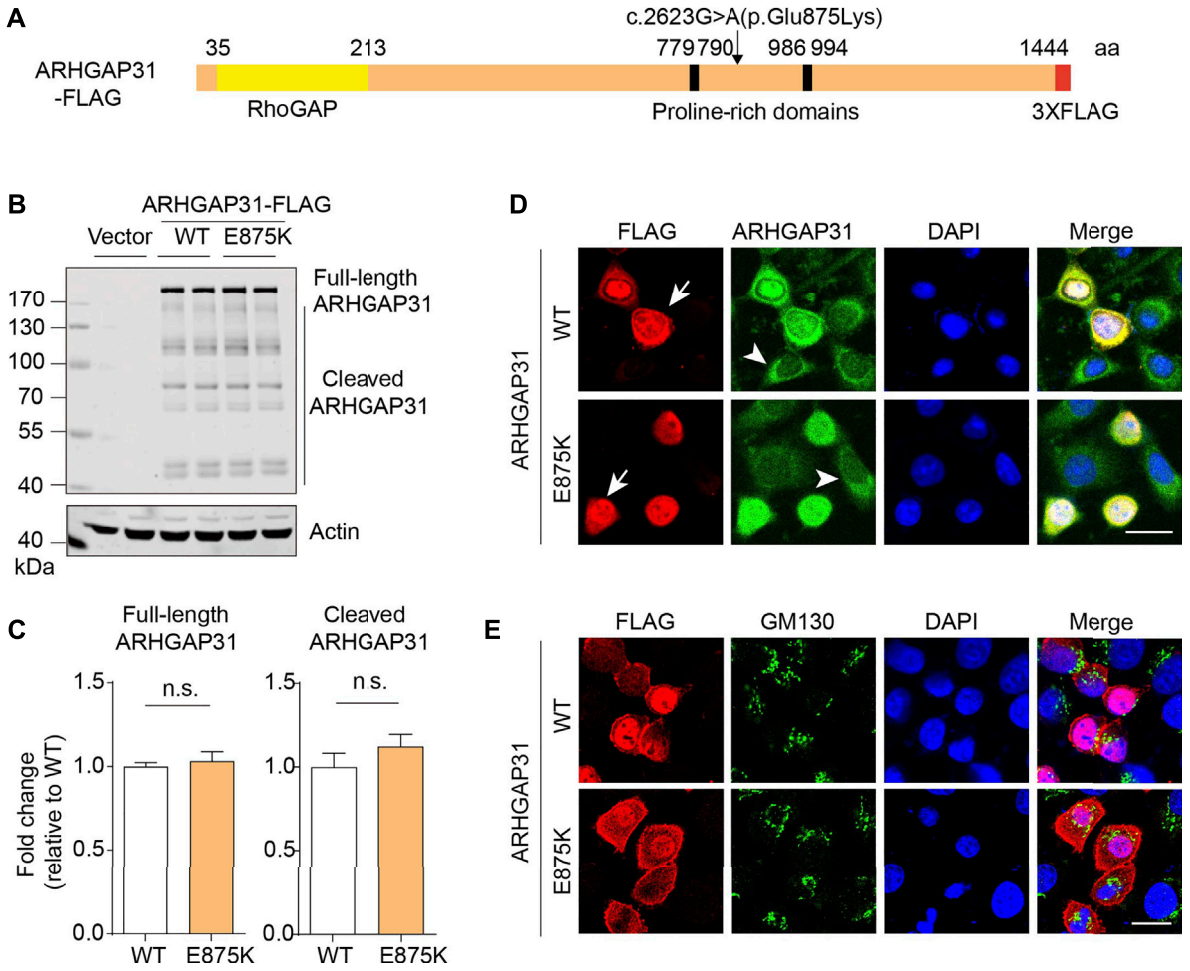


FIGURE 2
Exogenous expression and localization of ARHGAP31 in mammalian cells. **(A)** Schematic illustration of molecular structure of FLAG-tagged ARHGAP31. Missense mutation site was shown. **(B)** Representative immunoblots of exogenous ARHGAP31 expression in HeLa cells using anti-FLAG antibody. **(C)** Quantification of B (n = 4 per group). **(D)** Representative immunofluorescence co-staining of ARHGAP31 expression with anti-FLAG antibody and anti-ARHGAP31 antibody in transfected HeLa cells. Blue, DAPI-labeled nuclei. Arrow denotes ARHGAP31-overexpressing cells and arrowhead denote cells with endogenous ARHGAP31 expression. **(E)** Colocalization of ARHGAP31 and Golgi marker protein GM130 in transfected HeLa cells. Scale bar, 20 μm.

We further looked into the cellular localization of ARHGAP31 in HeLa cells. Anti-FLAG tag labeling and anti-ARHGAP31 staining indicated an obvious intracellular expression including a strong nuclear expression and diffuse cytosolic expression. Our ectopic ARHGAP31 was successfully detected (Figure 2D). Meanwhile, in both groups, cells bearing higher expression of ARHGAP31 appeared more roundish outline than those with lower expression. No obvious localization change was caused by mutation. Since Golgi-specific staining has been previously reported, we stained the Golgi maker protein GM130, and results show that ARHGAP31 was apparently not, least not exclusively, localized at Golgi department (Figure 2E).

Defective expression of FBLN1 mutant is caused by impaired protein stability

Similarly, we constructed FBLN1-HA plasmid as indicated (Figure 3A). Anti-HA immunoblotting revealed an obvious FBLN1 expression at around 90 kDa as predicated. Total protein expression of FBLN1 were significantly reduced by 25% (Figures 3B,C). Anti-HA immunostaining showed that FBLN1 expressed at perinuclear and cytoplasmic departments. A fibrillary-like pattern was observed in both WT and mutant FBLN1-transfected cells (Figure 3D).

To examine the reason for the defective FBLN1 mutant expression, we designed two pair of primers spanning the

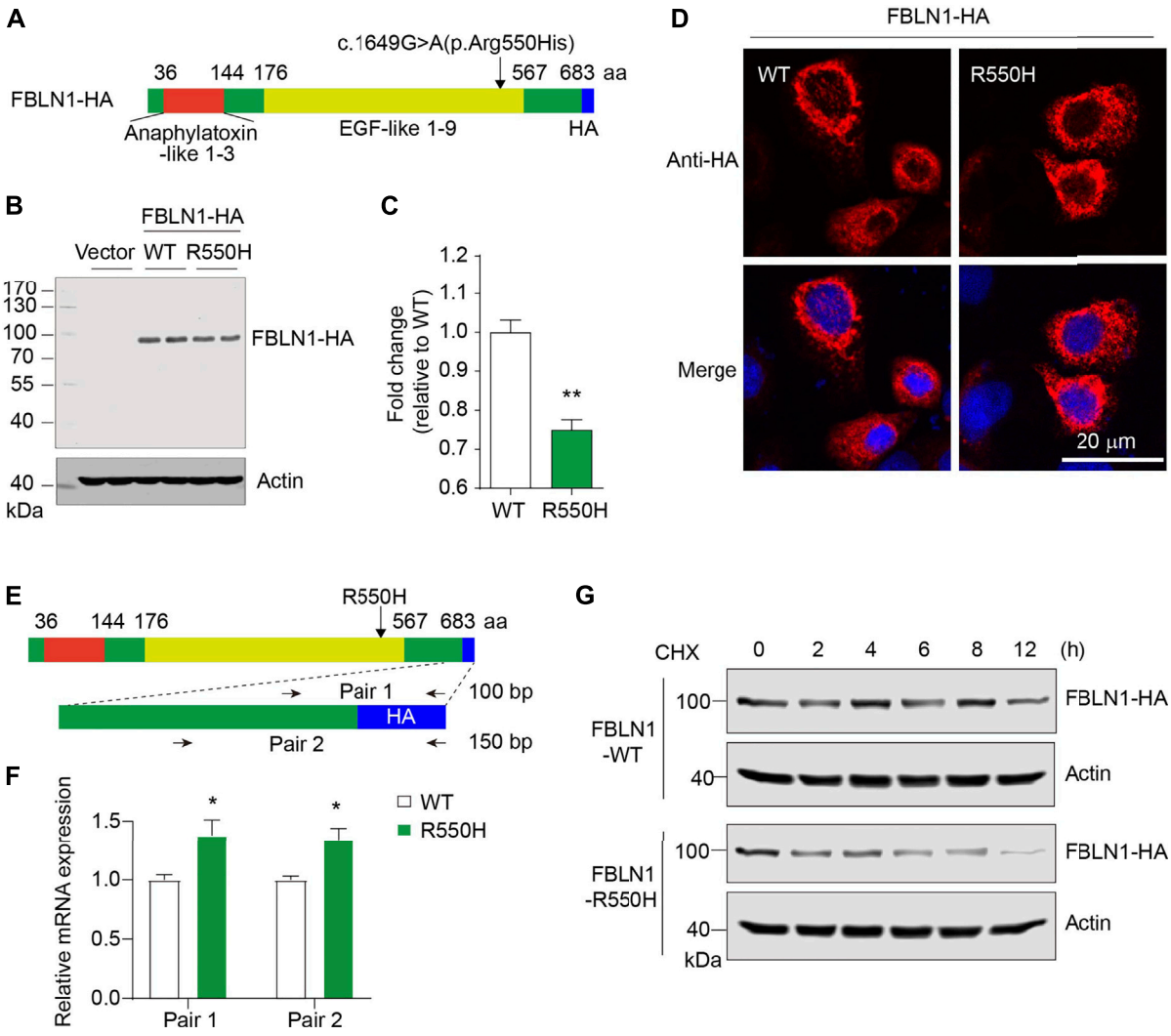


FIGURE 3
Defective expression of FBLN1 mutant is caused by impaired protein stability. **(A)** Schematic illustration of molecular structure of HA-tagged FBLN1. Missense mutation site was shown. **(B)** Representative immunoblots of exogenous FBLN1 expression in HeLa cells using anti-HA antibody. **(C)** Quantification of B (n = 3–4 per group). **$p < 0.01$. **(D)** Representative immunofluorescence staining of FBLN1-HA in transfected HeLa cells. **(E)** Primer design for quantitative real-time PCR at C-terminus of FBLN1 construct. **(F)** Quantitative real-time PCR results suggesting mutant FBLN1-transfected cells showed more mRNA expression. n = 3 per group. *$p < 0.05$. **(G)** Transfected HeLa cells were treated with protein translation inhibitor cycloheximide (CHX, 50 μg/ml) and subjected to immunoblotting.

coding sequence and the HA-encoding region in the construct (Figure 3E). Quantitative real-time PCR revealed that the R550H mutant form of FBLN1 have significantly higher mRNA levels (1.3-fold, Figure 3F). To further dissect the inconsistency, we investigated the protein stability by treating cells with protein translation inhibitor cycloheximide (CHX). As indicated, FBLN1 mutant showed apparently reduced total FBLN1 than the wild-type since 4 h after treatment, and the effect lasted until 12 h (Figure 3G). Hence, the decreased protein stability in the mutant may contribute to the deficiency of the ultimate protein expression.

Dysregulated protein expression by co-transfection of the variants

To examine possible effects by two mutant proteins together, we performed co-transfection in two types of mammalian cells. In FBLN1-WT-expressing HeLa cells, mutation in ARHGAP31 caused a mild but non-significant increase in full-length protein expression, and a significantly increased cleaved products ($p < 0.001$) (Figures 4A,B). Surprisingly, on the basis of ARHGAP31-E875K expression, all of the full-length, cleaved ARHGAP31 and FBLN1 expressions were significantly compromised in the

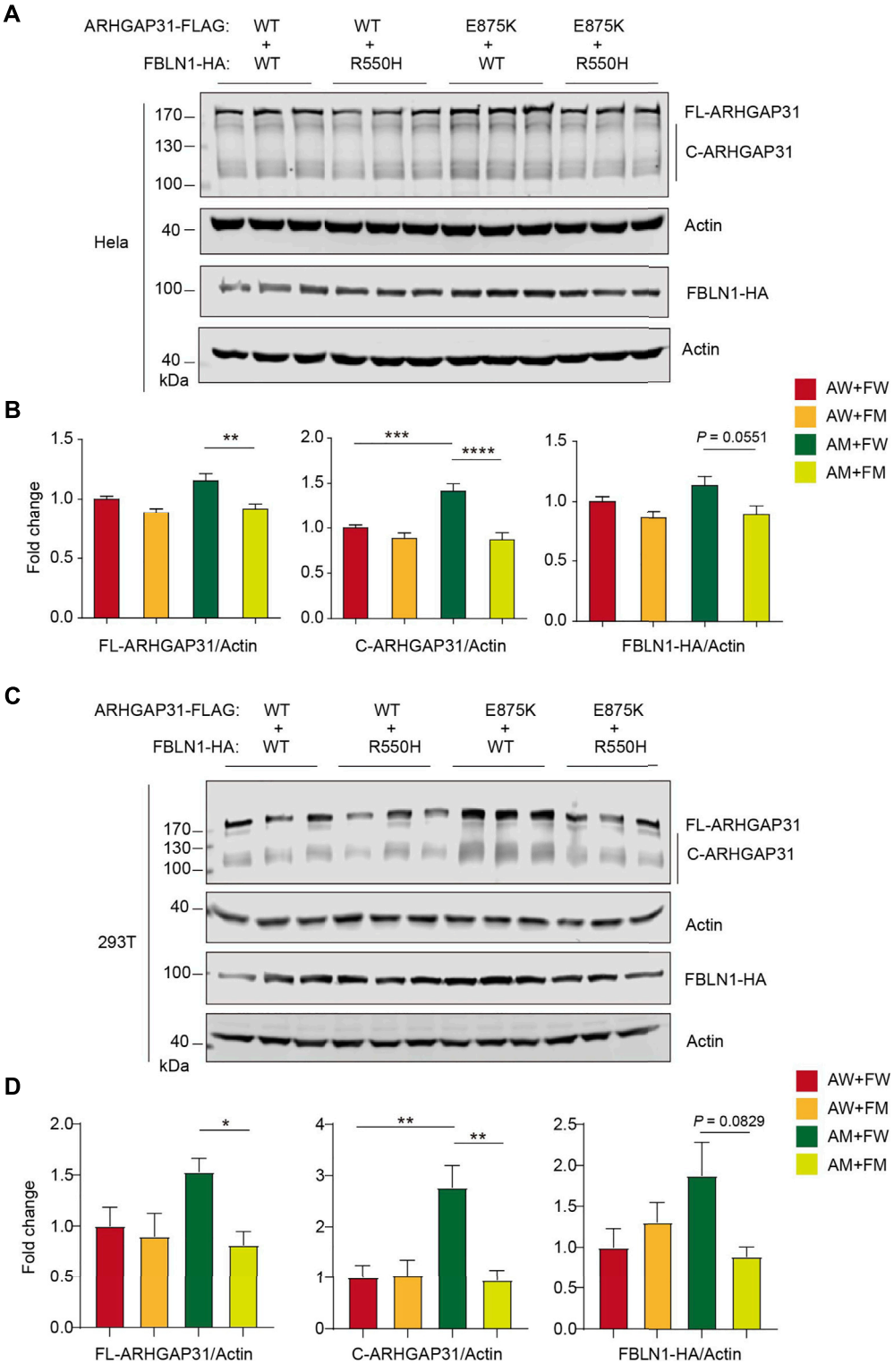


FIGURE 4
Protein expression with ARHGAP31 and FBLN1 variations *in vitro*. **(A)** Immunoblotting of exogenous ARHGAP31 and FBLN1 in co-transfected HeLa cells as indicated. **(B)** Quantification of **(A)**. **(C)** Immunoblotting of exogenous expression of ARHGAP31 and FBLN1 in co-transfected HEK293T cells. **(D)** Quantification of **(C)**. FL-ARHGAP31, full-length ARHGAP31; C-ARHGAP31, cleaved ARHGAP31; AW, wild-type ARHGAP31; AM, mutant ARHGAP31; FW, wild-type FBLN1; FM, and mutant FBLN1. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

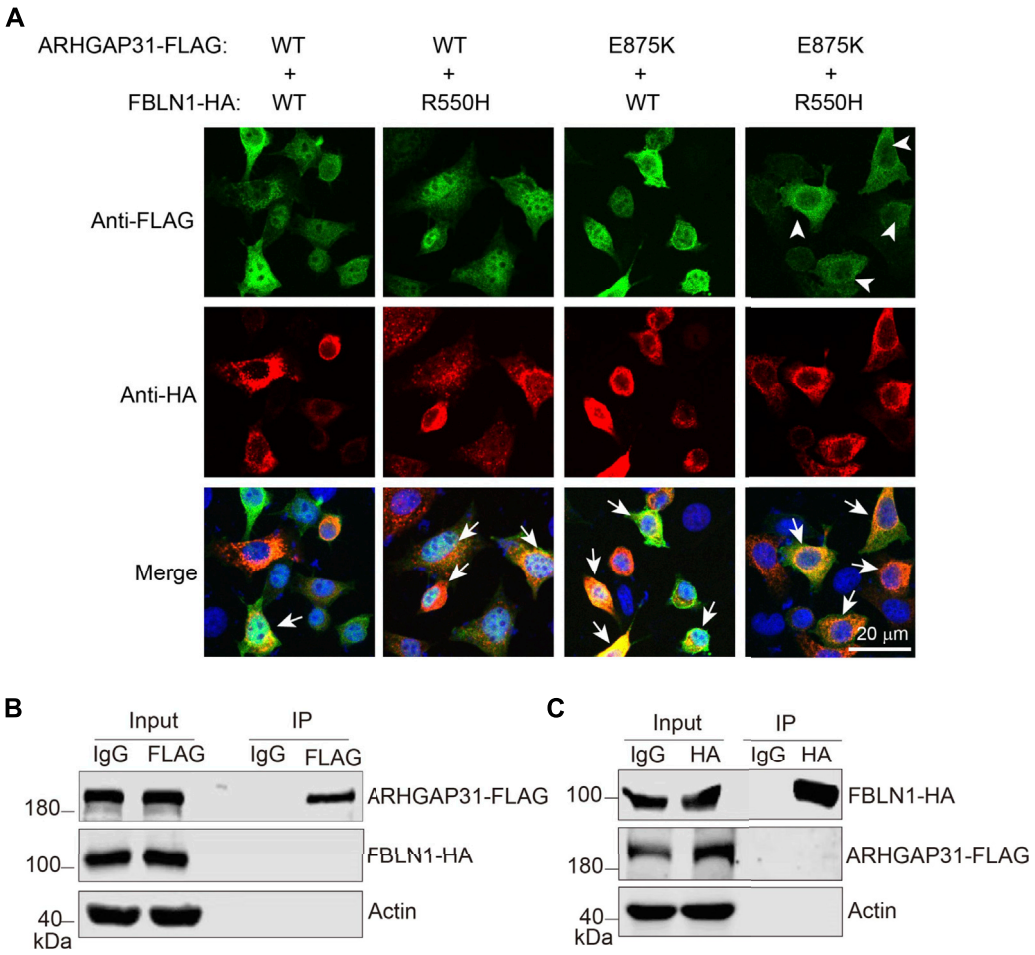


FIGURE 5
Increased colocalization with mutant ARHGAP31 and FBLN1. **(A)** Representative immunofluorescence staining of exogenous ARHGAP31 and FBLN1 in co-transfected HeLa cells. Green, anti-FLAG staining; red, anti-HA staining; and blue, DAPI-labeled nuclei. Arrowheads indicate cells with apparent non-nuclear expression of ARHGAP31, and arrows indicate recognizable cells with co-localized target proteins. **(B)** HeLa cells were co-transfected with ARHGAP31-FLAG and FBLN1-HA, and co-immunoprecipitation was performed using anti-FLAG antibody followed by detection with anti-HA antibody. **(C)** ARHGAP31-FLAG and FBLN1-HA co-transfected cells were lyzed, and co-immunoprecipitation was performed using anti-HA antibody followed by immunoblotting with anti-FLAG antibody.

FBLN1-R550H group compared with the FBLN1-WT group (Figures 4A,B), suggesting possible cytotoxic effects caused by FBLN1 mutant. More interestingly, when the same vector sets were co-transfected into HEK293T cells, we observed extremely similar alternation pattern as we have found in HeLa cells (Figures 4C,D).

Increased colocalization of *ARHGAP31* and FBLN1 mutants

We further investigated the subcellular localization of these variants in co-transfected HeLa cells. Compared to the intense nuclear localization of ARHGAP31 when it was transfected alone, apparent non-nuclear expression of ARHGAP31 appeared in both mutants-transfected cells. Overall, FBLN1 localization was not altered between groups. However, co-transfected of ARHGAP31-WT and FBLN1-WT showed only a few cells with relatively balanced expression of both proteins, as indicated by arrows; whereas in either of the mutants, colocalization of ARHGAP31 and FBLN1 were obviously increased (Figure 5A). Furthermore, enhanced co-localization of ARHGAP31 and FBLN1 mutants showed up in the cytosol, suggesting a potential interaction between the two proteins.

To examine if these two proteins directly interact with each other, we overexpressed two proteins in HeLa cells and performed co-immunoprecipitation. Although proteins were successfully immunoprecipitated by anti-tag antibodies, we

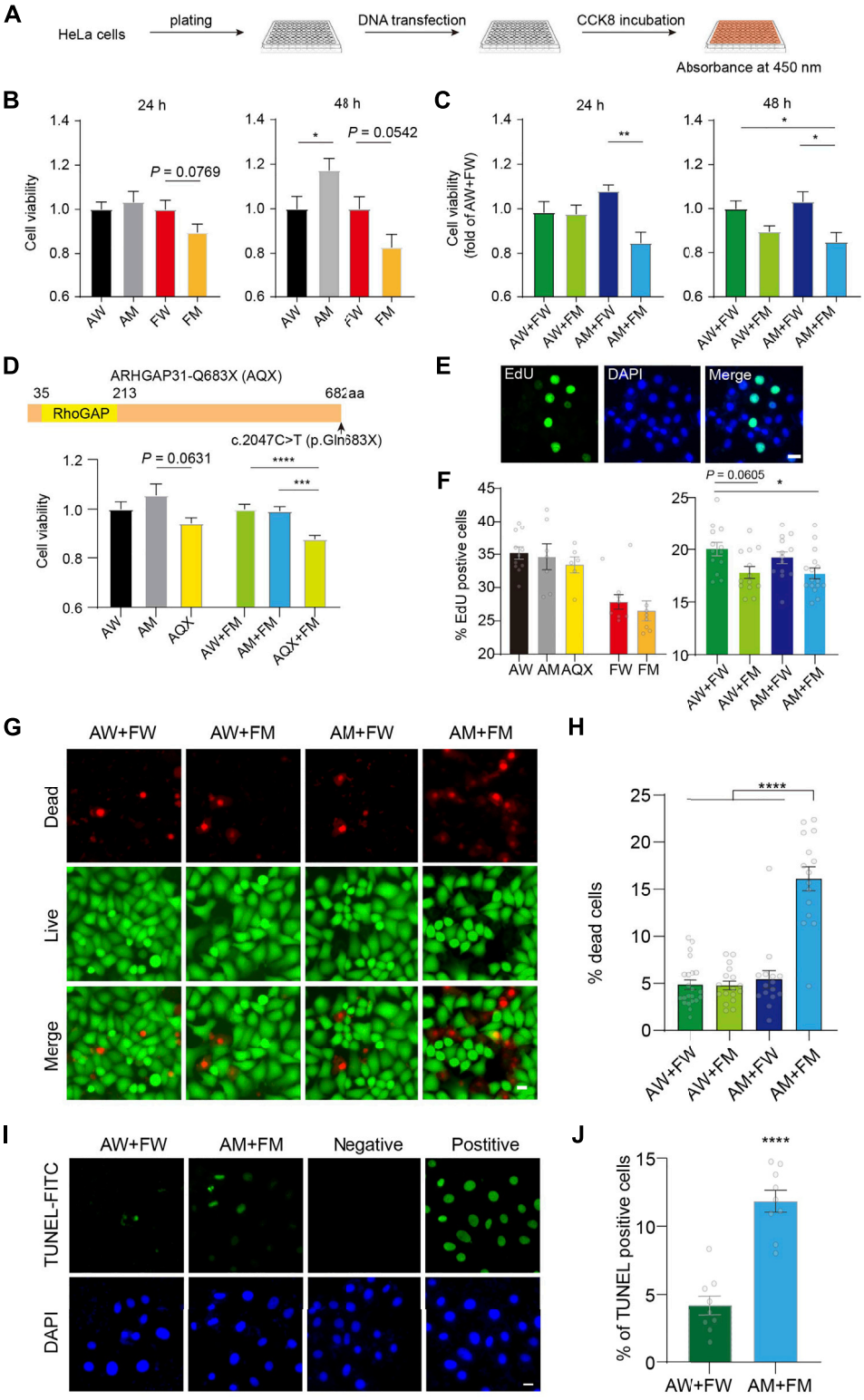


FIGURE 6
Enhanced cell toxicity and apoptosis with impaired cell proliferation induced by ARHGAP31 and FBLN1 mutants. **(A)** Flowchart demonstration of assessment of cell viability *in vitro* using CCK-8 assay. **(B)** Assessment of cell viability in HeLa cells 24 h and 48 h after single transfection of indicated vectors. **(C)** Assessment of cell viability in HeLa cells 24 h and 48 h after co-transfection of WT and mutant combinations. **(D)** Construction of ARHGAP31-p.Gln683X (AQX) truncation and CCK-8 assay at 24 h after indicated transfections in HeLa cells. **(E)** Representative images showing

(*Continued*)

FIGURE 6 (Continued)
EdU-positive cells, with DAPI labeling the nuclei. **(F)** Quantification of percentage of EdU-positive cells in transfected HeLa cells as indicated. n = 6–13 images including 1,700–5,000 cells per group. **(G)** Representative images showing co-stained live (green) and dead (red) HeLa cells after transfection with quantification in **(H)**. n = 15–22 images including 5,000–10,000 cells per group. **(I)** Representative images showing TUNEL-positive HeLa cells after transfection with quantification in **(J)**. n = 9 images including 800–900 cells per group. DAPI stains the nuclei. AW, wild-type ARHGAP31; AM, mutant ARHGAP31; FW, wild-type FBLN1; and FM, mutant FBLN1. Scale bar, 20 μm. Individual values are plotted in the bar chart. *p < 0.05, **p < 0.01. ***p < 0.001, and ****p < 0.0001.

were not able to detect any obvious interacting target (Figures 5B,C). Therefore, increased co-localization may be due to a direct interaction-independent manner.

Enhanced cytotoxicity and apoptosis with suppressed cell proliferation triggered by ARHGAP31 and FBLN1 mutants

To further test the functional consequences of the mutants, we performed CCK-8 assay, which examines combined effects of cell proliferation and cytotoxicity. As illustrated in Figure 6A, CCK-8 assays were conducted 24 h and 48 h after seeding and transfection. An increased cell viability with mutant ARHGAP31 (AM) was observed compared to the WT (AW) at 48 h ($p < 0.05$) but not 24 h, and mutant FBLN1 (FM) showed inhibitory effect trends toward significance on cell activity at both time points (p = 0.0769 and 0.0542, respectively, Figure 6B). After 24 h and 48 h with transfection combinations, combination of the mutant genes (AM + FM group) appeared significantly reduced cell viability compared to the AM + WT-FBLN1 (FW) group. In particular, a significant impairment of cell viability was observed in the AM + FM group at 48 h compared to the AW + FW group (Figure 6C). Other comparisons were not statistically significant.

Using the same method, we tested if a previously reported variant *p*. Gln683X of ARHGAP31 (AQX) in AOS exhibits toxicity in our *in vitro* model. Site-directed mutagenesis resulted truncated ARHGAP31 expression, and AQX-transfected HeLa cells showed decreased cell activity compared with AM-transfected cells (p = 0.0631) at 24 h post-transfection. More interestingly, when co-transfected with FM plasmid, AQX showed strong toxicity compared with AW and AM (Figure 6D), confirming that AM is less toxic than a known causative variant in AOS.

Given CCK-8 assay indicate a mixed reflection of cell activity, resulting from proliferation and cytotoxicity, we assessed cell proliferation directly using EdU labeling (Figure 6E). Statistical analysis suggested that neither ARHGAP31 nor FBLN1 is not affecting cell proliferation when transfected alone. However, when compared with the AW + FW group, AW + FM showed prone to significant effect in inhibition of cell proliferation; as expected, two mutants together (AM + FM) exhibit significant suppression in cell proliferation ($p < 0.05$, Figure 6F), suggesting a possible role of FBLN1 in regulating cell proliferation.

We next wondered whether cell death is also the cause for the declined cell activity. First, using cell membrane permeability-dependent dye, live and dead cells were stained in four co-transfected groups (Figure 6G). The number of dead cells in the AM + FM group were significantly higher than other groups ($p < 0.00001$, Figure 6H). Next, we directly labeled apoptotic cells using classic TUNEL assay, and the result showed apparently increased cell apoptosis in the AM + FM group compared to the AW + FW group (Figures 6I,J). Together, these results again suggest that ARHGAP31 and FBLN1 mutations may synergistically impair cell activity, which attributes to suppression of cell proliferation and augmented cell death particularly apoptosis.

Cdc42 inactivation and MAPK/ERK activation by ARHGAP31 and FBLN1 mutants

As Cdc42, a key regulator of cell cycle, is a putative target of GTPase-activating activity of ARHGAP31, we particularly examined Cdc42 expression and activity in mutants-transfected cells to further understand how mutant ARHGAP31 and FBLN1 contribute to apoptosis and cytotoxicity. Total Cdc42 levels were not altered by ARHGAP31 mutation, or FBLN1 mutation, or between co-transfected groups (Figures 7A,B). Next, we performed pull-down assay to analyze the level of active Cdc42 (GTP-bound form) in the cell lysate. In agreement with previous finding that AQX truncation showed decreased Cdc42 activity, E875K mutant in our study indicated similar inhibition of Cdc42 activation. Surprisingly, FBLN1 variation showed consistent inactivation of Cdc42. When co-transfected with AM, FBLN1 mutation further decreased the amount of active Cdc42 (Figures 7C,D).

In addition, we attempted to explore other possible targets for both ARHGAP31 and FBLN1. As a result, when comparing AM + FM to AW + FW groups, ARHGAP31 and FBLN1 mutants together showed significantly elevated the activation of MAPK/ERK pathway at both 24 h and 48 h after transfection (Figures 7E,F). Interestingly, a two-fold activation in the AW + FM group was observed at 24 h, indicating that MAPK/ERK activation may be initially driven by the FM variant. Since the MAPK/ERK signaling pathway is widely involved in cell activities including cell proliferation, differentiation, and cell apoptosis or anti-

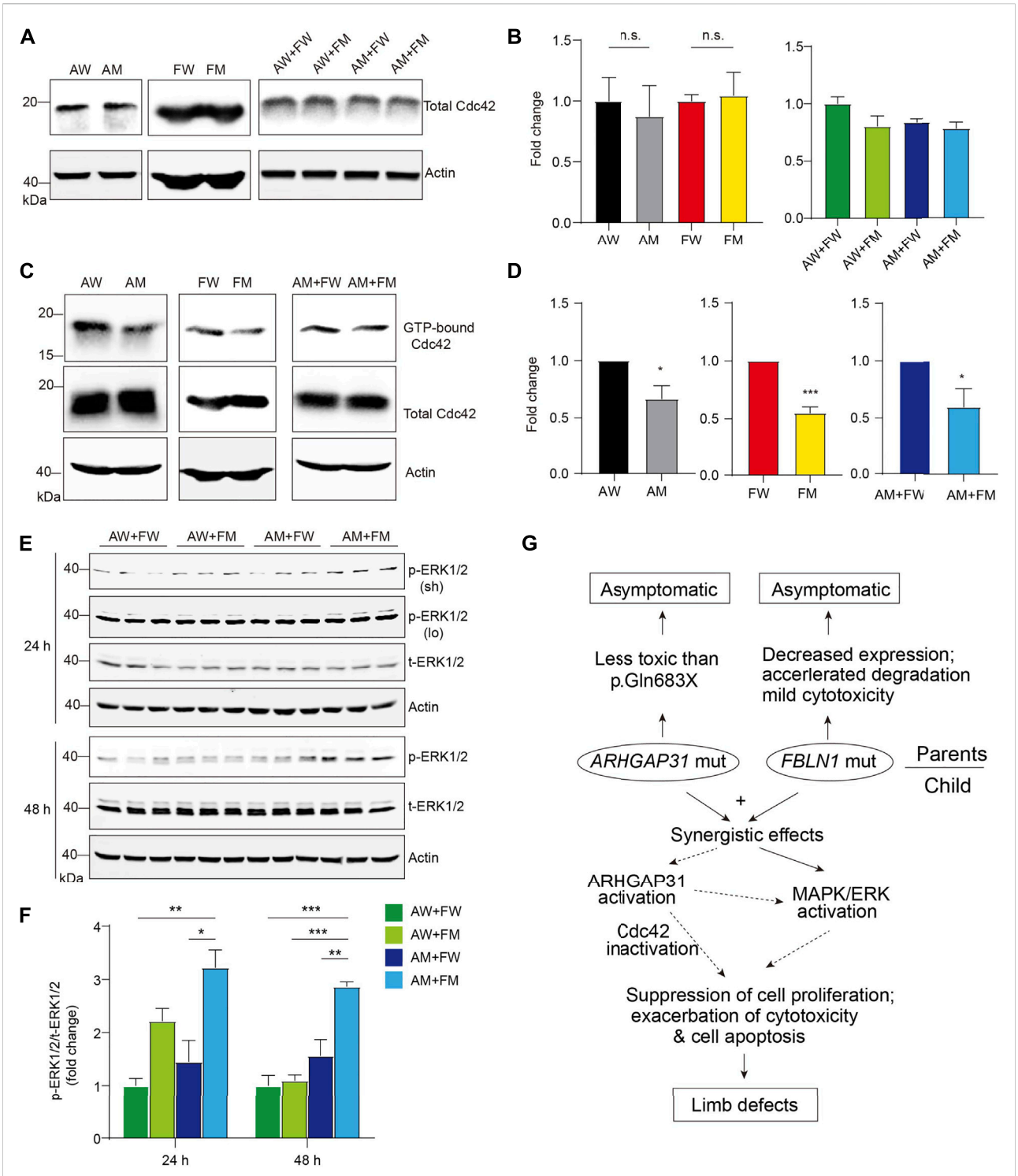


FIGURE 7
Enhanced activation of Cdc42 by ARHGAP31 and FBLN1 mutants. **(A)** Immunoblotting of Cdc42 expression in HeLa cell lysates after transfection. **(B)** Quantification of A (n = 3–6 per group). **(C)** Active Cdc42, that is, GTP-bound form, were detected using specific pull-down assay followed by immunoblotting of total Cdc42 in HeLa cell lysates post indicated transfections. All quantifications of respective densitometry are shown in **(D)** (n = 3–4 per group). **(E)** HeLa cells were transfected with indicated plasmids and immunoblotted with pERK1/2 and total-ERK1/2 (t-ERK1/2) antibodies. Sh, short exposure and lo, long exposure. **(F)** Quantification of E (results of short exposure were quantified). **(G)** Diagram of working hypothesis that the synergetic effects of variations in ARHGAP31 and FBLN1 cause penetrance of terminal transverse limb defects.

apoptotic functions, we thus speculate both Cdc42 and MAPK/ERK pathways play roles in synergistic effects of impaired cell activity and accelerated cell death *in vitro*, contributing to the clinical manifestation of limb defects (Figure 7G).

Discussion

Adams-Oliver syndrome is classically characterized by a combination of ACC and TTLDs, based on the diagnostic criteria proposed by Lehman in 2016 (Lehman et al., 1993). Six causative genes with variations have been identified in relation to AOS including *ARHGAP31*, *DOCK6*, *EOGT*, *RBPJ*, *NOTCH1*, and *DLL4*. In a four-generation family with *ARHGAP31* variation, all cases with *ARHGAP31* variations showed TTLD indicative of high penetrance, suggesting *ARHGAP31* variation may only cause isolated phenotype of TTLD in AOS (Isrie et al., 2014). Previously, all pathogenic variations in *ARHGAP31* reported are located within terminal exon 12, leading to premature termination of the translated protein or a missense mutation at C-terminus of the protein (Southgate et al., 2011; Isrie et al., 2014; Tao et al., 2021). Here in our case, the missense mutation (c.2623G > A (p.Glu875Lys)) occurs tween two proline-rich domains of the protein sequence in ARHGAP31 and is a rare variant ever been reported. The inherited copy in the daughter showed typical TTLD on here hands and feet, which are consistent with typical limb abnormalities in AOS (Isrie et al., 2014; Hassed et al., 2017). WES and Sanger sequencing excluded possibility of *de novo* variation or typical recessive inheritance of variants in single gene, therefore the transition of normal to abnormal expressivity in the family lead us to question the underlying cause. Since variation in *FBLN1* has been shown to be associated with complex forms of syndactyly (Debeer et al., 2002; Bohlega et al., 2014), the extra variation in *FBLN1* gene was therefore suspicious to be the other candidate as a causative gene. Due to the aforementioned points, we proposed that two gene variants together in the daughter may account for the dominant phenotype of limb malformations.

ARGHAP31 functions as a GAP by hydrating GTP to GDP, thereby reducing the level of GTP-bound Cdc42 and inactivating GTPase Cdc42/Rac1. Therefore, ARGHAP31 is required for cell spreading, polarized lamellipodia formation, and cell migration (Kurokawa et al., 2004; Tcherkezian et al., 2006). Southgate et al. (2011) demonstrated that truncated mutations in the last exon of *ARGHAP31* were pathogenic due to gain of function. Two such mutations, *p*. Gln683X (AQX named in our study) and *p*. Lys1087SerfsX4, led to generation of two fragments and decreased the amount of active Cdc42 levels. Here, we confirmed the toxicity of AQX using the CCK-8 assay and found that rare missense mutation E875K in ARHGAP31 showed less toxicity compared with AQX (Figure 6D). In addition, no effects on full-length ARHGAP31 expression, cleavage or cellular distribution were observed, although consistent spontaneous cleavage was observed throughout assays. To some degree, the unprevailing mutation may explain the normal phenotype in the mother as no apparent changes were observed by mutant ARHGAP31 in overexpressing cells. Labeling exogenous expression of ARHGAP31 with small FLAG tag, confirmed by anti-ARHGAP31 staining (Figure 2D), showed that the GAP protein is distributed at cytosolic prone to the plasma membrane, accompanied with nuclear localization, which was not altered by the E875K mutation. In disagreement, endogenous ARHGAP31 was reported to be localized to the Golgi (Southgate et al., 2011), and our result clearly showed that ARHGAP31 is not, least not exclusively, localized in the Golgi apparatus (Figure 2E). In addition, we believe that peri-membrane localization is more rational for the protein to function as a GAP when sensing broad extracellular signals.

Fibulins are extracellular glycoproteins secreted in elastic fibers and basement membranes of various tissues where they interact with several extracellular matrix components (De Vega et al., 2009). FBLN1-encoded Fibulin-1 is prominently expressed in the neural crest cells during development and is implicated in tissue organogenesis in developing myotomes, endocardial cushion and digits of the developing limbs (Cooley et al., 2008; De Vega et al., 2009). A translocation involving the last exon of FBLN1 isoform D has been associated with polysyndactyly possibly due to haploinsufficiency of isoform D (Debeer et al., 2002). Using a combination of homozygosity mapping and exome sequencing, the first point mutation in Fibulin-1 has also been confirmed in a novel autosomal recessive syndrome in development (Bohlega et al., 2014). In our study, HeLa cells expressing the rare variant of Fibulin-1 (R550H) showed compromised protein levels, but unaltered subcellular distribution. FBLN1-deficient embryos showed increased apoptosis in subpopulation of neural crest cells (Cooley et al., 2008). Inhibition of Fibulin-1 sensitized cancers cells to apoptotic signals, suggesting that FBLN1 serves a protective role from cell apoptosis (Gong et al., 2020). The R550H mutation-induced reduction of Fibulin-1 was moderated when co-expressed with wild-type ARHGAP31 but was consistently observed when mutant ARHGAP31 was co-transfected in both cell lines. In elucidation of the reason for reduced FBLN1 mutant expression, we found increased transcription levels for the mutant (Figures 3E,F); however, protein degradation rate was obviously faster in the mutant, suggesting the mutation-induced protein instability (Figure 3G). In conclusion, the reduction of Fibulin-1 widely observed in our study again is consistent with the indication that haploinsufficiency of Fibulin-1 may be able to initiate cell apoptosis during development.

Suggested by CCK-8 assay which indicates cell proliferation and cytotoxicity, FBLN1 mutation showed moderate cytotoxicity but has no effect on cell proliferation compared to the wild-type (Figures 6B,E,F); however, when cell viability and cell death were

investigated in co-transfections, synergistic effects of two mutants were consistently observed (AM + FM vs. AW + FW). In particular, cell apoptosis was confirmed in line with the cell death data. These findings are indeed intriguing and important in supporting the synergistic effects by two mutants in both *in vitro* and in the patient. Previously, few evidence elucidating functions of ARHGAP31 in AOS or FBLN1 in synpolydactyly have been reported. Combining our preliminary data of sub-localization distribution, it is suggested that these two proteins may have inter-molecular interactions, which may be enhanced by mutant epitopes, as elevated co-localizations with mutant proteins were found. *In vivo* evidence on their interactions might be supportive. More interestingly, functional assessments by CCK-8 assay, cell proliferation, and cell apoptosis together suggest a global loss of function in group co-expressing two mutants, coincided with the penetrance of TTLD in the mutant carrier.

As a GAP, ARHGAP31 inactivates Cdc42/Rac1 by dehydrating GTP to GDP that were bound by Cdc42 (Caron et al., 2016). The truncated form of C-terminus of ARHGAP31 boosts the activity of the RhoGAP domain by specific interaction with 1–221aa fragment, leading to decreased activity of Cdc42/Rac1 (Southgate et al., 2011). Rho GTPases such as Cdc42 regulate numerous cell functions, including cell cytoskeleton organization, migration, gene transcription, adhesion, cellular proliferation, and survival (Mosaddeghzadeh and Ahmadian, 2021). Pathogenic Cdc42 variant has been recently reported in AOS and other developmental disorders (Martinelli et al., 2018; Schnabel et al., 2021). Consistently, using PAK-PBD beads, we successfully pulled down active GTP-bound Cdc42 and found its activity was decreased by ARHGAP31 mutant, FBLN1 mutant, and the two mutants together (Figures 7C,D). It is possible that single variation induced overactivation of ARHGAP31, leading to Cdc42 inactivation, which was further augmented in the presence of mutant FBLN1, leading to a further reduction of activating Cdc42. Surprisingly and intriguingly, we disclosed significant activation of MAPK/ERK presenting the AM + FM group, further supporting the synergistic effects observed by the functional analysis (Figure 7E).

In summary, our findings identified two rare heterozygous variants in AOS risk gene *ARHGAP31* and synpolydactyly risk gene *FBLN1* in one case with TTLD deformation. Heterozygous carrier of individual variant shows no phenotype whereas two variants together resulted in typical limb defects. *In vitro* assays indicate that only two mutants could synergistically exhibit decreased cell viability, impaired cell proliferation, and increased apoptosis activation, possibly triggered by Cdc42 inactivation and MAPK/ERK activation, which both mutants have effects on (Figure 7G). Taken together, our study provides evidence of a synergistic disruption of cellular functions attributed by two independent risk gene variants, expanding the clinical spectrums of causal gene interactions in hereditary TTLD and AOS etiology.

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.ddbj.nig.ac.jp/, LC700997 https://www.ddbj.nig.ac.jp/, LC700998 https://www.ddbj.nig.ac.jp/, LC700999 https://www.ddbj.nig.ac.jp/, LC701000.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

HT and CL designed the study. HT and CL provided materials and funding. Data collection and analysis were performed by HT, FC, MX, YL, and WL. The draft of the manuscript was written by HT and CL, and all authors approved the final manuscript by reading the manuscript.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (31900692) to CL, and grants from the National Natural Science Foundation of China (31271043) to HT.

Acknowledgments

We thank the family for their interest and support. We thank Zhijun Liu for critically reading the manuscript.

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

Adams, F. O. C., and Oliver, C. P. (1945). Hereditary deformities in man. *J. Hered.* 36, 3–7. doi:10.1093/oxfordjournals.jhered.a105415

Bohlega, S., Al-Ajlan, H., and Al-Saif, A. (2014). Mutation of fibulin-1 causes a novel syndrome involving the central nervous system and connective tissues. *Eur. J. Hum. Genet.* 22, 640–643. doi:10.1038/ejhg.2013.210

Caron, C., DeGeer, J., Fournier, P., Duquette, P. M., Luangrath, V., Ishii, H., et al. (2016). CdGAP/ARHGAP31, a Cdc42/Rac1 GTPase regulator, is critical for vascular development and VEGF-mediated angiogenesis. *Sci. Rep.* 6, 27485. doi:10.1038/srep27485

Cooley, M. A., Kern, C. B., Fresco, V. M., Wessels, A., Thompson, R. P., McQuinn, T. C., et al. (2008). Fibulin-1 is required for morphogenesis of neural crest-derived structures. *Dev. Biol.* 319, 336–345. doi:10.1016/j.ydbio.2008.04.029

de Vega, S., Iwamoto, T., and Yamada, Y. (2009). Fibulins: Multiple roles in matrix structures and tissue functions. *Cell. Mol. Life Sci.* 66, 1890–1902. doi:10.1007/s00018-009-8632-6

Debeer, P., Schoenmakers, E. F., Twal, W. O., Argraves, W. S., De Smet, L., Fryns, J. P., et al. (2002). The fibulin-1 gene (FBLN1) is disrupted in a t(12;22) associated with a complex type of synpolydactyly. *J. Med. Genet.* 39, 98–104. doi:10.1136/jmg.39.2.98

Gong, J., Jie, Y., Xiao, C., Zhou, W., Li, X., Chen, Y., et al. (2020). Increased expression of fibulin-1 is associated with hepatocellular carcinoma progression by regulating the notch signaling pathway. *Front. Cell Dev. Biol.* 8, 478. doi:10.3389/fcell.2020.00478

Hassed, S., Li, S., Mulvihill, J., Aston, C., and Palmer, S. (2017). Adams-Oliver syndrome review of the literature: Refining the diagnostic phenotype. *Am. J. Med. Genet.* 173, 790–800. doi:10.1002/ajmg.a.37889

Isrie, M., Wuyts, W., Van Esch, H., and Devriendt, K. (2014). Isolated terminal limb reduction defects: Extending the clinical spectrum of Adams-Oliver syndrome andARHGAP31mutations. *Am. J. Med. Genet.* 164, 1576–1579. doi:10.1002/ajmg.a.36486

Kurokawa, K., Itoh, R. E., Yoshizaki, H., Nakamura, Y. O., and Matsuda, M. (2004). Coactivation of Rac1 and Cdc42 at lamellipodia and membrane ruffles induced by epidermal growth factor. *MBoC* 15, 1003–1010. doi:10.1091/mbc.e03-08-0609

Küster, W., Lenz, W., Kääriäinen, H., and Majewski, F. (1988). Congenital scalp defects with distal limb anomalies (Adams-Oliver syndrome): Report of ten cases and review of the literature. *Am. J. Med. Genet.* 31, 99–115. doi:10.1002/ajmg.1320310112

Lehman, A., Wuyts, W., Patel, M. S., Adam, M. P., Ardinger, H. H., Pagon, R. A., et al. (1993). "Adams-oliver syndrome," in *GeneReviews*. Seattle (WA)).

Martinelli, S., Krumbach, O. H. F., Pantaleoni, F., Coppola, S., Amin, E., Pannone, L., et al. (2018). Functional dysregulation of CDC42 causes diverse developmental phenotypes. *Am. J. Hum. Genet.* 102, 309–320. doi:10.1016/j.ajhg.2017.12.015

Meester, J. a. N., Sukalo, M., Schroder, K. C., Schanze, D., Baynam, G., Borck, G., et al. (2018). Elucidating the genetic architecture of Adams-Oliver syndrome in a large European cohort. *Hum. Mutat.* 39, 1246–1261. doi:10.1002/humu.23567

Melendez, J., Grogg, M., and Zheng, Y. (2011). Signaling role of Cdc42 in regulating mammalian physiology. *J. Biol. Chem.* 286, 2375–2381. doi:10.1074/jbc.R110.200329

Mosaddeghzadeh, N., and Ahmadian, M. R. (2021). The RHO family GTPases: Mechanisms of regulation and signaling. *Cells*10.doi:10.3390/cells10071831

Schnabel, F., Kamphausen, S. B., Funke, R., Kaulfuss, S., Wollnik, B., and Zenker, M. (2021). Aplasia cutis congenita in a CDC42-related developmental phenotype. *Am. J. Med. Genet. A* 185, 850–855. doi:10.1002/ajmg.a.62009

Snape, K. M., Ruddy, D., Zenker, M., Wuyts, W., Whiteford, M., Johnson, D., et al. (2009). The spectra of clinical phenotypes in aplasia cutis congenita and terminal transverse limb defects. *Am. J. Med. Genet. A* 149A, 1860–1881. doi:10.1002/ajmg.a.32708

Southgate, L., Machado, R. D., Snape, K. M., Primeau, M., Dafou, D., Ruddy, D. M., et al. (2011). Gain-of-function mutations of ARHGAP31, a Cdc42/Rac1 GTPase regulator, cause syndromic cutis aplasia and limb anomalies. *Am. J. Hum. Genet.* 88, 574–585. doi:10.1016/j.ajhg.2011.04.013

Tao, Z., Bu, S., and Lu, F. (2021). Two AOS genes attributed to familial exudative vitreoretinopathy with microcephaly: Two case reports. *Med. Baltim.* 100, e24633. doi:10.1097/MD.0000000000024633

Tcherkezian, J., Triki, I., Stenne, R., Danek, E. I., and Lamarche-Vane, N. (2006). The human orthologue of CdGAP is a phosphoprotein and a GTPase-activating protein for Cdc42 and Rac1 but not RhoA. *Biol. Cell* 98, 445–456. doi:10.1042/BC20050101

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fmolb.2022.991308

OPEN ACCESS

EDITED BY
Qi Wu,
Aarhus University, Denmark

REVIEWED BY
Hanting Yang,
Fudan University, China
Wenchao Cai,
Shihezi University, China

*CORRESPONDENCE
Andy T. Y. Lau,
andytylau@stu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Protein Biochemistry for Basic and Applied Sciences,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 11 July 2022
ACCEPTED 13 September 2022
PUBLISHED 18 October 2022

CITATION
Tan HW, Xu Y-M, Liang Z-L, Cai N-L, Wu Y-Y and Lau ATY (2022), Single-gene knockout-coupled omics analysis identifies C9orf85 and CXorf38 as two uncharacterized human proteins associated with ZIP8 malfunction. *Front. Mol. Biosci.* 9:991308. doi: 10.3389/fmolb.2022.991308

Single-gene knockout-coupled omics analysis identifies C9orf85 and CXorf38 as two uncharacterized human proteins associated with ZIP8 malfunction

Heng Wee Tan †, Yan-Ming Xu †, Zhan-Ling Liang †, Na-Li Cai , Yu-Yao Wu and Andy T. Y. Lau *

Laboratory of Cancer Biology and Epigenetics, Department of Cell Biology and Genetics, Shantou University Medical College, Shantou, Guangdong, China

Human transmembrane protein metal cation symporter ZIP8 (SLC39A8) is a member of the solute carrier gene family responsible for intracellular transportation of essential micronutrients, including manganese, selenium, and zinc. Previously, we established a ZIP8-knockout (KO) human cell model using the CRISPR/Cas9 system and explored how the expression of ZIP8 could possibly contribute to a wide range of human diseases. To further assess the biophysiological role of ZIP8, in the current study, we employed isobaric tags for relative and absolute quantitation (iTRAQ) and detected the changes of the proteome in ZIP8-KO cells (proteomic data are available *via* ProteomeXchange with identifier PXD036680). A total of 286 differentially expressed proteins (206 downregulated and 80 upregulated proteins) were detected in the ZIP8-KO cell model, and subsequent bioinformatics analyses (GO, KEGG, KOG, and PPI) were performed on these proteins. Interestingly, four "uncharacterized" proteins (proteins with unknown biological function) were identified in the differentially expressed proteins: C1orf198, C9orf85, C17orf75, and CXorf38—all of which were under-expressed in the ZIP8-KO cells. Notably, C9orf85 and CXorf38 were amongst the top-10 most downregulated proteins, and their expressions could be selectively induced by essential micronutrients. Furthermore, clinical-based bioinformatic analysis indicated that positive correlations between the gene expressions of *ZIP8* and *C9orf85* or *CXorf38* were observed in multiple cancer types. Overall, this study reveals the proteomic landscape of cells with impaired ZIP8 and uncovers the potential relationships between essential micronutrients and uncharacterized proteins C9orf85 and CXorf38. The differentially expressed proteins identified in ZIP8-KO cells could be the potential targets for diagnosing and/or treating human ZIP8-associated diseases, including but not limited to malnutrition, viral infection, and cancers.

KEYWORDS

C9orf85, CXorf38, iTRAQ, proteomics, ZIP8

Introduction

The human ZRT/IRT-like protein 8 (ZIP8), encoded by the solute carrier family 39 member 8 gene *SLC39A8*, is a metal cation symporter located mainly on the cell membrane (Zang et al., 2016). ZIP8 is well-known for its role in transporting divalent metal ions into the cells—these ions include several essential micronutrients (e.g., iron [Fe], manganese [Mn], and zinc [Zn]) and non-essential toxic heavy metal cadmium (Cd) (Wang et al., 2012; Xu et al., 2017).

A balanced level of essential micronutrients is crucial for maintaining human health (Mezzaroba et al., 2019; Tan et al., 2019). Studies have indicated that cells with impaired ZIP8 are often associated with disrupted micronutrient homeostasis, which may lead to a wide range of human disorders (Boycott et al., 2015; Park et al., 2015). It has been documented that impaired ZIP8, usually caused by mutations in the *SLC39A8* gene, is responsible for diseases such as type II congenital disorder of glycosylation (Riley et al., 2017; Choi et al., 2018), cardiovascular diseases (Zhang et al., 2016), severe idiopathic scoliosis (Haller et al., 2018), schizophrenia (McCoy et al., 2019), Crohn's disease (chronic inflammation of the digestive system) (Li et al., 2016), and Leigh syndrome (a severe inherited neurodegenerative disease) (Choi et al., 2018). These disease-associated *SLC39A8* gene mutations include 97G>A, 112G>C, 338G>C, 610G>T, 1004G>C, 1019T>A, and 1171G>A (Boycott et al., 2015; Park et al., 2015; Park et al., 2018).

Recently, we used CRISPR/Cas9 genome editing technology to knockout (KO) the *SLC39A8* gene in the human cervical cancer HeLa cell line (Liang et al., 2021). We then studied the micronutrient transport ability of the single-gene KO cell model and discovered that the elimination of ZIP8 could result in not only the reduced cellular uptake of the above-mentioned metals, but also reduced uptake of another essential micronutrient, selenium (Se) (Liang et al., 2021). Furthermore, by utilizing clinical datasets of 40 different types of cancers from The Cancer Genome Atlas (TCGA) database, we showed that *SLC39A8* gene expressions tend to be upregulated in a great number of tumor types (Liang et al., 2021). This data suggested that ZIP8 could be a novel molecular target in preventing or treating human cancers and/or illnesses related to Cd exposure. However, the proteome in cells with impaired ZIP8 has remained uninvestigated.

In this study, we sought to better elucidate the biophysiological role of ZIP8 by carrying out a single-gene KO-coupled omics (SGKOmics) analysis on our established ZIP8-KO cell model. Specifically, proteomes of the established ZIP8-KO and ZIP8-wildtype (WT) cell lines were quantified using isobaric tags for relative and absolute quantitation (iTRAQ) and liquid chromatography-tandem mass spectrometry (LC-MS/MS). Differentially expressed proteins between the two cell lines were identified, and subsequent bioinformatics analyses were performed. Among the differentially expressed proteins, four were recognized as the "uncharacterized proteins" with unknown biological functions: C1orf198, C9orf85, C17orf75, and CXorf38. Notably, C9orf85 and CXorf38 were two of the top-10 most downregulated proteins. It has so far remained unknown whether the expressions of C9orf85 and CXorf38 could be affected by ZIP8-transportable essential micronutrients (e.g., Mn, Se, and Zn), and therefore, we further examined the expressions of these two uncharacterized proteins upon essential micronutrient treatments. In addition, we performed clinical-based bioinformatic analysis to evaluate the potential connections between the gene expressions of *ZIP8* and *C9orf85* or *CXorf38* in multiple cancer types. Overall, findings from this study provide insights into the underlying mechanisms of ZIP8 deficiency-associated diseases by indicating that cell line with impaired ZIP8 contains an aberrant protein profile, which may be due in part to imbalanced levels of intracellular micronutrients.

Materials and methods

Cell lines and culture conditions

Human cervical cancer (HeLa) cell line was purchased from the American Type Culture Collection (ATCC) (Rockville, MD, United States). A ZIP8 single-gene KO HeLa cell model was established using the CRISPR/Cas9 genome editing technology, as described in Liang et al. (2021). All cell lines used in the current study were routinely cultured in MEM medium containing 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C in a 5% CO_2 incubator as recommended by ATCC.

Protein sample preparation

HeLa parental cell line (with ZIP8-WT) and ZIP8-KO HeLa cell line were grown to 80%–85% confluency (approximately 1×10^7 cells per sample) in 10 cm diameter Petri dishes prior to protein extraction. Samples ready to be extracted were placed on ice and washed with 5 ml pre-chilled phosphate-buffered saline (PBS) for six times. Then, 1 ml PBS per dish was added, and cells were scratched with a sterile plastic scraper. The cell suspension was then transferred to a 1.5 ml Eppendorf tube and centrifuged at 1,200 × *g* for 5 min at 4°C. After centrifugation, the supernatant was carefully removed, and the sample was stored at −80°C.

Protein extraction and digestion for iTRAQ analysis

Extraction and digestion of protein were performed by Wininnovate Bio Co., Ltd (Shenzhen, China). Briefly, frozen

cell samples were lysed in RIPA lysis buffer (0.1% SDS, 1% Triton X-100, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 50 mM Tris-HCl pH 7.4) containing PhosSTOP™ protease inhibitor cocktail. The samples were subsequently homogenized by sonication on ice with 15% ultrasound power under the repeated settings of "2 s on and 3 s off" for 4 min. The homogenate was then cleared using centrifugation at 12,000 rpm for 15 min at 4°C. Next, supernatants were transferred into clean tubes, and protein concentrations were determined using the Pierce™ BCA Protein Assay Kit. For protein digestion, 150 µg of protein from each lysate was mixed with solutions of 8 M urea, 0.1 M Tris-HCl, 0.1 M dithiothreitol, and 5 mM iodoacetamide, as described in Wiśniewski et al. (2009). Nanosep® centrifugal devices were used to centrifuge the samples.

iTRAQ labeling and LC-MS/MS analysis

iTRAQ and LC-MS/MS analysis were performed by Wininnovate Bio Co., Ltd (Shenzhen, China). Briefly, samples were labeled with iTRAQ reaction reagents according to the manufacturer's instructions. Each iTRAQ reagent was dissolved in 70 µl of isopropanol and added to the respective peptide mixture for 120 min. The labeling reaction was quenched by the addition of 100 µl of Milli-Q® H_2O, and all labeled samples were then pooled into one sample. The ZIP8-WT control samples (H1, H2, and H3) were labeled with tag-113, -114, and -115, while the ZIP8-KO samples (Z1, Z2, and Z3) were labeled with tag-116, -117, and -118.

LC-MS/MS detection was performed using data-dependent acquisition (DDA) MS techniques on a Thermo Scientific™ Q Exactive™ MS fitted with a Nanospray Flex™ ion source. Data was acquired using an ion spray voltage of 1.9 kV and an interface heater temperature of 275°C. The MS was operated with FULL-MS scans. For DDA, survey scans were acquired in 250 msec, and up to 20 product ion scans (50 msec) were collected. Only spectra with a charge state of 2–4 were selected for fragmentation by higher-energy collision. The MS/MS data were analyzed for protein identification and quantification using the Proteome Discoverer™ (v2.1.0.81). The local false discovery rate was 1.0% after searching against the *Homo sapiens* protein database with a maximum of two missed cleavages and one missed termini cleavage. The following settings were selected: Oxidation (M), Acetylation (Protein N-term), Deamidation (NQ), Pyro-glu from E, and Pyro-glu from Q for variable modifications; Carbamidomethylation (C), iTRAQ 8plex (K), and iTRAQ 8plex (Peptide N-term) for fixed modifications. Precursor and fragment mass tolerance were set to 10 ppm and 0.05 Da, respectively. The MS proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD036680 (10.6019/PXD036680).

Bioinformatics analysis

For ZIP8-WT and ZIP8-KO samples, two of the most closely related biological replicates were selected (tag-114, -115, -117, and -118) for further bioinformatic analysis. Proteins with $p \leq 0.05$ and expression fold-change ≥ 1.2 or ≤ -1.2 were considered differentially expressed, and these proteins were compared by a hierarchical cluster analysis using hclust function (complete linkage method) in R software. All differentially expressed proteins were then subjected to Gene Ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, Eukaryotic Orthologous Groups (KOG) annotation, and subcellular localization analyses. The KOG annotation analysis was performed using automated construction and annotation of orthologous groups of genes (eggNOG) database. Subcellular localization of proteins was analyzed using Bologna Unified Subcellular Component Annotator (BUSCA). Protein-protein interaction (PPI) network analysis was carried out on all or specifically selected differentially expressed proteins using the STRING database and visualized using Cytoscape 3.9.1; the minimum required interaction score was set at 0.4 medium confidence. Clinical-based bioinformatic analysis was performed to assess the co-expression of *ZIP8* and top-10 upregulated or downregulated ZIP8-KO-associated protein genes in 40 cancer types; co-expression data was obtained from TCGA database and analyzed by Spearman's correlation.

Immunoblot analysis

HeLa cells or ZIP8-KO cells were treated with various concentrations of Mn, Se, or Zn for 24 h prior to immunoblot analysis. Specifically, $MnCl_2$, Na_2SeO_3, or $ZnCl_2$ were dissolved directly in MEM medium and sterilized with 0.45 µm filters. Micronutrient-treated samples were then lysed with sample buffer at 95°C for 10 min. After that, cell lysates were centrifuged at 16,900 × *g* for 10 min at room temperature, and the extracted proteins were used for immunoblot analysis as described previously (Qin et al., 2020). Three antibodies were used: C9orf85 (PA5-65639; 1:1,000) and CXorf38 (PA5-62139; 1:1,000) were purchased from Invitrogen and β-actin (A5441; 1:10,000) was purchased from Sigma-Aldrich. Quantifications of immunoblot results were performed using ImageJ (v1.51j8) and GraphPad Prism® 8 (v8.0.2, GraphPad Software Inc.).

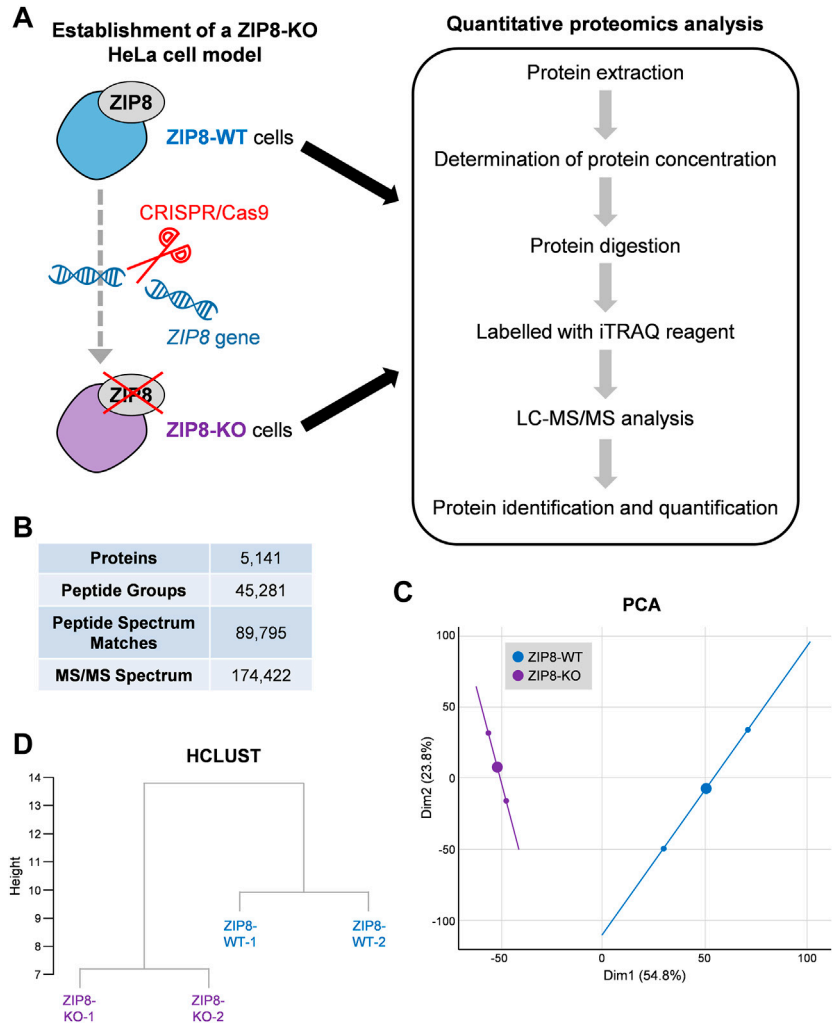


FIGURE 1
Quantitative proteomic analysis of human cells with ZIP8-wildtype (WT) or -knockout (KO). **(A)** Schematic diagram of the cell models subjected to iTRAQ-based comparative proteomic analysis in this study. **(B)** LC-MS/MS data after searching against Homo sapiens protein database for protein identification and quantification. **(C,D)** Principal component analysis (PCA) **(C)** and cluster dendrogram (HCLUST) **(D)** of all identified proteins between ZIP8-WT and ZIP8-KO cells.

Statistical analysis

Unconstrained principal component analysis (PCA) was used to determine the distributions of protein expressions in ZIP8-WT control samples and ZIP8-KO samples. Quantitative data of immunoblotting was expressed as mean ± standard deviation (SD) of at least three replicates. Student's t-test was used for statistical analysis between the untreated control and treated samples, and a probability of $p \leq 0.05$ was used as the criterion for statistical significance. Statistical analysis was performed using GraphPad Prism® 8 software (v8.0.2, GraphPad Software Inc.).

Results

Proteome of human cells with ZIP8 deficiency obtained by iTRAQ-Based comparative proteomic analysis

We previously established a ZIP8-KO HeLa cell model using the CRISPR/Cas9 genome editing technology and showed that ZIP8 is a transmembrane protein responsible for the cellular uptake of Zn, Mn, Se, and Cd (Liang et al., 2021). Here, protein profiles of ZIP8-WT and ZIP8-KO HeLa cell lines were obtained by iTRAQ-coupled LC-MS/MS analysis (Figure 1A; Supplementary Table S1).

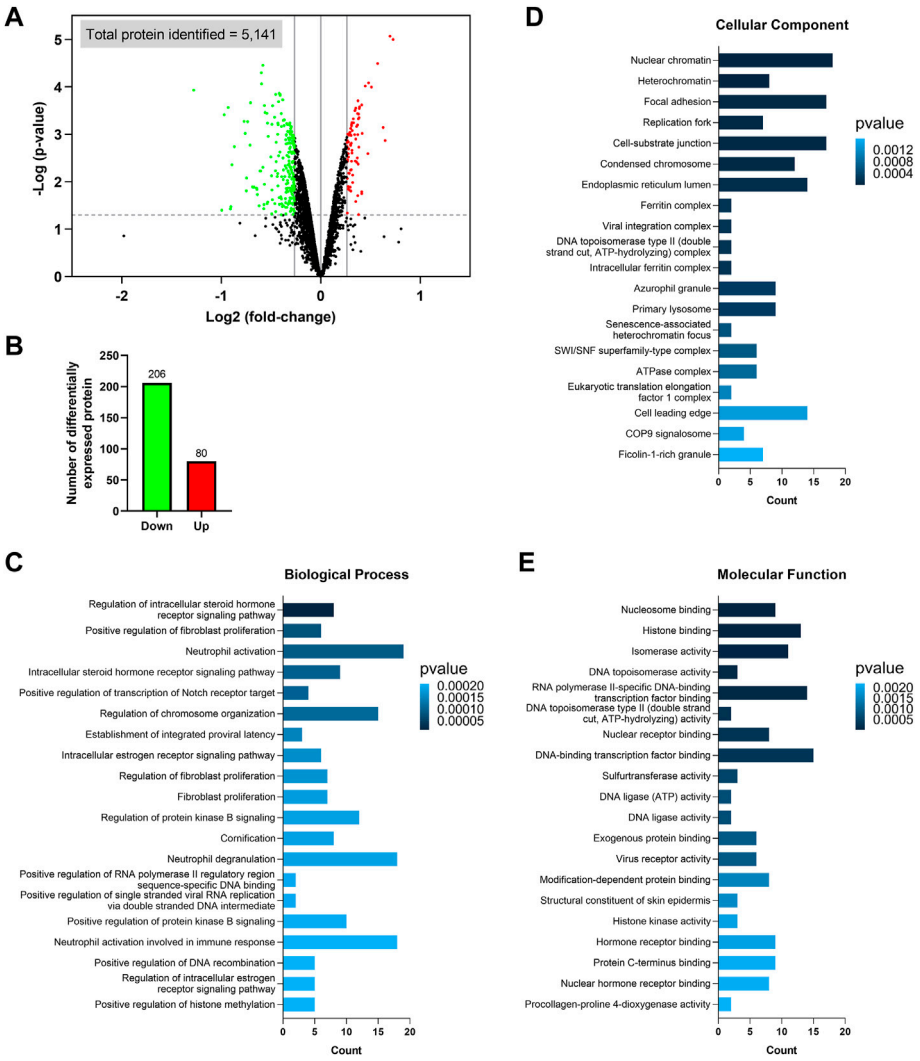


FIGURE 2
Identification of differentially expressed proteins and Gene Ontology (GO) enrichment analysis. **(A)** Volcano plot showing the differentially expressed proteins identified in the ZIP8-KO cells. Only proteins with $p \leq 0.05$ and expression fold-change ≥ 1.2 or ≤ -1.2 are considered significant, and these proteins are highlighted in green (downregulated) or red (upregulated). Vertical lines on the x-axis represent 1.2 or −1.2 fold-change; dash line on y-axis represents $p = 0.05$. **(B)** Number of upregulated and downregulated proteins. **(C–E)** GO enrichment analysis of the differentially expressed proteins classified based on biological process **(C)**, cellular component **(D)**, and molecular function **(E)**.

A total of 174,422 spectra and 45,281 unique peptides corresponding to 5,141 proteins were identified in the tested samples at a false discovery rate of 1% (Figure 1B; Supplementary Figure S1). PCA of the identified proteins showed that the ZIP8-WT and ZIP8-KO cells were generally closer to their biologically replicated counterparts ($n = 2$) but were more distant between different cell lines, indicating an overall dissimilarity of the two cell lines (Figure 1C). Also, hierarchical clustering based on Euclidean distance of the two cell lines showed similar results (Figure 1D).

GO enrichment analysis of differentially expressed proteins in the ZIP8-KO cell model

Among the 5,141 identified proteins, 286 were differentially expressed between the ZIP8-KO and ZIP8-WT cells ($p \leq 0.05$ and expression fold-change ≥ 1.2 or ≤ -1.2) (Figure 2A; Supplementary Table S1). Specifically, the ZIP8-KO cells contained 206 downregulated and 80 upregulated proteins as compared to the ZIP8-WT cells, as shown in Figures 2A,B. Noteworthily, four of the 206 downregulated proteins were

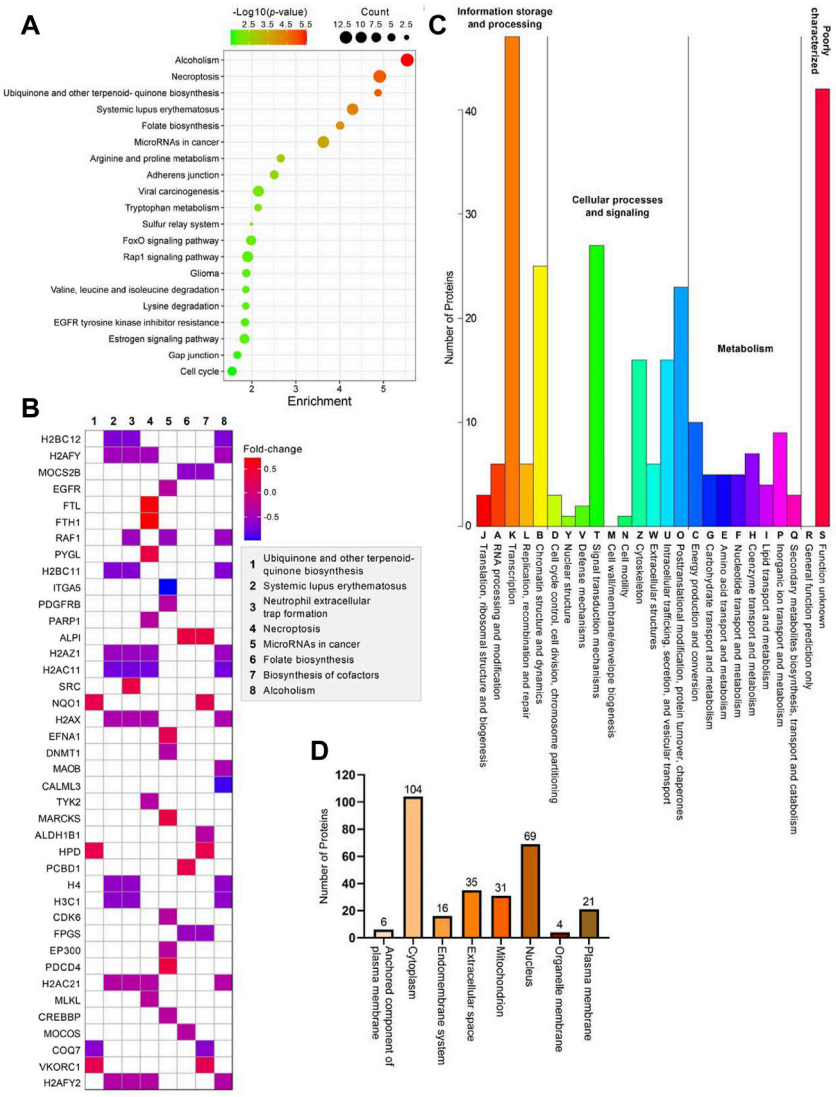


FIGURE 3
Bioinformatics analyses of 286 differentially expressed proteins identified in the ZIP8-KO cell model. **(A)** Bubble plot of KEGG pathway enrichment analysis using KEGGPATHWAY database. **(B)** Heatmap of genes and their relative pathways analyzed using KEGG pathway enrichment analysis. **(C)** KOG annotation analysis using eggNOG database. **(D)** Main subcellular distribution of the differentially expressed proteins analyzed using BUSCA.

classified as the "uncharacterized" proteins: C1orf198, C9orf85, C17orf75, and CXorf38.

The differentially expressed proteins were subsequently classified using the GO categories (Figures 2C–E). Overall, results indicated that the most important biological processes of these proteins were related to fibroblast proliferation, neutrophil activation and degranulation, viral RNA replication, and receptor signaling pathways involving the regulations of steroid hormone, estrogen, Notch, and protein kinase B (Figure 2C; Supplementary Table S2). For GO cellular component enrichment analysis, it was revealed that most proteins were mainly associated with chromatin/chromosome, focal adhesion, replication fork, endoplasmic reticulum lumen, primary lysosome, azurophil granule, and complexes involving viral integration, ferritin, DNA topoisomerase type II, SWI/SNF superfamily-type, ATPase, and eukaryotic translation elongation factor 1 (Figure 2D; Supplementary Table S3). The GO terms of molecular function showed that the differentially expressed proteins mostly participated in the bindings of DNA-binding transcription factor, exogenous protein, histone, hormone receptor, modification-dependent protein, nuclear hormone receptor, nuclear receptor, nucleosome, and protein C-terminus, as well as activities of DNA ligase, DNA

topoisomerase, histone kinase, isomerase, procollagen-proline 4-dioxygenase, sulfurtransferase, and virus receptor (Figure 2E; Supplementary Table S4). Furthermore, since ZIP8 was best known for its role in transporting divalent metal ions, we also looked at which GO categories related to "transporter regulation," "metal response," and "stress response" were significantly enriched, and these categories, along with their associated differential expressed proteins, were summarized in Supplementary Figure S2.

KEGG pathway, eggNOG annotation, and subcellular localization analyses of differentially expressed proteins

Further bioinformatics analyses were carried out on the differentially expressed proteins to better resolve the affected molecular pathways or biological systems in cells without the ZIP8 protein. Briefly, KEGG enrichment analysis indicated that alcoholism, necroptosis, biosynthesis of cofactors (e.g., ubiquinone), systemic lupus erythematosus, folate biosynthesis, microRNAs in cancer were among the most significantly enriched pathways (Figures 3A,B; Supplementary Table S5). In addition, other KEGG pathways that might also involve in ZIP8-KO including those that were related to viral carcinogenesis, EGFR tyrosine kinase inhibitor resistance, degradation of lysine, valine, leucine, and isoleucine, and signaling of FoxO, Rap1, and estrogen (Figure 3A).

KOG annotation analysis revealed that the pathways of transcription, chromatin structure and dynamics, signal transduction mechanisms, posttranslational modification, protein turnover, and chaperones encompassing the highest numbers of differentially expressed proteins, whereas those involved in nuclear structure, cell motility, and cell wall/membrane/envelope biogenesis contained little to no proteins (Figure 3C; Supplementary Table S6). Most of the differentially expressed proteins were proteins mainly located in the cytoplasm, followed by the nucleus (Figure 3D; Supplementary Table S7). Only 6 (ALPI, CD109, CPM, EFNA1, NEGR1, and TCTN3), 4 (CCNB1, CISD1, MPC1, and MRS2), and 21 (ABCG2, ANPEP, ATP11A, ATP13A1, CD83, CD9, CEMIP2, CSPG4, DNAJB11, EGFR, HLA-B, ITGA5, ITGB1, PCSK9, PDGFRB, PTPRF, SCAMP1, SLC7A2, TFRC, TGFBR1, and TSC2) proteins were located in the anchored component of the plasma membrane, organelle membrane, and plasma membrane, respectively (Figure 3D).

PPI network analysis of differentially expressed proteins

We then assessed the PPIs among all the 286 differentially expressed proteins, and the results indicated that a total of 819 interactions were generated from 243 proteins (Figure 4A). The remaining 43 proteins, including three uncharacterized proteins (C1orf198, C17orf75, and CXorf38), did not directly interact with any of the analyzed proteins (not displayed in Figure 4A). However, another uncharacterized protein, C9orf85, was shown to interact with TMEM30A (Figure 4A). Furthermore, we found that a few proteins were able to interact with more than 30 different proteins, and these "super interactors" were CREBBP, EP300, EGFR, H2AFZ, H2AFX, PARP1, SMARCA5, and SRC (Figure 4A). On the other hand, we checked if ZIP8 could directly interact with any of the differentially expressed proteins and found that ZIP8 only interacted with one protein, the transferrin receptor protein 1 (TFRC) (data not shown). In addition, proteins that could interact directly with the four uncharacterized proteins, C1orf198, C9orf85, C17orf75, and CXorf38, were examined and shown in Supplementary Figure S3.

Top-10 upregulated and downregulated proteins associated with ZIP8-KO human cells

We then retrieved the proteins with the most significant changes in their expression based on fold-change values ($p \leq 0.05$). The top-10 most upregulated proteins, from highest to lowest, were FTH1, FTL, CD9, PALM, DAB2, CRIP2, SYNGR1, ABCG2, COL15A1, and GUSB (Table 1). On the other hand, the top-10 most downregulated proteins, from lowest to highest, were KRT9, CXorf38, ITGA5, KRT1, CALML3, NT5C3B, CDKN2A, PPIAL4D, C9orf85, and HMGA1 (Table 2). Notably, two of the four uncharacterized proteins (C9orf85 and CXorf38) were among the top-10 most downregulated proteins.

We also performed a PPI network analysis on the top-10 upregulated and downregulated proteins, and the results indicated that these proteins generally did not interact with each other, with the exception of FTL–FTH1 and ITGA5–CD9 (Figure 4B). When proteins within five interactions were included in the PPI network analysis, five proteins (CDK4, CDK6, FN1, MDM2, and TP53) were identified as the "intermediate interactors" for some of the analyzed proteins (Figure 4B).

Relationships between essential micronutrients levels and uncharacterized proteins C9orf85 and CXorf38

Next, we investigated the potential role of micronutrients on the expressions of C9orf85 and CXorf38. Immunoblot analysis was performed to verify the iTRAQ data, and the results indicated that the protein levels of C9orf85 and CXorf38 were indeed lower in the ZIP8-KO cells compared with the HeLa

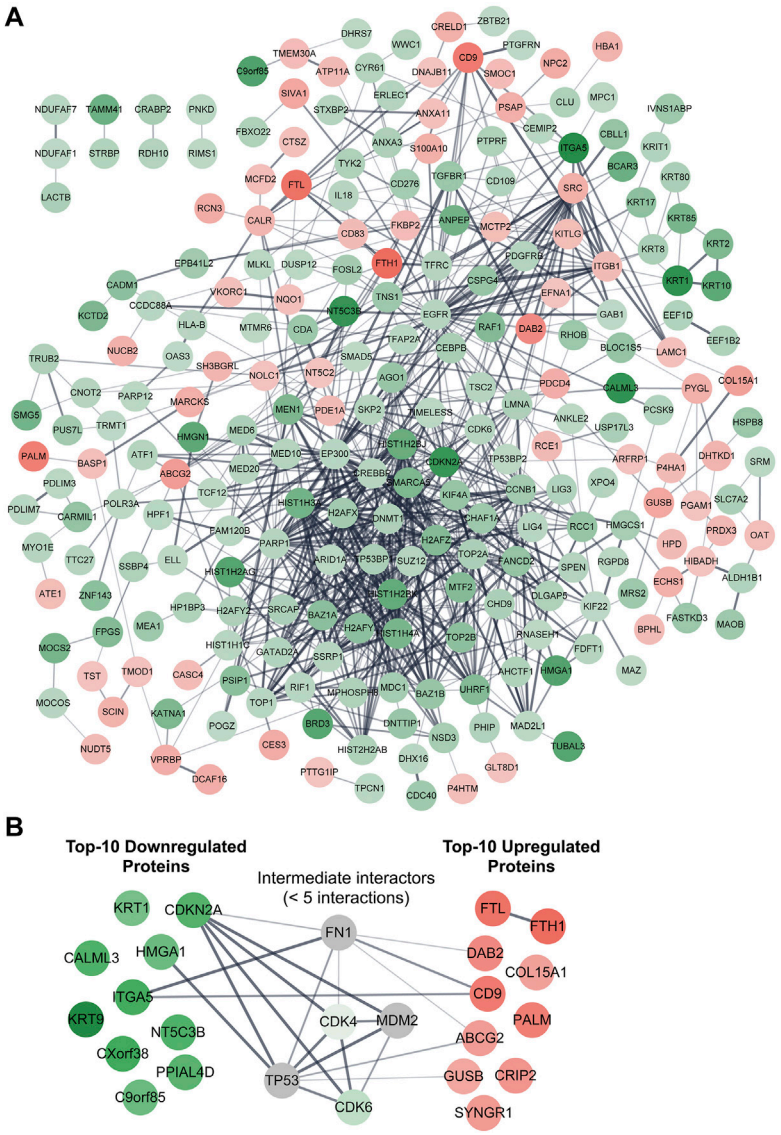


FIGURE 4
PPI network analysis of differentially expressed proteins identified in the ZIP8-KO cell model. Functional and physical interactions of all 286 **(A)** or 20 top-10 upregulated/downregulated **(B)** proteins were subjected to STRING analysis. Disconnected proteins are excluded in the network of **(A)** but included in **(B)**. Proteins within five interactions (max number of interactors to show = <5) are included in **(B)**. The thicker the line between two connected proteins indicates a higher confidence level of the interaction. Downregulated and upregulated proteins are highlighted in green and red, respectively; proteins highlighted in grey represent proteins that are not differentially expressed.

parental cells (Figure 5A; Supplementary Figure S4). Then, ZIP8-KO cells were treated with various concentrations of micronutrients (Mn, Se, and Zn), and the expressions of C9orf85 and CXorf38 were assessed. Overall, we found that the expression of C9orf85 could be induced only briefly by Mn and Se, and it could not be influenced by Zn (Figures 5B–D). However, for CXorf38, its expression could be induced effectively by Mn (Figure 5B), not so effectively by Se (Figure 5C), and not induced by Zn (Figure 5D). These findings suggest that the downregulation of C9orf85 and CXorf38 may be associated with the dysregulation of intracellular micronutrient levels in the ZIP8-KO cells.

Co-expression of ZIP8 and Top-10 upregulated or downregulated ZIP8-KO-associated protein genes in cancers

Lastly, we evaluated the possible connections between the gene expressions of *ZIP8* and the top-10 upregulated or

TABLE 1 Top-10 upregulated proteins identified in the ZIP8-KO cells.

Protein (UniProt ID)	Protein name	Fold-change	*p*-value	Description
FTH1 (P02794)	Ferritin heavy chain	1.6552	0.0000	• Stores iron in a soluble, non-toxic, readily available form • Important for iron homeostasis shows ferroxidase activity
FTL (P02792)	Ferritin light chain	1.6216	0.0000	• Stores iron in a soluble, non-toxic, readily available form • Important for iron homeostasis
CD9 (P21926)	CD9 antigen	1.5677	0.0014	• Involves in cell adhesion, cell motility, and tumor metastasis
PALM (O75781)	Paralemmin-1	1.5455	0.0007	• Involves in plasma membrane dynamics and cell process formation
DAB2 (P98082)	Disabled homolog 2	1.4864	0.0000	• Expression is downregulated in numerous aggressive cancers (e.g., prostate cancer and medulloblastoma) • Involves in several processes such as innate immune response, inflammation and cell growth inhibition, apoptosis, cell survival, angiogenesis, cell migration, and maturation • Induces G0/G1 cell cycle arrest and promotes apoptosis
CRIP2 (P52943)	Cysteine-rich protein 2	1.4234	0.0001	• Zinc ion binding
SYNGR1 (O43759)	Synaptogyrin-1	1.3960	0.0001	• Regulates exocytosis
ABCG2 (Q9UNQ0)	Broad substrate specificity ATP-binding cassette transporter ABCG2	1.3884	0.0026	• Regulates multidrug resistance • Acts as human drug efflux transporter
COL15A1 (P39059)	Collagen alpha-1 (XV) chain	1.3633	0.0001	• Structural protein that stabilizes microvessels and muscle cells, both in the heart and in skeletal muscle • Associates with drug resistance
GUSB (P08236)	Beta-glucuronidase	1.3351	0.0010	• Degradation of dermatan and keratan sulfates

TABLE 2 Top-10 downregulated proteins identified in the ZIP8-KO cells.

Protein (UniProt ID)	Protein name	Fold-change	*p*-value	Description
KRT9 (P35527)	Keratin, type I cytoskeletal 9	−2.4219	0.0001	• Involves in keratin filament assembly • A cytoskeleton intermediate filament protein
CXorf38 (Q8TB03)	Uncharacterized protein CXorf38	−1.9912	0.0397	• Unknown biological function
ITGA5 (P08648)	Integrin alpha-5	−1.9599	0.0004	• Belongs to the integrin alpha chain family • Vital for promoting cancer cell invasion and metastasis
KRT1 (P04264)	Keratin, type II cytoskeletal 1	−1.9039	0.0003	• May regulate the activity of kinases such as PKC and SRC *via* binding to integrin beta-1 (ITB1) and the receptor of activated protein C kinase 1 (RACK1)
CALML3 (P27482)	Calmodulin-like protein 3	−1.8770	0.0375	• May function as a specific light chain of unconventional myosin-10 (MYO10) • Enhances MYO10 translation, possibly by acting as a chaperone for the emerging MYO10 heavy chain protein • May compete with calmodulin by binding, with different affinities, to cellular substrates • Promotes JNK1/2 and ERK1/2 pathway
NT5C3B (Q969T7)	7-methylguanosine phosphate-specific 5′-nucleotidase	−1.8717	0.0336	• Plays a role in inflammation and tissue remodeling • May result in airway wall thickening
CDKN2A (Q8N726)	Tumor suppressor ARF	−1.8563	0.0044	• Acts as a negative regulator of the proliferation of normal cells by interacting strongly with CDK4 and CDK6, inhibiting their ability to interact with cyclins D • A tumor suppressor gene
PPIAL4D (F5H284)	Peptidyl-prolyl cis-trans isomerase A-like 4D	−1.8258	0.0018	• Accelerates the folding of proteins • Catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides
C9orf85 (Q96MD7)	Uncharacterized protein C9orf85	−1.7043	0.0005	• Unknown biological function
HMGA1 (P17096)	High mobility group protein HMG-I/HMG-Y	−1.6915	0.0010	• A coactivator and coregulator of transcriptional activity

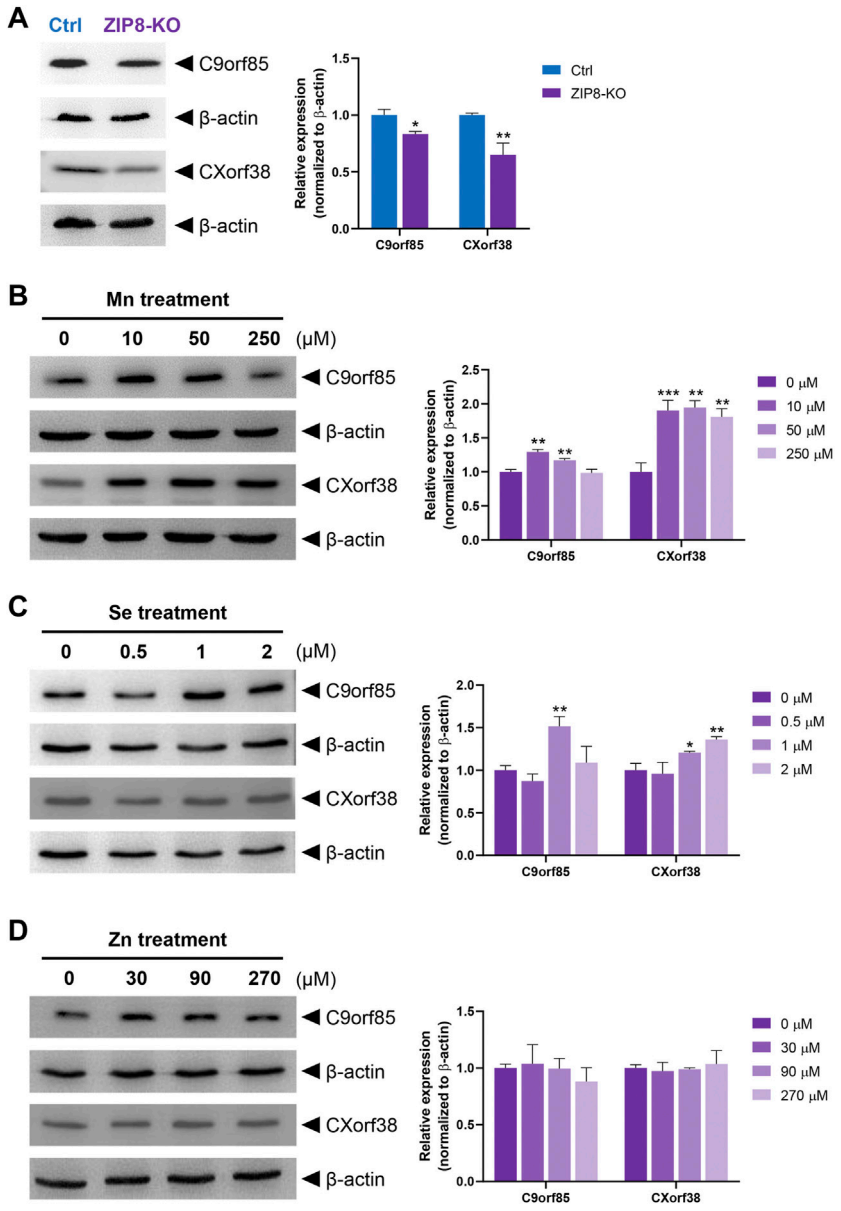


FIGURE 5
Immunoblot analysis of uncharacterized human proteins C9orf85 and CXorf38 treated with essential micronutrients (see Supplementary Figure S4 for full blot images). **(A)** Protein levels of C9orf85 and CXorf38 in HeLa parental cells (with WT ZIP8) and ZIP8-KO HeLa cells. **(B–D)** Protein levels of C9orf85 and CXorf38 in ZIP8-KO HeLa cells exposed to various concentrations of essential micronutrients Mn **(B)**, Se **(C)**, and Zn **(D)** for 24 h. Images are representative of independent experiments with similar expression trends of at least three independent experiments, and error bars represent mean ± SD of triplicate samples. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$.

downregulated ZIP8-KO-associated protein genes in a range of cancers using the TCGA database. We have previously shown that upregulated *ZIP8* (*SLC39A8*) gene expression was common across multiple cancer types (Liang et al., 2021). Here, to determine if the expression of *ZIP8* is significantly correlated to the gene expressions of the top-10 upregulated or downregulated proteins in the ZIP8-KO cells, the following criteria must be met: a negative correlation between the *ZIP8* and an upregulated protein gene; or a positive correlation between the *ZIP8* and a downregulated protein gene. Based on the above criteria, we identified ZIP8-KO-associated protein genes that were significantly co-expressed with *ZIP8* in different cancer types (Figure 6). Specifically, *CXorf38* was significantly co-expressed with *ZIP8* in 18 out of 40 cancer types, whereas *C9orf85* was only in nine out of 40 cancer types (Figure 6).

Cancers	Top 10 up-regulated genes										Top 10 down-regulated genes									
	FTH1	*FTL*	*CD9*	*PALM*	*DAB2*	*CRIP2*	*SYNGR1*	*ABCG2*	*COL15A1*	*GUSB*	*KRT9*	*CXorf38*	*ITGA5*	*KRT1*	*CALML3*	*NT5C3B*	*CDKN2A*	*PPIAL4D*	*C9orf85*	*HMGA1*
Adrenocortical Carcinoma	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Acral Melanoma	ns	-0.389	ns	ns	0.665	-0.689	ns	0.666	0.433	ns	ns	0.427	0.398	ns	ns	-0.524	ns	ns	-0.331	-0.389
Acute Myeloid Leukemia	ns	ns	-0.159	ns	ns	-0.217	ns	ns	ns	ns	N/A	ns	ns	ns	N/A	ns	0.244	ns	ns	ns
Bladder Urothelial Carcinoma	0.121	0.224	-0.269	ns	0.307	ns	-0.320	ns	0.362	0.172	0.251	ns	0.433	0.125	ns	ns	0.230	ns	-0.104	0.228
Brain Lower Grade Glioma	0.195	0.254	0.278	-0.109	ns	ns	-0.128	0.610	0.347	0.207	ns	ns	0.404	0.142	ns	-0.188	ns	ns	-0.224	-0.143
Cervical Squamous Cell Carcinoma	0.174	ns	-0.436	ns	0.220	0.151	-0.135	ns	0.148	0.283	-0.280	ns	ns	-0.307	-0.499	ns	-0.242	ns	-0.161	ns
Cholangiocarcinoma	ns	ns	ns	ns	ns	ns	ns	0.44	ns	ns	ns	ns	-0.344	0.373	ns	ns	ns	ns	ns	ns
Colorectal Adenocarcinoma	ns	ns	0.131	-0.367	0.101	-0.211	-0.146	ns	ns	ns	ns	0.081	-0.195	ns	ns	ns	-0.158	ns	0.286	-0.113
Diffuse Large B-Cell Lymphoma	0.513	0.479	ns	ns	0.402	ns	ns	ns	0.337	ns	ns	0.387	0.606	ns	ns	ns	ns	ns	ns	ns
Esophagogastric Adenocarcinoma	ns	0.193	-0.448	ns	0.442	ns	-0.321	0.163	0.431	0.404	-0.356	0.291	ns	-0.353	-0.446	-0.168	0.258	ns	ns	ns
Glioblastoma Multiforme	0.421	0.381	ns	-0.294	0.524	ns	-0.326	ns	0.418	0.315	ns	0.265	0.490	0.166	ns	-0.350	ns	ns	ns	-0.263
Head and Neck Squamous Cell Carcinoma	0.128	0.168	-0.277	ns	0.502	ns	ns	0.185	0.408	0.187	-0.310	0.329	0.236	ns	-0.220	-0.112	ns	ns	-0.145	-0.234
Invasive Breast Carcinoma	ns	0.112	-0.078	-0.256	0.214	-0.205	-0.150	ns	0.139	-0.087	0.097	0.164	0.094	ns	ns	ns	0.078	ns	0.138	0.063
Kidney Renal Papillaryy Cell Carcinoma	-0.145	-0.125	ns	0.187	0.146	0.186	0.358	-0.165	-0.372	0.191	-0.161	ns	-0.407	0.123	0.172	0.130	ns	ns	ns	-0.217
Liver Hepatocellular Carcinoma	-0.181	ns	ns	-0.336	ns	-0.166	-0.281	0.510	ns	0.168	-0.121	ns	-0.306	ns	0.114	-0.144	ns	ns	ns	-0.366
Lung Adenocarcinoma	ns	0.199	ns	ns	0.360	ns	ns	0.244	-0.184	ns	-0.136	0.166	-0.124	0.237	-0.242	-0.123	ns	ns	0.106	-0.345
Lung Squamous Cell Carcinoma	0.097	0.296	-0.250	0.099	0.485	0.312	-0.305	0.213	0.217	0.236	-0.219	0.203	0.267	ns	-0.230	-0.243	ns	ns	-0.133	-0.300
Melanoma	ns	0.581	ns	ns	-0.448	ns	ns	0.556	0.447	0.448	ns	ns	ns	ns	ns	ns	ns	N/A	-0.653	ns
Mesothelioma	ns	ns	0.259	ns	0.379	-0.346	ns	ns	-0.371	ns	ns	0.393	-0.348	ns	ns	-0.474	0.271	ns	ns	-0.303
Metastatic Prostate Adenocarcinoma	ns	0.337	0.198	ns	0.149	ns	ns	ns	ns	ns	ns	ns	0.230	ns	-0.221	0.247	ns	ns	ns	ns
Miscellaneous Neuroepithelial Tumor	0.177	0.145	ns	ns	0.391	ns	ns	0.269	0.392	ns	0.267	ns	0.356	0.242	ns	ns	0.177	ns	ns	ns
Ovarian Serous Cystadenocarcinoma	0.247	0.191	0.210	ns	0.157	ns	ns	-0.196	ns	ns	ns	ns	0.284	ns	ns	ns	ns	ns	ns	ns
Pancreatic Adenocarcinoma	ns	ns	ns	ns	0.413	ns	ns	0.152	0.388	ns	ns	ns	0.215	0.155	ns	ns	ns	ns	0.175	ns
Pediatric Acute Lymphoid Leukemia (Phase II)	ns	ns	ns	ns	ns	ns	ns	-0.217	ns	ns	ns	ns	0.145	ns	ns	-0.193	ns	ns	ns	-0.236
Pediatric Neuroblastoma	0.333	0.471	ns	-0.202	0.583	ns	-0.366	0.368	0.516	0.166	-0.191	0.168	0.658	0.225	ns	-0.283	ns	ns	ns	ns
Pediatric Rhabdoid Tumor	ns	0.459	ns	ns	0.312	ns	ns	ns	0.325	0.376	ns	ns	ns	ns	ns	ns	ns	N/A	ns	ns
Pediatric Wilms' Tumor	ns	ns	ns	-0.294	ns	ns	ns	0.285	ns	ns	-0.263	ns	ns	ns	ns	ns	ns	ns	ns	ns
Pheochromocytoma and Paraganglioma (Miscellaneous Neuroepithelial Tumor)	0.177	0.145	ns	ns	0.391	ns	ns	0.269	0.392	ns	0.267	ns	0.356	0.242	ns	ns	0.177	ns	ns	ns
Pheochromocytoma and Paraganglioma (Soft Tissue)	0.197	0.149	ns	ns	0.394	ns	ns	0.260	0.389	ns	0.270	ns	0.353	0.242	ns	ns	0.203	ns	ns	ns
Prostate Adenocarcinoma	ns	ns	0.155	-0.152	0.248	-0.116	ns	ns	ns	0.206	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Renal Call Carcinoma	ns	-0.120	-0.110	-0.114	0.238	-0.099	ns	0.325	0.159	ns	-0.109	0.217	-0.121	ns	-0.117	ns	-0.199	-0.155	0.093	-0.192
Skin Cutaneous Melanoma	-0.116	ns	ns	-0.168	ns	-0.184	-0.245	0.200	0.178	0.118	ns	0.314	0.106	ns	ns	ns	ns	ns	ns	-0.286
Small Cell Lung Cancer	ns	0.352	0.254	ns	0.635	ns	0.287	0.354	0.498	ns	ns	ns	0.624	0.252	-0.222	-0.264	ns	N/A	ns	ns
Stomach Adenocarcinoma	0.188	ns	ns	-0.179	0.099	ns	-0.162	ns	ns	0.207	ns	ns	ns	-0.112	ns	ns	ns	ns	0.143	ns
Testicular Germ Cell Tumors	0.449	0.308	ns	ns	0.523	ns	ns	0.394	0.317	0.477	-0.313	0.290	0.547	ns	0.208	-0.314	ns	0.195	0.421	0.212
Thymoma	-0.415	-0.297	ns	-0.226	0.225	-0.290	ns	0.288	ns	ns	ns	0.257	ns	ns	-0.260	-0.318	ns	ns	ns	-0.324
Upper Tract Urothelial Carcinoma	0.597	0.588	0.652	ns	0.797	0.207	ns	0.516	0.837	0.524	ns	0.504	0.418	ns	ns	ns	0.377	N/A	0.516	0.353
Uterine Carcinosarcoma	0.284	ns	0.504	ns	ns	0.358	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.298	ns	ns	ns	ns
Uterine Corpus Endometrial Carcinoma	ns	-0.143	0.153	-0.181	0.175	ns	ns	ns	0.146	ns	ns	0.120	ns	ns	ns	-0.150	-0.098	ns	0.131	-0.093
Uveal Melanoma	ns	ns	0.462	-0.393	-0.440	0.236	-0.356	0.281	ns	0.305	-0.284	0.399	0.360	ns	ns	-0.317	ns	ns	ns	-0.519

FIGURE 6
Co-expression of *ZIP8* (*SLC39A8*) and top-10 upregulated or downregulated ZIP8-KO-associated protein genes in 40 cancer types. Clinical data were obtained from TCGA database and analyzed by Spearman's correlation. A Spearman's correlation value greater than 0 indicates a positive correlation, and a value less than 0 indicates a negative correlation. Co-expressions that signify the proteome of ZIP8-KO were highlighted: negative correlations between the *ZIP8* and upregulated protein genes (highlighted in red) or positive correlations between the *ZIP8* and downregulated protein genes (highlighted in green). Only values with $p \leq 0.05$ were shown. N/A: data not available; ns: non-significant.

Discussion

Recent advances in omics technologies have enabled us to rapidly investigate the status of the biomolecules of interest in a selected biological system (Tan et al., 2018). Single-gene KO cell lines are useful models to study the function of a particular gene, and when an omics analysis (SGKOmics analysis) is performed on such cell models, the overall biophysiological effects of the gene on the entire cell can be revealed.

The ZIP8 is a transmembrane protein that involves in the uptake of Fe, Mn, Zn, and Se, and therefore, cells with impaired ZIP8 are often associated with dysregulated intracellular levels of these essential micronutrients (McDermott et al., 2016; Zang et al., 2016; Lin et al., 2017; Liang et al., 2021). Although studies have indicated that the expression of ZIP8 can be affected by different physiological conditions (e.g., during lactation, Cd exposure, ZnO nanoparticle treatment, and inflammation) (Kelleher et al., 2012; Gao et al., 2017; Xu et al., 2017; Pan et al., 2020; Hatano et al., 2021), it is largely unclear how impaired ZIP8 will affect the cells other than having an impact on the intracellular homeostasis of micronutrients.

Here, using an iTRAQ-based quantitative proteomics approach, the proteome of a ZIP8-KO cell model was examined for the first time. This ZIP8-KO stable cell line was previously established from the HeLa cells using the CRISPR/Cas9 genome editing technology (Liang et al., 2021). An endogenous KO of the *ZIP8* gene could allow us to study the functions of this gene more explicitly by

mimicking the conditions of patients with non-functional ZIP8—the physiological roles of ZIP8 in relation to a range of human diseases have been extensively reviewed (Zang et al., 2016; Fujishiro and Himeno, 2019; Nebert and Liu, 2019). However, only a ZIP8-KO cell model cannot fully demonstrate or explain the connections between ZIP8 and human diseases. Thus, further research should be carried out on additional ZIP8-KO cell and animal models, as well as patient samples with impaired ZIP8.

Bioinformatics analysis of the proteomic changes in the ZIP8-KO cells revealed proteins and their associated pathways that were potentially affected by the dysregulation of intracellular micronutrient levels, including a range of signaling pathways that are long-known for their roles in human diseases. For example, the FoxO-related signaling pathway is involved in sarcopenia (a skeletal muscle disorder) and cancer development, especially upon nutrient starvation (Jiramongkol and Lam, 2020; Liu et al., 2021); Signaling pathway of estrogen is highly regulated by the nutrient availability, and the abnormal actions of estrogen can lead to various metabolic syndromes as well as affect the immune and inflammatory conditions of the cells (Mauvais-Jarvis et al., 2013); Notch-related signaling pathway is involved in the regulation of neurogenesis and central nervous system, which is linked to neurological disorders (e.g., Parkinson's disease and Alzheimer's disease) and neuroendocrine tumors (e.g., glioma) (Kim et al., 2020; Zhou et al., 2022).

In the ZIP8-KO cells, it was shown that many of the differentially expressed proteins were viral infection-related, as pathways such as viral RNA replication, integration, carcinogenesis, and receptor activity were significantly enriched. Evidence is clear that an imbalanced level of intracellular micronutrients can disrupt normal immune function and increase the risk of microbial infection (Gombart et al., 2020). It has also been suggested that ZIP8 is one of the key proteins in the pathogenesis of viral infections, especially during respiratory viral infections, since ZIP8 is highly expressed in the human lungs (Sadeghsoltani et al., 2022). Among the ZIP8-transportable essential micronutrients, Zn and Se are the two main elements that exhibit antiviral properties and play vital roles in the immune response against viral infections (Kar et al., 2019; Tan et al., 2020). In addition to preventing viral infection, Se and many Se compounds are also potent anticancer agents that can be utilized in cancer prevention and therapy (Tan et al., 2019). Furthermore, previous studies have suggested that ZIP8 is associated with the development of cancers, probably due to disrupted Se homeostasis (Liu et al., 2018; Liang et al., 2021). However, the intertwining relationships between the differentially expressed proteins identified in the ZIP8-KO cells, intracellular micronutrient levels, and human diseases have remained unclear and warrant further studies.

Humans have approximately 22,000 protein-coding genes, which could be translated into hundreds of thousands of proteins through alternative splicing (Wilhelm et al., 2014). To date, the characteristics of many proteins have remained uninvestigated, and they are collectively known as the "uncharacterized" protein and commonly named as the "open reading frame (ORF)" proteins (Duek and Lane, 2019). Early studies of uncharacterized proteins usually focused only on whether the atypical ORF regions found within the genome could be translated into proteins. Further research on uncharacterized proteins, especially those encoded by short-ORFs, showed that they were indeed involved in diverse biological processes, including but not limited to cell proliferation (Galindo et al., 2007), DNA damage repair (Slavoff et al., 2014), muscle activities (Anderson et al., 2015; Nelson et al., 2016), cell signaling (Chng et al., 2013), mRNA degradation (D'Lima et al., 2017), and phagocytosis (Pueyo et al., 2016). Therefore, uncharacterized proteins have attracted increasing attention in recent years, but systemic research on these proteins is still lacking (Delcourt et al., 2018; Wu et al., 2022).

In the current study, one of the most interesting findings in the altered proteome of ZIP8-KO cells was the discovery of two uncharacterized proteins among the top-10 significantly downregulated proteins: C9orf85 and CXorf38. Although we did not test the potential biological functions of these two uncharacterized proteins, we did find that their expressions could be induced by either Mn or Se, thus unveiling the possible roles of these proteins in nutrition. These findings also indicate that C9orf85 and CXorf38 have the potential to be targeted as biomarkers and/or therapeutic targets for micronutrient deficiency diseases. Nevertheless, further studies are required to investigate the functions of these two uncharacterized proteins in human diseases.

Conclusion

To conclude, this study shows that a functional ZIP8 is important in maintaining normal cellular physiology. In particular, the differentially expressed proteins (e.g., C9orf85 and CXorf38) identified in the ZIP8-KO cells could be potential targets for diagnosing and/or treating a wide range of human ZIP8-associated diseases, including but not limited to malnutrition, viral infection, and cancers.

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: http://www.proteomexchange.org/, PXD036680.

Author contributions

Conceptualization: Y-MX and ATYL; Data curation: HWT, Z-LL, N-LC, and Y-YW; Formal analysis: HWT, Z-LL, Y-MX, and ATYL; Funding acquisition: Y-MX and ATYL; Investigation: HWT, Z-LL, Y-MX, and ATYL; Methodology: HWT, Z-LL, Y-MX, and ATYL; Project administration: Y-MX and ATYL; Resources: Y-MX and ATYL; Supervision: Y-MX and ATYL; Validation: HWT, Z-LL, N-LC, and Y-YW; Visualization: HWT and Z-LL; Writing—original draft: HWT, Y-MX, and ATYL; Writing—review & editing: HWT, Y-MX, and ATYL. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (31771582 and 31271445), the Guangdong Natural Science Foundation of China (2017A030313131), the "Thousand, Hundred, and Ten" Project of the Department of Education of Guangdong Province of China, the Basic and Applied Research Major Projects of Guangdong Province of China (2017KZDXM035 and 2018KZDXM036), the "Yang Fan" Project of Guangdong Province of China (Andy T. Y. Lau-2016 and Yan-Ming Xu-2015), and the Shantou Medical Health Science and Technology Plan (200624165260857).

Acknowledgments

We would like to thank members of the Lau And Xu laboratory for critical reading of this manuscript.

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/fmolb.2022.991308/full#supplementary-material

SUPPLEMENTARY TABLE S1
Data of iTRAQ-coupled LC-MS/MS analysis.

SUPPLEMENTARY TABLE S2
Data of gene ontology (biological process) enrichment analysis of differentially expressed proteins.

SUPPLEMENTARY TABLE S3
Data of gene ontology (cellular component) enrichment analysis of differentially expressed proteins.

SUPPLEMENTARY TABLE S4
Data of gene ontology (molecular function) enrichment analysis of differentially expressed proteins.

SUPPLEMENTARY TABLE S5
Data of Kyoto encyclopedia of genes and genomes pathway enrichment analysis of differentially expressed proteins.

SUPPLEMENTARY TABLE S6
Data of eukaryotic orthologous groups annotation analysis of differentially expressed proteins.

SUPPLEMENTARY TABLE S7
Data of subcellular location of differentially expressed proteins.

SUPPLEMENTARY FIGURE S1
Heatmap of identified proteins in the ZIP8-WT or ZIP8-KO HeLa cells based on hierarchical cluster analysis of individual samples **(A)** or combined samples **(B)**.

SUPPLEMENTARY FIGURE S2
Significantly enriched Gene Ontology (GO) categories **(A)** and differentially expressed proteins **(B)** related to "transporter regulation", "metal response", and "stress response." BP: Biological process; MF: Molecular function.

SUPPLEMENTARY FIGURE S3
PPI network analysis of uncharacterized proteins C1orf198, C9orf85, C17orf75, and CXorf38.

SUPPLEMENTARY FIGURE S4
Immunoblotting of HeLa parental cells (with ZIP8-WT) or ZIP8-KO cells probed with C9orf85 (PA5-65639; Invitrogen), CXorf38 (PA5-62139; Invitrogen), or β-actin (A5441; Sigma-Aldrich) antibody, displayed in full blots.

References

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi:10.1016/j.cell.2015.01.009

Boycott, K. M., Beaulieu, C. L., Kernohan, K. D., Gebril, O. H., Mhanni, A., Chudley, A. E., et al. (2015). Autosomal-recessive intellectual disability with cerebellar atrophy syndrome caused by mutation of the manganese and zinc transporter gene *SLC39A8*. *Am. J. Hum. Genet.* 97, 886–893. doi:10.1016/j.ajhg.2015.11.002

Chng, S. C., Ho, L., Tian, J., and Reversade, B. (2013). ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev. Cell* 27, 672–680. doi:10.1016/j.devcel.2013.11.002

Choi, E. K., Nguyen, T. T., Gupta, N., Iwase, S., and Seo, Y. A. (2018). Functional analysis of *SLC39A8* mutations and their implications for manganese deficiency and mitochondrial disorders. *Sci. Rep.* 8, 3163. doi:10.1038/s41598-018-21464-0

Delcourt, V., Staskevicius, A., Salzet, M., Fournier, I., and Roucou, X. (2018). Small proteins encoded by unannotated ORFs are rising stars of the proteome, confirming shortcomings in genome annotations and current vision of an mRNA. *Proteomics* 18, e1700058. doi:10.1002/pmic.201700058

D'Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., et al. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* 13, 174–180. doi:10.1038/nchembio.2249

Duek, P., and Lane, L. (2019). Worming into the uncharacterized human proteome. *J. Proteome Res.* 18, 4143–4153. doi:10.1021/acs.jproteome.9b00435

Fujishiro, H., and Himeno, S. (2019). New insights into the roles of ZIP8, a cadmium and manganese transporter, and its relation to human diseases. *Biol. Pharm. Bull.* 42, 1076–1082. doi:10.1248/bpb.b18-00637

Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A., and Couso, J. P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5, e106. doi:10.1371/journal.pbio.0050106

Gao, Y., Xu, Y., Wu, D., Yu, F., Yang, L., Yao, Y., et al. (2017). Progressive silencing of the zinc transporter Zip8 (Slc39a8) in chronic cadmium-exposed lung epithelial cells. *Acta Biochim. Biophys. Sin.* 49, 444–449. doi:10.1093/abbs/gmx022

Gombart, A. F., Pierre, A., and Maggini, S. (2020). A review of micronutrients and the immune system–working in harmony to reduce the risk of infection. *Nutrients* 12, 236. doi:10.3390/nu12010236

Haller, G., McCall, K., Jenkitkasemwong, S., Sadler, B., Antunes, L., Nikolov, M., et al. (2018). A missense variant in *SLC39A8* is associated with severe idiopathic scoliosis. *Nat. Commun.* 9, 4171. doi:10.1038/s41467-018-06705-0

Hatano, N., Matsubara, M., Suzuki, H., Muraki, Y., and Muraki, K. (2021). HIF-1α dependent upregulation of ZIP8, ZIP14, and TRPA1 modify intracellular Zn^{2+} accumulation in inflammatory synoviocytes. *Int. J. Mol. Sci.* 22, 6349. doi:10.3390/ijms22126349

Jiramongkol, Y., and Lam, E. W. (2020). FOXO transcription factor family in cancer and metastasis. *Cancer Metastasis Rev.* 39, 681–709. doi:10.1007/s10555-020-09883-w

Kar, M., Khan, N. A., Panwar, A., Bais, S. S., Basak, S., Goel, R., et al. (2019). Zinc chelation specifically inhibits early stages of dengue virus replication by activation of NF-κB and induction of antiviral response in epithelial cells. *Front. Immunol.* 10, 2347. doi:10.3389/fimmu.2019.02347

Kelleher, S. L., Velasquez, V., Croxford, T. P., McCormick, N. H., Lopez, V., and MacDavid, J. (2012). Mapping the zinc-transporting system in mammary cells: Molecular analysis reveals a phenotype-dependent zinc-transporting network during lactation. *J. Cell. Physiol.* 227, 1761–1770. doi:10.1002/jcp.22900

Kim, S., Lee, M., and Choi, Y. K. (2020). The role of a neurovascular signaling pathway involving hypoxia-inducible factor and Notch in the function of the central nervous system. *Biomol. Ther.* 28, 45–57. doi:10.4062/biomolther.2019.119

Li, D., Achkar, J. P., Haritunians, T., Jacobs, J. P., Hui, K. Y., D'Amato, M., et al. (2016). A pleiotropic missense variant in *SLC39A8* is associated with crohn's disease and human gut microbiome composition. *Gastroenterology* 151, 724–732. doi:10.1053/j.gastro.2016.06.051

Liang, Z. L., Tan, H. W., Wu, J. Y., Chen, X. L., Wang, X. Y., Xu, Y. M., et al. (2021). The impact of ZIP8 disease-associated variants G38R, C113S, G204C, and S335T on selenium and cadmium accumulations: The first characterization. *Int. J. Mol. Sci.* 22, 11399. doi:10.3390/ijms222111399

Lin, W., Vann, D. R., Doulias, P. T., Wang, T., Landesberg, G., Li, X., et al. (2017). Hepatic metal ion transporter ZIP8 regulates manganese homeostasis and manganese-dependent enzyme activity. *J. Clin. Invest.* 127, 2407–2417. doi:10.1172/jci90896

Liu, L., Geng, X., Cai, Y., Copple, B., Yoshinaga, M., Shen, J., et al. (2018). Hepatic ZIP8 deficiency is associated with disrupted selenium homeostasis, liver pathology, and tumor formation. *Am. J. Physiol. Gastrointest. Liver Physiol.* 315, G569–G579. doi:10.1152/ajpgi.00165.2018

Liu, P. J., Hu, Y. S., Wang, M. J., and Kang, L. (2021). Nutrient weight Against sarcopenia: Regulation of the IGF-1/PI3K/akt/FOXO pathway in quinoa metabolites. *Curr. Opin. Pharmacol.* 61, 136–141. doi:10.1016/j.coph.2021.10.001

Mauvais-Jarvis, F., Clegg, D. J., and Hevener, A. L. (2013). The role of estrogens in control of energy balance and glucose homeostasis. *Endocr. Rev.* 34, 309–338. doi:10.1210/er.2012-1055

McCoy, T. H., Jr., Pellegrini, A. M., and Perlis, R. H. (2019). Using phenome-wide association to investigate the function of a schizophrenia risk locus at SLC39A8. *Transl. Psychiatry* 9, 45. doi:10.1038/s41398-019-0386-9

McDermott, J. R., Geng, X., Jiang, L., Gálvez-Peralta, M., Chen, F., Nebert, D. W., et al. (2016). Zinc- and bicarbonate-dependent ZIP8 transporter mediates selenite uptake. *Oncotarget* 7, 35327–35340. doi:10.18632/oncotarget.9205

Mezzaroba, L., Alfieri, D. F., Colado Simão, A. N., and Vissoci Reiche, E. M. (2019). The role of zinc, copper, manganese and iron in neurodegenerative diseases. *Neurotoxicology* 74, 230–241. doi:10.1016/j.neuro.2019.07.007

Nebert, D. W., and Liu, Z. (2019). *SLC39A8* gene encoding a metal ion transporter: Discovery and bench to bedside. *Hum. Genomics* 13, 51. doi:10.1186/s40246-019-0233-3

Nelson, B. R., Makarewich, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 351, 271–275. doi:10.1126/science.aad4076

Pan, C. Y., Lin, F. Y., Kao, L. S., Huang, C. C., and Liu, P. S. (2020). Zinc oxide nanoparticles modulate the gene expression of ZnT_1 and ZIP_8 to manipulate zinc homeostasis and stress-induced cytotoxicity in human neuroblastoma SH-SY5Y cells. *PLoS One* 15, e0232729. doi:10.1371/journal.pone.0232729

Park, J. H., Hogrebe, M., Grüneberg, M., DuChesne, I., von der Heiden, A. L., Reunert, J., et al. (2015). SLC39A8 deficiency: A disorder of manganese transport and glycosylation. *Am. J. Hum. Genet.* 97, 894–903. doi:10.1016/j.ajhg.2015.11.003

Park, J. H., Hogrebe, M., Fobker, M., Brackmann, R., Fiedler, B., Reunert, J., et al. (2018). SLC39A8 deficiency: Biochemical correction and major clinical improvement by manganese therapy. *Genet. Med.* 20, 259–268. doi:10.1038/gim.2017.106

Perez-Riverol, Y., Bai, J., Bandla, C., Hewapathirana, S., García-Seisdedos, D., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. doi:10.1093/nar/gkab1038

Pueyo, J. I., Magny, E. G., Sampson, C. J., Amin, U., Evans, I. R., Bishop, S. A., et al. (2016). Hemotin, a regulator of phagocytosis encoded by a small ORF and conserved across metazoans. *PLoS Biol.* 14, e1002395. doi:10.1371/journal.pbio.1002395

Qin, S. H., Lau, A. T. Y., Liang, Z. L., Tan, H. W., Ji, Y. C., Zhong, Q. H., et al. (2020). Resveratrol promotes tumor microvessel growth via endoglin and extracellular signal-regulated kinase signaling pathway and enhances the anticancer efficacy of gemcitabine against lung cancer. *Cancers* 12, 974. doi:10.3390/cancers12040974

Riley, L. G., Cowley, M. J., Gayevskiy, V., Roscioli, T., Thorburn, D. R., Prelog, K., et al. (2017). A *SLC39A8* variant causes manganese deficiency, and glycosylation and mitochondrial disorders. *J. Inherit. Metab. Dis.* 40, 261–269. doi:10.1007/s10545-016-0010-6

Sadeghsoltani, F., Mohammadzadeh, I., Safari, M. M., Hassanpour, P., Izadpanah, M., Qujeq, D., et al. (2022). Zinc and respiratory viral infections: Important trace element in anti-viral response and immune regulation. *Biol. Trace Elem. Res.* 200, 2556–2571. doi:10.1007/s12011-021-02859-z

Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., and Saghatelian, A. (2014). A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *J. Biol. Chem.* 289, 10950–10957. doi:10.1074/jbc.C113.533968

Tan, H. W., Xu, Y. M., Wu, D. D., and Lau, A. T. Y. (2018). Recent insights into human bronchial proteomics – how are we progressing and what is next? *Expert Rev. Proteomics* 15, 113–130. doi:10.1080/14789450.2017.1417847

Tan, H. W., Mo, H. Y., Lau, A. T. Y., and Xu, Y. M. (2019). Selenium species: Current status and potentials in cancer prevention and therapy. *Int. J. Mol. Sci.* 20, 75. doi:10.3390/ijms20010075

Tan, H. W., Xu, Y. M., and Lau, A. T. Y. (2020). Angiotensin-converting enzyme 2: The old door for new severe acute respiratory syndrome coronavirus 2 infection. *Rev. Med. Virol.* 30, e2122. doi:10.1002/rmv.2122

Wang, C. Y., Jenkitkasemwong, S., Duarte, S., Sparkman, B. K., Shawki, A., Mackenzie, B., et al. (2012). ZIP8 is an iron and zinc transporter whose cell-surface expression is up-regulated by cellular iron loading. *J. Biol. Chem.* 287, 34032–34043. doi:10.1074/jbc.M112.367284

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., et al. (2014). Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587. doi:10.1038/nature13319

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi:10.1038/nmeth.1322

Wu, C., Lu, X., Lu, S., Wang, H., Li, D., Zhao, J., et al. (2022). Efficient detection of the alternative spliced human proteome using translatome sequencing. *Front. Mol. Biosci.* 9, 895746. doi:10.3389/fmolb.2022.895746

Xu, Y. M., Gao, Y. M., Wu, D. D., Yu, F. Y., Zang, Z. S., Yang, L., et al. (2017). Aberrant cytokine secretion and zinc uptake in chronic cadmium-exposed lung epithelial cells. *Prot. Clin. Appl.* 11, 1600059. doi:10.1002/prca.201600059

Zang, Z. S., Xu, Y. M., and Lau, A. T. Y. (2016). Molecular and pathophysiological aspects of metal ion uptake by the zinc transporter ZIP8 (SLC39A8). *Toxicol. Res.* 5, 987–1002. doi:10.1039/c5tx00424a

Zhang, R., Witkowska, K., Afonso Guerra-Assunção, J., Ren, M., Ng, F. L., Mauro, C., et al. (2016). A blood pressure-associated variant of the *SLC39A8* gene influences cellular cadmium accumulation and toxicity. *Hum. Mol. Genet.* 25, 4117–4126. doi:10.1093/hmg/ddw236

Zhou, B., Lin, W., Long, Y., Yang, Y., Zhang, H., Wu, K., et al. (2022). Notch signaling pathway: Architecture, disease, and therapeutics. *Signal Transduct. Target. Ther.* 7, 95. doi:10.1038/s41392-022-00934-y

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fgene.2022.1014836

OPEN ACCESS

EDITED BY
Yan-Ming Xu,
Shantou University, China

REVIEWED BY
Rongcai Yue,
Fujian Medical University, China
Hui Li,
Guangxi University, China
Alex Boye,
University of Cape Coast, Ghana

*CORRESPONDENCE
Jun Ma,
majun@shtrhospital.com
Qin Lu,
LQ1982@shtrhospital.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 09 August 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Chen T, Zhang Z, Lu Q and Ma J (2022), Screening and functional analysis of the differential peptides from the placenta of patients with healthy pregnancy and preeclampsia using placental peptidome.
Front. Genet. 13:1014836.
doi: 10.3389/fgene.2022.1014836

Screening and functional analysis of the differential peptides from the placenta of patients with healthy pregnancy and preeclampsia using placental peptidome

Tingting Chen[1†], Zhongxiao Zhang[2†], Qin Lu[1*] and Jun Ma[1,3*]

[1]Department of Gynaecology and Obstetrics, Tongren Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, [2]Hongqiao International Institute of Medicine, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3]Department of General Practitioners, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Molecular peptides play an extensive range of functions in the human body. However, no previous study has performed placental peptidome profiling. In the present study, 3,941 peptides from human placental tissues were identified using peptidomics. Compared to healthy pregnant women, there were 87 and 129 differentially expressed peptides (DEPs) in the mild and severe preeclampsia groups, respectively. In the mild PE group, 55 and 34 DEPs had high and low expressions, respectively. In comparison, in the severe PE group, 82 and 47 DEPs had high and low expressions, respectively. Functional analysis of the precursor proteins of DEPs by gene ontology suggested that they are primarily involved in focal adhesion, extracellular matrix-receptor interaction, tight junction, and extracellular matrix. Network analysis using ingenuity pathway analysis software showed that the precursor proteins of DEPs were primarily related to the transforming growth factor-β (TGF-β)/Smad signaling pathway. Further molecular docking experiments showed that the AASAKKKNKKGKTISL peptide (placenta-derived peptide, PDP) derived from the precursor protein IF4B could bind to TGF-β1. Therefore, our preliminary results suggest that the actions of PDP may be mediated through the TGF-β1/Smad signaling pathway. Our results demonstrate that the placental bioactive peptides may regulate the placental function during PE progression.

KEYWORDS

peptides, placenta, preeclampsia, peptidomics, TGF-β/Smad signaling pathway

Introduction

Preeclampsia (PE) is characterized by hypertension and proteinuria after 20 weeks of gestation (Rana et al., 2019). PE is the leading cause of maternal and fetal diseases and death. PE affects 5%–8% of pregnant women in developing countries (Ramos et al., 2017). The progression of PE to eclampsia is associated with severe complications, leading to adverse maternal and fetal outcomes. There is no effective treatment for eclampsia and its definitive treatment is the termination of pregnancy (Phipps et al., 2019). Premature birth and even abortion caused by premature termination of pregnancy are associated with significant burden on the family and society. Therefore, it is necessary to develop additional treatments for PE. PE is prevented by the use of aspirin (Atallah et al., 2017); however, studies indicate that 30%–40% of high-risk pregnant women develop PE despite the use of aspirin (Walsh et al., 2020). Due to the late onset of symptoms and limited treatment options for PE, biomarkers are being developed for earlier detection and treatment (Ma'ayeh and Costantine 2020).

In the previous decade, several studies have focused on the development of diagnostic markers of PE (Stepan et al., 2020). Placental ischemia triggers the pathogenesis of PE and stimulates the release of placental anti-angiogenic factors, such as soluble fms-like tyrosine kinase-1 (sFlt-1) and soluble endoglin (Herraiz et al., 2018). Given the importance of Flt-1 in pregnancy, the *Flt-1* gene may serve as a genetic marker for PE susceptibility. Placental protein 13 (PP13) levels are also elevated in PE (Wu et al., 2021). In addition, blood pressure during pregnancy is positively related to the level of pregnancy-associated plasma protein A (PAPP-A) (Serra et al., 2020). However, it is still unknown whether these markers are effective for screening for PE.

In recent years, endogenous peptides, a class of naturally occurring bioactive molecules, were identified. Several studies have reported their wide tissue distribution and strong biological effects (Czuba et al., 2018). As important regulators of cardiovascular homeostasis, endogenous peptides enter the blood by autocrine or paracrine pathways, participate in blood pressure regulation, and protect the body and reduce injury in pathological states. Human β-defensin 2 selectively acts on calcium-activated potassium (BK) channels to reduce the diastolic and systolic blood pressure and blood flow velocity; because of these effects, β-defensin 2 is expected to be useful as a new drug treatment of hypertension (Liu et al., 2013). Glucagon-like peptide 1 (GLP-1) improves vascular endothelial cell injury and treats diabetic atherosclerosis by regulating the NF-Kb signal pathway (Ma et al., 2021). So far, no study has examined the difference in placental peptidome between normotensive pregnant women and those with mild and severe PE. PE is resolved with the delivery of the placenta (Shaheen et al., 2020), indicating that the placenta is the primary contributor to the pathogenesis of PE (Wang et al., 2019). Therefore, it is essential to screen placental bioactive peptides.

In this study, placental tissues from women with a healthy pregnancy, mild PE, and severe PE were used for peptidomics. The purpose was to screen the differentially expressed peptides (DEPs) of PE and identify the functional peptides that may be useful for the prevention and treatment of PE.

Materials and methods

Sample collection

This study was approved by the ethical research committee of Shanghai Tongren Hospital (2018-060-01). The median age of the patients was 31 years (range: 28–33). The study participants were pregnant singleton women, without previous organic diseases or hypertensive disorders before the pregnancy. The patients fulfilled the criteria for mild or severe PE proposed by the guidelines of the Chinese Society of Obstetrics and Gynecology. PE refers to high blood pressure (systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg on two consecutive occasions) occurring after 20 weeks of pregnancy in women with preterm babies. The control group included healthy women with no signs of hypertension or proteinuria. The placental tissues from three patients with severe PE ($n = 3$), mild PE ($n = 3$), and healthy, gestational age-matched pregnant women ($n = 3$) were collected from the department of Obstetrics and Gynecology at Shanghai Tongren Hospital and stored at −80°C until analysis.

Preparation of samples for LC-MS/MS

Tris-HCl was added at a volume ratio of 1:3, boiled for 10 min, and then cooled in an ice bath. Next, a 100-Hz ultrasonic wave was administered for 5 s with an interval of 5 s followed by 2 min of ultrasonic wave. The sample tube was filled with glacial acetic acid at a final concentration of 1 M, followed by 2 min of vortex oscillation. The supernatant was collected and frozen. A solution of 80% acetone was added, vortexed, shaken, and added to the water bath. Then, the solution was subjected to ultrasonic waves for 2 min at 4°C and 20,000 rpm. Next, the solution was subjected to high-speed centrifugation for 30 min. The supernatant was collected, transferred to a clean centrifugal tube, and freeze-dried. Next, 200 μl of 0.1% TFA solution was added to the solution and desalting was performed on a C18 column using a sample loading of 80 μg. The sample was freeze-dried and subsequently analyzed by LC-MS.

LC-MS/MS analysis

Mass spectrometry was performed using Q Exactive with C18 (3 μm, 150 mm × 5 μm). Mass spectrometry: positive ion

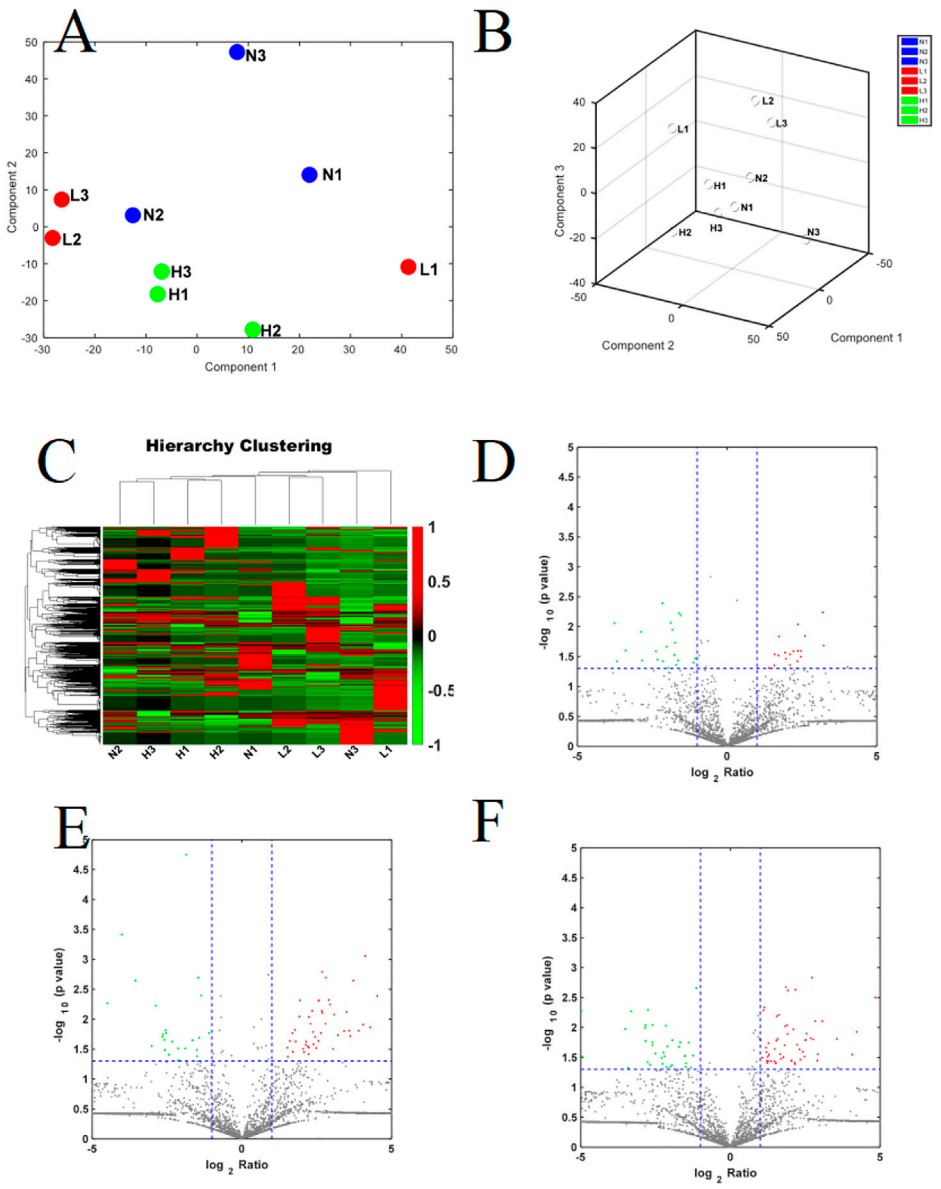


FIGURE 1
(A,B) The PCA score plot of severe PE/control groups, mild PE/control groups; L, mild PE. H, severe PE; N, control group; **(C)** Heatmap of DEPs in the comparison of severe PE, mild PE and control groups; L, mild PE. H, severe PE; N, control group. **(D)** Volcano plot of DEPs in the comparison of mild PE/control group; **(E)** Volcano plot of DEPs in the comparison of severe PE/control group; **(F)** Volcano plot of DEPs in the comparison of severe PE/mild PE; each point in the figure represents a peptide. The red region is the upregulated peptide, and the green region is the downregulated peptide.

detection mode, first-order resolution of 70,000, AGC set to 3e6, and scanning range of 300–1,400 m/z. Top 20 ions were selected for MS/MS analysis. The secondary resolution was 17,500, AGC was set to 5e4, and the isolation window was 3 m/z. For liquid chromatography, the chromatographic column was C18 (3 μm, and 250 mm × 75 μm). The mobile phase A was 0.1% formic acid; mobile phase B was acetonitrile with 0.1% formic acid. The flow rate was 300 nl/min and the injection volume was 6 μl. The mobile phase gradient was 0–8 min for 6%–10% B, 8–60 min for 8%–30% B; 60–79 min for 30%–42% B; 79–80 min for 42%–95% B; 80–85 min for 95% B; 85–86 min for 95%–6% B; and 86–90 min for 6% B. The total running time was 90 min.

Bioinformatics

MAXQUANT (Bruker Daltonics) was used to process the data. Mascot (version 2.5.1) and Swissprot *Homo sapiens*

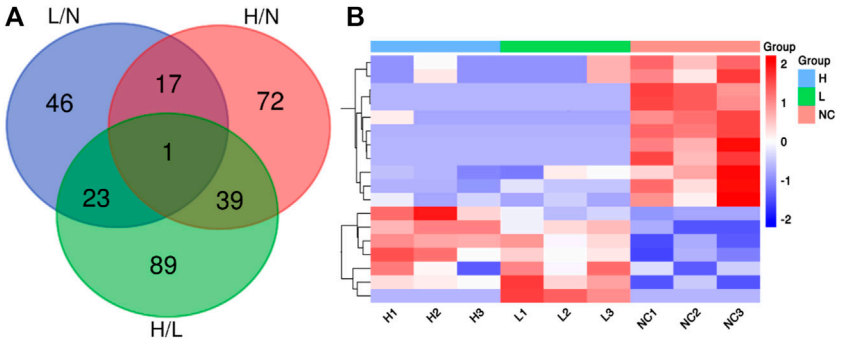


FIGURE 2
(A) Venn diagram showing the overlap of DEPs in the comparison of severe PE/control groups, mild PE/control group and severe PE/mild PE; **(B)** Heatmap analysis of 18 DEPs both in the comparison of severe PE/control group and mild PE/control group. L, mild PE; H, severe PE; N, control group.

TABLE 1 DEPs both in the comparison of mild PE/control and severe PE/control.

Sequence	Entry name	Log ratio (L/N)	*p*-value (L/N)	Log ratio (H/N)	*p*-value (H/N)	Log ratio (H/L)	*p*-value (H/L)
AASAKKKNKKGKTISL	IF4B	−11.10	8.53E-06	−10.69	8.54E-06	0.42	0.15
ADIQTERAYQKQPTIFQNKKRVLL	RS11	−1.12	4.00E-02	−1.35	4.02E-03	−0.23	0.69
AHVDDMPNALSA	HBA	2.60	1.44E-02	3.51	7.54E-03	0.90	0.06
ANRGPAYGLSREVQQKI	TAGL2	−6.76	2.09E-03	−6.34	2.12E-03	0.42	0.15
DQEAIQDLWQWRKSL	NPM	2.37	2.55E-02	2.21	2.86E-02	−0.16	0.54
DVFLGMFLYEYAR	ALBU	−2.16	4.09E-03	−3.55	2.25E-03	−1.39	0.07
GHFTEEDKATI	HBG1	1.94	2.71E-02	4.12	8.84E-04	2.17	0.00
IENEEQEYVQTVK	ANXA1	2.22	2.60E-02	2.95	5.68E-03	0.73	0.05
IENPGFEASPPAQGIPEAKVRHPLS	VISTA	−7.73	1.21E-02	−7.31	1.22E-02	0.42	0.15
LDPITGRSRGFGFVLF	HNRPD	−7.96	1.47E-03	−2.53	1.72E-02	5.43	0.37
QMGTNRGASQAGMTGYGMPRQIL	TAGL2	−2.14	3.80E-02	−2.55	1.53E-02	−0.41	0.85
SGDAAIVDMVPGKPM	EF1A1	−6.27	6.52E-04	−5.85	6.63E-04	0.42	0.15
SGGKYVDSEGHL	CAV1	6.07	1.33E-03	0.25	4.43E-01	−5.82	0.00
SLSPFYLRPPSFLRA	CRYAB	−1.96	4.91E-02	−2.13	3.10E-02	−0.17	0.94
SRNGMVLKPHFHKDWQRRVATWF	RL13	−3.76	8.76E-03	−1.90	3.07E-02	1.86	0.27
STMAFKQMEQISQFLQAAERY	TAGL2	−6.14	9.99E-03	−5.72	1.02E-02	0.42	0.15
VETRDGQVINETSQ	VIME	1.72	3.02E-02	2.55	4.94E-03	0.84	0.01
VSESSDVLPK	K2C8	1.57	4.42E-02	1.17	4.16E-01	−0.40	0.61

L/N, represent the comparison of mild PE/normal; H/N, represent the comparison of severe PE/normal; H/L, represent the comparison of severe PE/mild PE.

database were used for the database search, considering the Oxidation (M) Acetyl (Protein N-term) as variable modifications. The search was conducted to identify non-specifc peptides with at least six amino acids. Other parameters were set as default. There were 3,941 peptide matches. The peak intensity values were normalized by summing the peaks in each sample. The Mann-Whitney U test was used to estimate the differences in peptide representation among the groups. The data were uploaded to the website (http://bioinformatics.psb.ugent.be/webtools/venn/) to construct a Venn diagram. Volcano plots and heatmaps were constructed using the R package. Using the R package, principal component analysis (PCA) was performed to create a classification model: control, mild PE, and severe PE.

Function analyses

Gene ontology (GO) was used to analyze the properties and functions of precursor proteins of DEPs in terms of biological

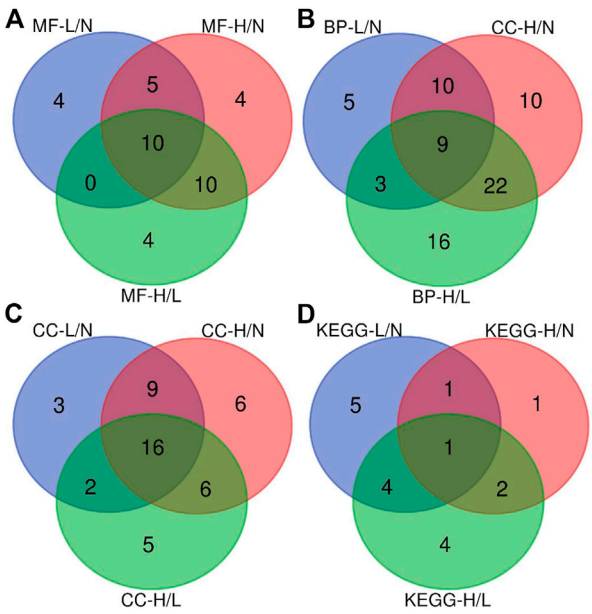


FIGURE 3
Venn diagram showing the overlap of the enriched biological pathways identified using GO and KEGG pathway databases. **(A)**, Venn diagram of the overlap of significant regulated molecular function categories; **(B)**, Venn diagram of the overlap of significant regulated biological process categories; **(C)**, Venn diagram of the overlap of significant regulated cellular component categories; **(D)**, Venn diagram of the overlap of significant regulated KEGG pathway categories. L/N, mild PE/control; H/N, severe PE/control group; H/L, severe PE/mild PE.

processes (BPs), molecular functions (MFs), and cellular components (CCs). The KEGG pathway database was used to identify the most important signaling pathways and metabolic pathways related to the precursor proteins. The interaction analysis of precursor proteins of DEPs was carried out using IPA software. The interaction network of precursor proteins was constructed, and the enrichment analysis of related signal pathways was carried out.

Molecular docking

The peptide (AASAKKKNKKGKTISL) structure was obtained from Maestro 11.9. Then the peptide was imported to the Chem3D software for optimize and minimize the energy by using the MM2 module, and saved as a sdf file as a ligand molecule for molecular docking. The TGF-β1 (PDB ID: 6P7J) protein structure was downloaded from the RCSB database (https://www.rcsb.org/). The protein structure is processed on Maestro11.9 platform. The protein is processed with Schrodinger's Protein Preparation Wizard to remove crystal water, add missing hydrogen atoms, repair missing bond information, and repair missing peptide segments. Finally, the protein was then minimized using the OPLS3 force field. The processing and optimization of molecular docking was completed by Glide module in Schrödinger Maestro software (Schrödinger, New York, NY, United States). While performing the docking of this peptide to TGF-β1, five poses of the ligand, were produced by the SP mode of Glide. The ligand interaction diagram module of Glide was used to analyze ligand-protein interactions.

Results

Identification of differentially expressed peptides

We identified 3,941 peptides. These data were used for multivariate analysis. The three groups were separated by PCA analysis, as shown in Figures 1A,B. Compared to healthy pregnant women, 51 peptides were increased, whereas 36 were decreased, in patients with mild PE (fold change > 2, $p < 0.05$; Supplementary Table S1). Additionally, 82 peptides were increased and 47 peptides were decreased in patients with severe PE (fold change > 2, $p < 0.05$; Supplementary Table S2). Compared to healthy pregnant women, 78 peptides were upregulated and 74 peptides were downregulated in severe PE patients (fold change > 2, $p < 0.05$; Supplementary Table S3). A heatmap and Volcano map were used to display the DEGs (Figures 1C–F). Compared to healthy pregnant women, 18 peptides were significantly changed both in mild and severe PE groups, as shown in the Venn plot and heatmap (Figures 2A,B). Only one peptide (GHFTEEDKATI) was significantly altered in the abovementioned three comparisons (Table 1). GO analyses revealed that many peptides were related to the negative regulation of the apoptotic process, translational initiation, translation, and other extracellular matrices (Supplementary Figure S1).

Bioinformatics

The CCs related to the DEPs included extracellular exosome, focal adhesion, extracellular space, extracellular region, and extracellular matrix. The precursor proteins of DEPs were enriched in oxidative phosphorylation, arrhythmogenic right ventricular cardiomyopathy, oxidative phosphorylation, hypertrophic cardiomyopathy, dilated cardiomyopathy, and tight junction, as well as with the occurrence of PE (Figure 3). The contents of the related PEPs in these pathways among different groups are shown in Figure 4.

Network analysis using IPA showed that the changes in precursor proteins of DEPs were mainly related to the TGF-β/Smad signaling pathway (Figure 5). Further molecular docking experiments showed that the AASAKKKNKKGKTISL peptide

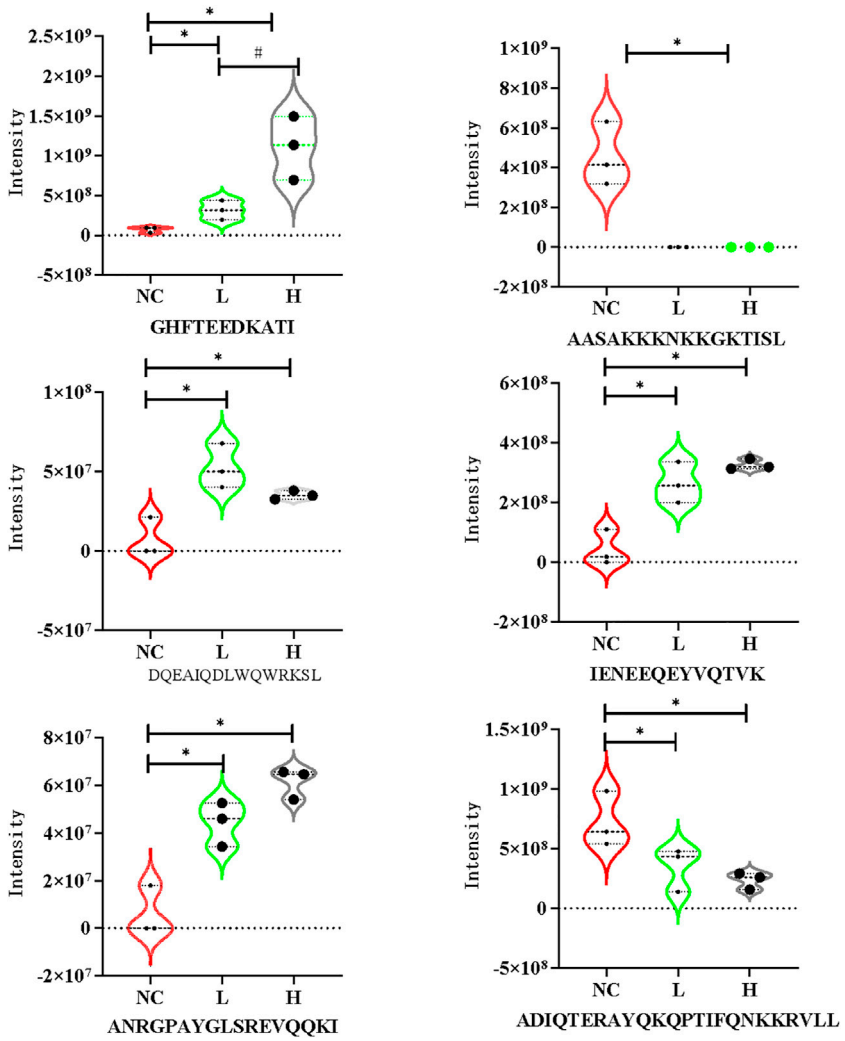


FIGURE 4
Violin plots of DEPs in the mild PE, severe PE and control group. L, mild PE; H, severe PE; NC, control group. *, represent $p < 0.05$ compared to the control group #, represent $p < 0.05$ compared to the mild PE group.

(PDP) derived from the precursor protein IF4B could bind to TGF-β1, suggesting that TGF-β1 may be the target protein of PDP (Figure 6). The docking score of PDP was −28.57 kcal/mol. PDP could be docked with the molecule of TGF-β1 Serine residue at the position of ASN 224.

Discussion

Peptidomics is used to study the peptides in biological samples, which are important because the soluble molecules can provide valuable information about an individual's physical status, changes in behavior, and severity of illness (Foreman et al., 2021). Several studies have evaluated the urine and serum peptides in pregnant women with PE (Law et al., 2015; Dai et al., 2017; Qian et al., 2017; Kononikhin et al., 2020; Starodubtseva et al., 2020). However, no previous study has investigated placental peptides. PE is resolved with the delivery of the placenta (Shaheen et al., 2020); thus, the placenta is the primary contributor to the pathogenesis of PE (Wang et al., 2019). There is evidence that placenta and placental peptides may play an important role in PE. Thus, it is worth investigating the peptidome profiles in the placenta of PE patients.

In our study, several biologically active peptides derived from 480 precursor proteins were altered during PE development. Compared to a previous study, a greater number of peptides was identified in the placenta than in the serum and urine of PE patients (Kononikhin et al., 2020). This may be due to the release of a greater number of endogenous peptides from the placenta, indicating that the greater number of peptides may provide

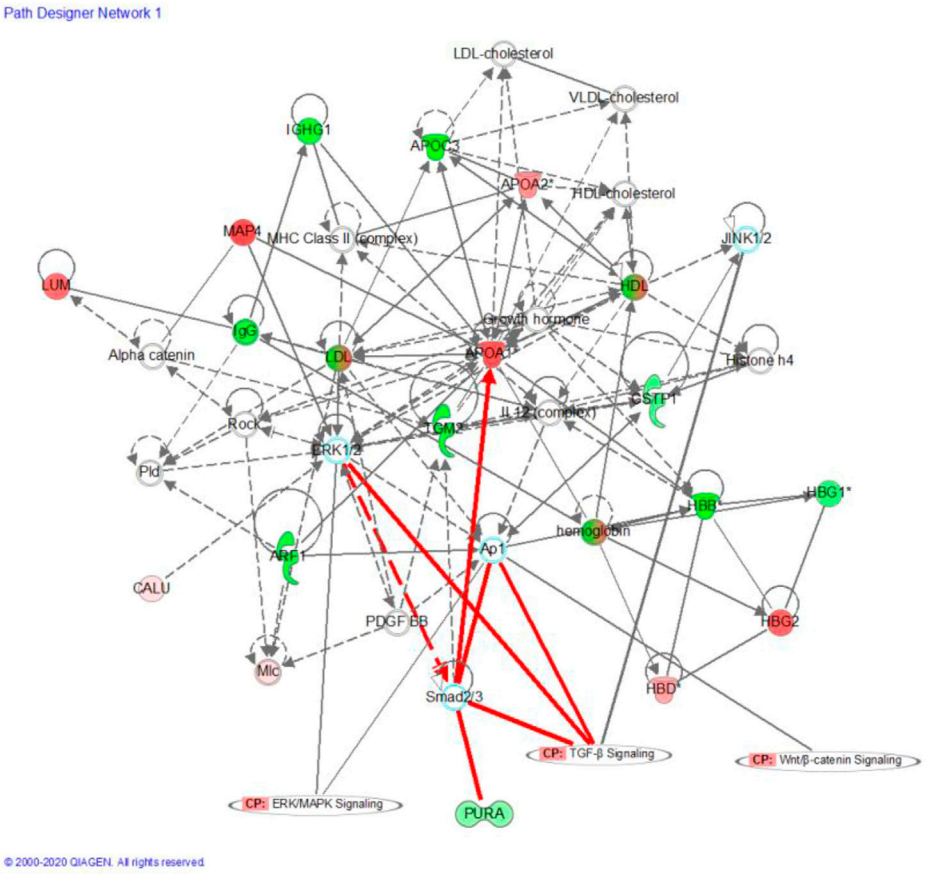


FIGURE 5
Precursor protein of DEPs analysis by Ingenuity Pathway Analysis (IPA); Red and blue dots represent the significantly up- and downregulated DEPs related Precursor Proteins, respectively; "CP" is an abbreviation for "canonical pathway," which represents a signaling pathway related to the highly linked molecules. Indirect relationships are represented by dotted lines and direct relationships by solid lines.

additional options for PE treatment. Compared to healthy pregnant women, those with mild PE showed upregulation of 51 peptides and downregulation of 36 peptides. In patients with severe PE, 82 peptides were upregulated and 47 peptides were downregulated. Of these DEPs, 18 peptides were altered between the severe and mild PE compared to the control groups. The panel peptides were derived from six protein precursors: ANXA1, CAV1, IF4B, RS11, HBA, and TAGL2.

Notably, IENEEQEYVQTVK was increased in severe and mild PE patients; this protein was derived from the protein Annexin A1 (ANXA1). A previous study revealed that the level of the anti-inflammatory protein ANXA1 was also increased in early PE (Perucci et al., 2015; Feng et al., 2018). In our study, the level of the peptide IENEEQEYVQTVK derived from AnxA1 was higher in mild and severe PE compared to healthy pregnant women. The increased AnxA1-derived peptide level may counter the inflammatory response in PE patients. No previous study has evaluated the AnxA1-derived peptide in the placenta from PE and normotensive patients. A prospective study is required to explore the use of AnxA1-derived peptide as a PE biomarker.

The pathways significantly associated with the precursor protein of DEPs primarily included "focal adhesion", "tight junction", and "extracellular matrix". Evidence indicates that PE develops due to the presence of the placenta (Lim et al., 2015; Mohammadpour-Gharehbagh et al., 2018). Placenta maintains the transport between the fetus and mother (Jena et al., 2020; Melchiorre et al., 2022). For fetal development and placental embedding, extravillous trophoblasts must invade the decidua of the mother, while it can be disrupted in conditions related to pregnancy, such as PE (Ma et al., 2020). Abnormal placentation caused by impaired trophoblast function causes pregnancy-associated syndromes, such as PE (Schoots et al., 2018). There is growing evidence to suggest that ECM interactions play a significant role in trophoblast proliferation and differentiation (Wong et al., 2018). In trophoblasts, ECM thickness affects the production of mRNA and proteins related to fusion. Abnormal expression of placental ECM components is

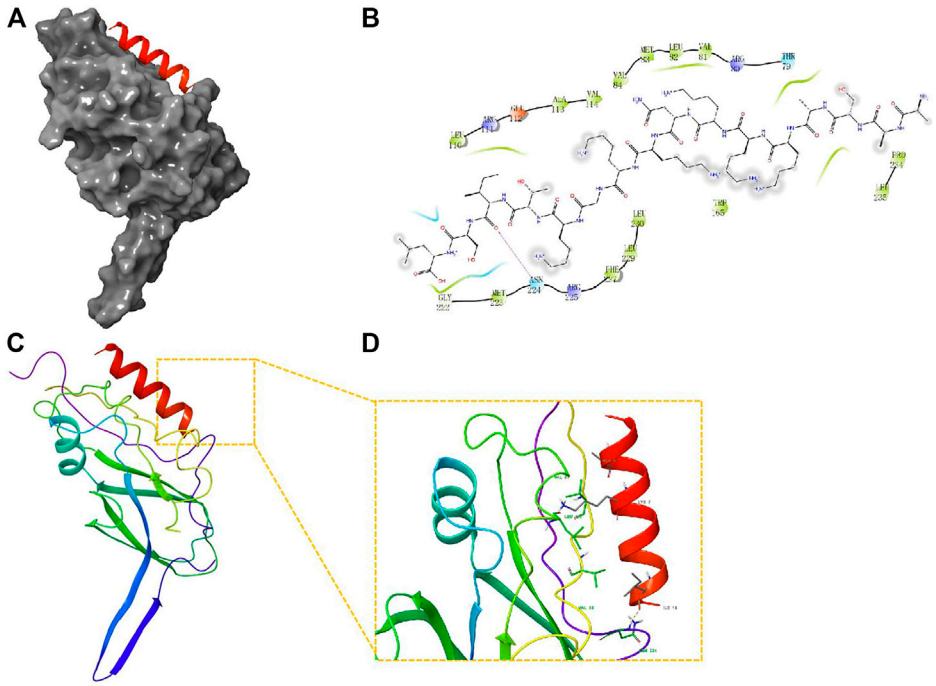


FIGURE 6
Molecular docking study of PDP to TGF-β1. **(A)** The 3D structure of docked molecule binding to TGF-β1; **(B)**, Binding sites of PDP and TGF-β1, the binding site amino acid residue ASN (224); **(C,D)**; Close-up view of binding site of PDP docked with TGF-β.

associated with abnormal migration/invasion of human trophoblast cells (Romanowicz and Galewska 2011; Ma et al., 2020; Liu et al., 2021; Parameshwar et al., 2021). Consequently, PE may develop due to ECM changes.

To uncover the potential efficacy of peptides in PE, network analysis was performed using IPA, which showed that the changes in the precursor proteins of DEPs were mainly related to the TGF-β/Smad signaling pathway. Further molecular docking experiments showed that AASAKKKNKKGKTISL peptide (PDP) derived from the precursor protein IF4B could bind to TGF-β1, suggesting that TGF-β1 may be the target protein of PDP. Thus, targeting the TGF-β/Smad signal pathway is an effective strategy for the treatment of PE. Ligand binding activates the receptors, and TGF-β phosphorylates Smad2 and Smad3 (Xu et al., 2016). Activated Smad2 and Smad3 bind to Smad4 to regulate the transcription of several genes that contribute to trophoblast invasion and proliferation (Brkić et al., 2020). PDP is an endogenous peptide that was identified by peptidomics screening; no functional report has evaluated PDP so far. PDP may act as a potential bioactive peptide to improve PE by acting on the TGF-β signal pathway. PDP has the advantages of low molecular weight, high stability, good lipophilicity, and easy entry into cells and nuclei. PDP may be a new method for the clinical treatment of PE. Therefore, future studies are required to study its function and mechanism.

Conclusion

In the present study, we identified DEPs from the placenta of patients with mild and severe PE. Functional analysis suggested that the precursor proteins of DEPs were mainly involved in PE progression. Further molecular docking experiments showed that PDP peptides derived from the precursor protein IF4B could bind to TGF-β1, suggesting that PDP may play its biological function by affecting the TGF-β1/Smad signaling pathway. Our results suggest some strategies for the discovery of active peptides, which may be helpful for the diagnosis and treatment of PE.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the ethical research committee of Shanghai Tongren Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

TC and ZZ were responsible for paper writing. QL and JM designed this study.

Funding

The grant from the funds for Discipline Construction of Tongren Hospital and the National Natural Science Foundation of China (82101773) supported this research.

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.2022.1014836/full#supplementary-material

References

Atallah, A., Lecarpentier, E., Goffinet, F., Gaucherand, P., and Tsatsaris, V. (2017). Aspirin for prevention of preeclampsia. *Drugs* 77, 1819–1831. doi:10.1007/s40265-017-0823-0

Brkić, J., Dunk, C., Shan, Y., O'Brien, J. A., Lye, P., Qayyum, S., et al. (2020). Differential role of smad2 and smad3 in the acquisition of an endovascular trophoblast-like phenotype and preeclampsia. *Front. Endocrinol.* 11, 436. doi:10.3389/fendo.2020.00436

Czuba, L. C., Hillgren, K. M., and Swaan, P. W. (2018). Post-translational modifications of transporters. *Pharmacol. Ther.* 192, 88–99. doi:10.1016/j.pharmthera.2018.06.013

Dai, X., Song, X., Rui, C., Meng, L., Xue, X., Ding, H., et al. (2017). Peptidome analysis of human serum from normal and preeclamptic pregnancies. *J. Cell. Biochem.* 118, 4341–4348. doi:10.1002/jcb.26087

Feng, J., Wang, X., Li, H., Wang, L., and Tang, Z. (2018). Silencing of annexin a1 suppressed the apoptosis and inflammatory response of preeclampsia rat trophoblasts. *Int. J. Mol. Med.* 42, 3125–3134. doi:10.3892/ijmm.2018.3887

Foreman, R. E., George, A. L., Reimann, F., Gribble, F. M., and Kay, R. G. (2021). Peptidomics: A review of clinical applications and methodologies. *J. Proteome Res.* 20, 3782–3797. doi:10.1021/acs.jproteome.1c00295

Herraiz, I., Llurba, E., Verlohren, S., and Galindo, A. (2018). Update on the diagnosis and prognosis of preeclampsia with the aid of the sFlt-1/PlGF ratio in singleton pregnancies. *Fetal diagn. Ther.* 43, 81–89. doi:10.1159/000477903

Jena, M. K., Sharma, N. R., Petitt, M., Devika, M, and Nihar, RN (2020). Pathogenesis of preeclampsia and therapeutic approaches targeting the placenta Biomolecules. *Biomolecules* 10, 953. doi:10.3390/biom10060953

Kononikhin, A. S., Zakharova, N. V., Sergeeva, V. A., Indeykina, M. I., Starodubtseva, N. L., Bugrova, A. E., et al. (2020). Differential diagnosis of preeclampsia based on urine peptidome features revealed by high resolution mass spectrometry. *Diagnostics* 10, E1039. doi:10.3390/diagnostics10121039

Law, K. P., Han, T. L., Tong, C., and Baker, P. N. (2015). Mass spectrometry-based proteomics for pre-eclampsia and preterm birth. *Int. J. Mol. Sci.* 16, 10952–10985. doi:10.3390/ijms160510952

Lim, R., Acharya, R., Delpachitra, P., Hobson, S., Sobey, C. G., Drummond, G. R., et al. (2015). Activin and nadph-oxidase in preeclampsia: Insights from *in vitro* and murine studies. *Am. J. Obstet. Gynecol.* 212, 86.e1–12. doi:10.1016/j.ajog.2014.07.021

Liu, C., Hu, Y., Wang, Z., Pan, H., Ren, Y., Li, X., et al. (2021). The downregulation of placental lumican promotes the progression of preeclampsia. *Reprod. Sci.* 28, 3147–3154. doi:10.1007/s43032-021-00660-w

Liu, R., Zhang, Z., Liu, H., Hou, P., Lang, J., Wang, S., et al. (2013). Human β-defensin 2 is a novel opener of ca2+-activated potassium channels and induces vasodilation and hypotension in monkeys. *Hypertension* 62, 415–425. doi:10.1161/HYPERTENSIONAHA.111.01076

Ma, X., Liu, Z., Ilyas, I., Little, P. J., Kamato, D., Sahebka, A., et al. (2021). Glp-1 receptor agonists (glp-1ras): Cardiovascular actions and therapeutic potential. *Int. J. Biol. Sci.* 17, 2050–2068. doi:10.7150/ijbs.59965

Ma, Z., Sagrillo-Fagundes, L., Mok, S., Vaillancourt, C., and Moraes, C. (2020). Mechanobiological regulation of placental trophoblast fusion and function through extracellular matrix rigidity. *Sci. Rep.* 10, 5837. doi:10.1038/s41598-020-62659-8

Ma'ayeh, M., and Costantine, M. M. (2020). Prevention of preeclampsia. *Semin. Fetal Neonatal Med.* 25, 101123. doi:10.1016/j.siny.2020.101123

Melchiorre, K., Giorgione, V., and Thilaganathan, B. (2022). The placenta and preeclampsia: Villain or victim? *Am. J. Obstet. Gynecol.* 226, S954–s962. doi:10.1016/j.ajog.2020.10.024

Mohammadpour-Gharehbagh, A., Teimoori, B., Narooei-Nejad, M., Mehrabani, M., Saravani, R., and Salimi, S. (2018). The association of the placental mthfr 3'-utr polymorphisms, promoter methylation, and mthfr expression with preeclampsia. *J. Cell. Biochem.* 119, 1346–1354. doi:10.1002/jcb.26290

Parameshwar, P. K., Sagrillo-Fagundes, L., Fournier, C., Girard, S., Vaillancourt, C., and Moraes, C. (2021). Disease-specific extracellular matrix composition regulates placental trophoblast fusion efficiency. *Biomater. Sci.* 9, 7247–7256. doi:10.1039/d1bm00799h

Perucci, L. O., Carneiro, F. S., Ferreira, C. N., Sugimoto, M. A., Soriani, F. M., Martins, G. G., et al. (2015). Annexin a1 is increased in the plasma of preeclamptic women. *PLoS One* 10, e0138475. doi:10.1371/journal.pone.0138475

Phipps, E. A., Thadhani, R., Benzing, T., and Karumanchi, S. A. (2019). Pre-eclampsia: Pathogenesis, novel diagnostics and therapies. *Nat. Rev. Nephrol.* 15, 275–289. doi:10.1038/s41581-019-0119-6

Qian, Y., Zhang, L., Rui, C., Ding, H., Mao, P., Ruan, H., et al. (2017). Peptidome analysis of amniotic fluid from pregnancies with preeclampsia. *Mol. Med. Rep.* 16, 7337–7344. doi:10.3892/mmr.2017.7582

Ramos, J. G. L., Sass, N., and Costa, S. H. M. (2017). Preeclampsia, *Rev. Bras. Ginecol. Obstet.* 39, 496–512. doi:10.1055/s-0037-1604471

Rana, S., Lemoine, E., Granger, J. P., and Karumanchi, S. A. (2019). Preeclampsia: Pathophysiology, challenges, and perspectives. *Circ. Res.* 124, 1094–1112. doi:10.1161/CIRCRESAHA.118.313276

Romanowicz, L., and Galewska, Z. (2011). Extracellular matrix remodeling of the umbilical cord in pre-eclampsia as a risk factor for fetal hypertension. *J. Pregnancy* 2011, 542695. doi:10.1155/2011/542695

Schoots, M. H., Gordijn, S. J., Scherjon, S. A., Harry van, G., and Jan-Luuk, H. (2018). Oxidative stress in placental pathology. *Placenta* 69, 153–161. doi:10.1016/j.placenta.2018.03.003

Serra, B., Mendoza, M., Scazzocchio, E., Meler, E., Nolla, M., Sabria, E., et al. (2020). A new model for screening for early-onset preeclampsia. *Am. J. Obstet. Gynecol.* 222, e1–e608. doi:10.1016/j.ajog.2020.01.020

Shaheen, G., Jahan, S., Ain, Q. U., Ullah, A., Afsar, T., Almajwal, A., et al. (2020). Placental endothelial nitric oxide synthase expression and role of oxidative stress in susceptibility to preeclampsia in pakistani women. *Mol. Genet. Genomic Med.* 8, e1191. doi:10.1002/mgg3.1191

Starodubtseva, N., Nizyaeva, N., Baev, O., Bugrova, A., Gapaeva, M., Muminova, K., et al. (2020). Serpina1 peptides in urine as a potential marker of preeclampsia severity. *Int. J. Mol. Sci.* 21, E914. doi:10.3390/ijms21030914

Stepan, H., Hund, M., and Andraczek, T. (2020). Combining biomarkers to predict pregnancy complications and redefine preeclampsia: The angiogenic-placental syndrome. *Hypertension* 75, 918–926. doi:10.1161/HYPERTENSIONAHA.119.13763

Walsh, S. W., Reep, D. T., Alam, S. M. K., Washington, S. L., Al Dulaimi, M., Lee, S. M., et al. (2020). Placental production of eicosanoids and sphingolipids in women who developed preeclampsia on low-dose aspirin. *Reprod. Sci.* 27, 2158–2169. doi:10.1007/s43032-020-00234-2

Wang, R., Yu, R., Zhu, C., Lin, H. Y., Lu, X., and Wang, H. (2019). Tubulin detyrosination promotes human trophoblast syncytium formation. *J. Mol. Cell Biol.* 11, 967–978. doi:10.1093/jmcb/mjz084

Wong, M. K., Shawky, S. A., Aryasomayajula, A., Green, M. A., Ewart, T., Selvaganapathy, P. R., et al. (2018). Extracellular matrix surface regulates self-assembly of three-dimensional placental trophoblast spheroids. *PLoS One* 13, e0199632. doi:10.1371/journal.pone.0199632

Wu, Y., Liu, Y., and Ding, Y. (2021). Predictive performance of placental protein 13 for screening preeclampsia in the first trimester: A systematic review and meta-analysis. *Front. Med.* 8, 756383. doi:10.3389/fmed.2021.756383

Xu, J., Sivasubramaniyam, T., Yinon, Y., Tagliaferro, A., Ray, J., Nevo, O., et al. (2016). Aberrant tgfβ signaling contributes to altered trophoblast differentiation in preeclampsia. *Endocrinology* 157, 883–899. doi:10.1210/en.2015-1696

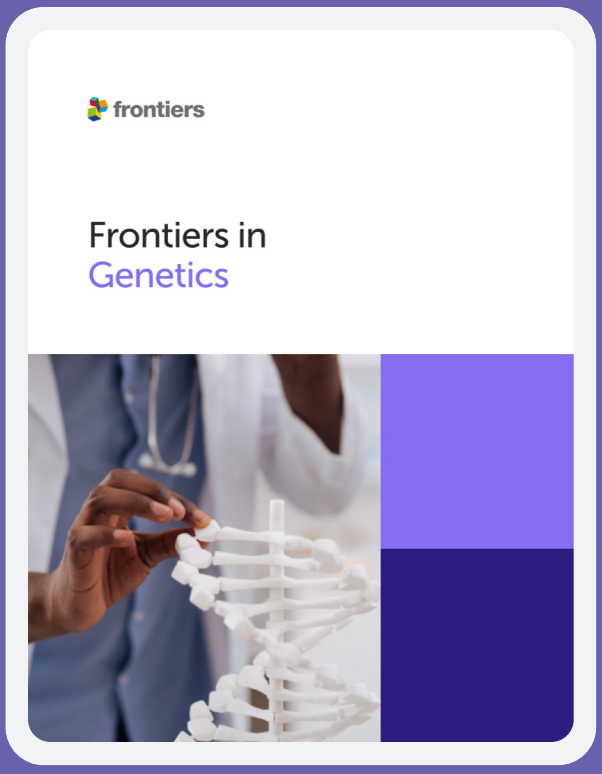
Frontiers in Genetics

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