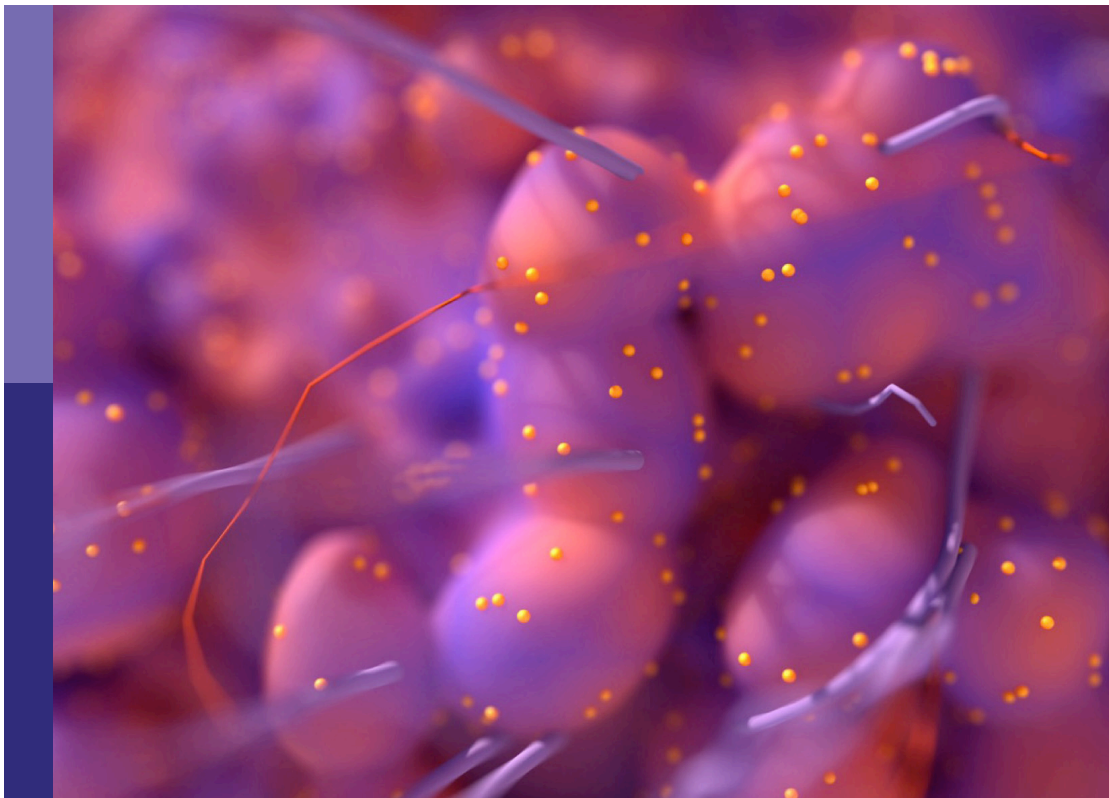
frontiers | Research Topics

Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer.

Edited by
Carlos A. Vaccaro, Mev Dominguez-Valentin and Walter Hernán Pavicic

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-83251-212-8
DOI 10.3389/978-2-83251-212-8

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

Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer.

Topic editors

Carlos A. Vaccaro — Italian Hospital of Buenos Aires, Argentina
Mev Dominguez-Valentin — Oslo University Hospital, Norway
Walter Hernán Pavicic — National Scientific and Technical Research Council (CONICET), Argentina

Citation

Vaccaro, C. A., Dominguez-Valentin, M., Pavicic, W. H., eds. (2023). *Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer.* Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-212-8

Table of contents

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 19 December 2022
DOI 10.3389/fonc.2022.1104671

OPEN ACCESS

EDITED AND REVIEWED BY
Heather Cunliffe,
University of Otago,
New Zealand

*CORRESPONDENCE
Walter Hernán Pavicic
walter.pavicic@
hospitalitaliano.org.ar

SPECIALTY SECTION
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

RECEIVED 21 November 2022
ACCEPTED 30 November 2022
PUBLISHED 19 December 2022

CITATION
Vaccaro CA, Soarez J, Dominguez-Valentin M and Pavicic WH (2022) Editorial: Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer.
Front. Oncol. 12:1104671.
doi: 10.3389/fonc.2022.1104671

Editorial: Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer

Carlos Alberto Vaccaro[1,2], Julieta Soarez[1,2], Mev Dominguez-Valentin[3] and Walter Hernán Pavicic[1,2]*

[1]Instituto de Medicina Traslacional e Ingeniería Biomédica (IMTIB) Hospital Italiano de Buenos Aires, Instituto Universitario Hospital Italiano de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (HIBA-IUHI-CONICET), Ciudad Autónoma de Buenos Aires (CABA), Argentina, [2]Programa de Cáncer Hereditario (Pro.Can.He.), Hospital Italiano de Buenos Aires (HIBA), Ciudad Autónoma de Buenos Aires (CABA), Argentina, [3]Department of Tumor Biology, Institute of Cancer Research, The Norwegian Radium Hospital, Oslo, Norway

KEYWORDS

gastrointestinal cancer, colorectal cancer, pancreatic cancer, genetic predisposition, hereditary syndrome

Editorial on the Research Topic
Advances in genetics and molecular diagnosis in colorectal, stomach, and pancreatic cancer

Gastrointestinal (GI) cancer accounts for over one-quarter (26%) of all cancer cases and over one-third (35%) of cancer deaths globally (1). According to worldwide incidence and mortality rates, the most common cancers include colorectal, stomach, liver, esophageal, and pancreatic cancers (1). Although their incidence has decreased, this group of malignancies continues to pose major challenges to public health. Driven by population growth and aging worldwide, it is estimated that the number of new cases and deaths will increase by 58% and 73%, respectively, by 2040 (1).

Cancer is a complex, heterogeneous, and multifactorial disease, in which both genetic and non-genetic factors are involved. A complete identification of the risk factors, and an understanding of their role at different stages of tumor development, is essential for achieving high-quality comprehensive cancer care. In recent years, the overwhelming revolution in omics technologies, from genomics to transcriptomics, has expanded our understanding of the molecular pathogenesis of GI cancers (2). Despite the advances made in this field, extensive work is still needed to fully understand the molecular and genetic nature that leads to the development of such cancers.

This Research Topic focuses attention on advances in genetic predisposition and novel biomarkers linked to GI cancers, seeking to bring some light to this research field.

Colorectal cancer (CRC) is the most frequently diagnosed GI malignancy, and it is estimated that up to ~35% of all cases are associated with an inherited form attributable

to moderate and high penetrance germline genetic alterations (2, 3). Lynch syndrome and familial adenomatous polyposis (FAP) syndrome are the two most common disorders, accounting for 2-4% and 1% of all CRCs, respectively (4). Although new predisposition genes for FAP have been identified recently, to date, up to 20% of patients with a clear FAP phenotype and as much as 80% with the attenuated form remain genetically unexplained (Olkinuora et al.). Olkinuora et al. confirmed the importance of extending the coverage of genomic studies by NGS to different populations to achieve the goal of identifying new genes that predispose to hereditary CRC. A recurrent variant in the DNA glycosylase gene *NEIL1* was observed in two non-related families from Finland and Argentina, suggesting that pathogenic germline variants in this gene can explain polyposis in FAP cases without *APC* pathogenic alterations. To date, for certain genes, such as *BRCA1*, the existing data to prove their role in CRC development are contradictory. In that regard, the article by Freire et al. reported three non-related CRC cases (families) carrying a heterozygous pathogenic germline variant in the *BRCA1* gene, with a clear family history of cancer. These new data provide more insight to support the role of *BRCA1* in CRC development.

Despite pancreatic cancer ranking fifth among five major GI cancers, its mortality and incidence rates are almost equal, making it an extremely deadly malignancy. Heritability is estimated to account for 21.5% of pancreatic cancer cases, thus it is important to identify the genetic and molecular background behind this population at risk. By performing a GWAS analysis and a recessive genetic model, Lu et al. identified a set of SNPs showing an association with pancreatic cancer risk (specific recessive effect; OR, 0.75-1.42, $p<10^{-5}$). This study also emphasized the importance of choosing the right bioinformatic tool to improve the definition of polygenic risk scores.

For the treatment of most cancers, in particular for GI malignancies such as CRC, a combination of surgery with adjuvant therapy is widely used. However, the prognostic and risk factors for these patients have not yet been clarified. Furthermore, owing to a lack of effective markers, it is also not possible to infer their prognosis by only having data on genetic alterations in primary tumor tissue. To solve these problems, Yu et al. investigated the expression and protein levels of matrix metalloproteinase (MMP) gene family members. They found that a subgroup of MMP genes could have the potential to work as a target for precision therapies in CRC patients. More precisely, modifying their expression levels could significantly improve both progression-free survival and relapse-free time. In the same line of prognostic research, but for pancreatic adenocarcinomas (PAAD), Xu et al. investigated invasive-related genes with the aim of identifying different molecular subtypes of PAAD. Based on a molecular signatures analysis, a three-gene model was constructed and validated as a marker for assessing the prognostic risk of such patients.

Latest epidemiological data support the notion that at least 90% of cancer deaths from solid tumors are caused by metastases (5). Cancer metastases to the liver over a course of CRC occur in more than 50% of patients (6). To date, only a small number of cases have been candidates for a multimodal treatment, for which high drug resistance and relapse rates have been observed (7). As the mechanisms linked to those events are poorly understood, Kyrochristos et al. conducted a pilot study to assess the potential translational implications of intrapatient heterogeneity as a tool for defining appropriate matched drug therapy. Although it was a proof-of-concept approach, they provided strong evidence to support further evaluation, even in future clinical trials.

To fully understand GI cancer carcinogenesis, uncovering molecular data at each stage is essential. Several authors approached this topic by submitting review and original articles. Lv et al. described that, in CRC, Fibrillin-1 gene methylation can serve as a biomarker for disease development, as its transcriptional inactivation might start from normal colonic epithelium and increase through adenoma to CRC. Using a meta-analysis, Zhang et al. revealed that the Leptin G19A polymorphism may decrease the risk for CRC and esophageal GI cancers, as well as urinary tract cancer. Furthermore, Gao et al. demonstrated that a polymorphism (rs7158663) in the maternally expressed three-lncRNAs can be used as a genetic marker for predicting the risk of GI (CRC and gastric) and breast cancers. Finally, Mulder et al., in their submitted article, underlined that the development of orphan medicinal products for PAAD treatment should be closely monitored to increase the success rate of drugs reaching the market.

All of the collected articles support the notion that a deepened understanding of the molecular and genetic factors linked to the development of GI cancers could help to improve the translation of basic scientific research into clinical services and practices, pointing towards personalized medicine, and thus, more effective treatment of GI malignancies.

Author contributions

All of the listed authors equally, substantially, and intellectually contributed to this Editorial and approved it for publication.

Funding

The authors acknowledge funding from the Agencia Nacional de Promoción Científica y Tecnológica (grant number PICT_2017 -3210 to CV and -3197 to WP). The Agencia Nacional de Promoción Científica y Tecnológica offered a paid doctoral student fellowship to JS. MD-V

acknowledges funding from the Norwegian Cancer Society (contract 194751-2017).

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

1. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global burden of 5 major types of gastrointestinal cancer. *Gastroenterology* (2020) 159:335–49.e15. doi: 10.1053/j.gastro.2020.02.068

2. Olkinuora AP, Peltomäki PT, Aaltonen LA, Rajamäki K. From APC to the genetics of hereditary and familial colon cancer syndromes. *Hum Mol Genet* (2021) 30:R206–24. doi: 10.1093/hmg/ddab208

3. Katabathina VS, Menias CO, Khanna L, Murphy L, Dasyam AK, Lubner MG, et al. Hereditary gastrointestinal cancer syndromes: Role of imaging in screening, diagnosis, and management. *RadioGraphics* (2019) 39:1280–301. doi: 10.1148/rg.2019180185

4. Samadder NJ, Baffy N, Giridhar KV, Couch FJ, Riegert-Johnson D. Hereditary cancer syndromes–a primer on diagnosis and management, part 2: Gastrointestinal cancer syndromes. *Mayo Clin Proc* (2019) 94:1099–116. doi: 10.1016/j.mayocp.2019.01.042

5. Dillekås H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? *Cancer Med* (2019) 8:5574–6. doi: 10.1002/cam4.2474

6. Zarour LR, Anand S, Billingsley KG, Bisson WH, Cercek A, Clarke MF, et al. Colorectal cancer liver metastasis: Evolving paradigms and future directions. *Cell Mol Gastroenterol Hepatol* (2017) 3:163–73. doi: 10.1016/j.jcmgh.2017.01.006

7. Nordlinger B, Sorbye H, Glimelius B, Poston GJ, Schlag PM, Rougier P, et al. Perioperative FOLFOX4 chemotherapy and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC 40983): long-term results of a randomised, controlled, phase 3 trial. *Lancet Oncol* (2013) 14:1208–15. doi: 10.1016/S1470-2045(13)70447-9

frontiers in Oncology

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fonc.2021.771099

Comprehensive Analysis of the Expression and Prognosis for MMPs in Human Colorectal Cancer

Jing Yu[1,2†], *Zhen He*[1,2†], *Xiaowen He*[1,2], *Zhanhao Luo*[2], *Lei Lian*[1,2], *Baixing Wu*[3], *Ping Lan*[1,2*] *and Haitao Chen*[4,5*]

[1] Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2] Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Guangdong Institute of Gastroenterology, Guangzhou, China, [3] Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Guangdong-Hong Kong Joint Laboratory for Ribose Nucleic Acid (RNA) Medicine, Ribose Nucleic Acid (RNA) Biomedical Institute, Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China, [4] School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China, [5] School of Public Health, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China

Background: Previous study implicated that genes of matrix metalloproteinase (MMP) family play an important role in tumor invasion, neoangiogenesis, and metastasis. However, the diverse expression patterns and prognostic values of 24 MMPs in colorectal cancer are yet to be analyzed.

Methods: In this study, by integrating public database and our data, we first investigated the expression levels and protein levels of MMPs in patients with colorectal cancer. Then, by using TCGA and GEO datasets, we evaluated the association of MMPs with clinicopathological parameters and prognosis of colorectal cancer. Finally, by using the cBioPortal online tool, we analyzed the alterations of MMPs and did the network and pathway analyses for MMPs and their nearby genes.

Results: We found that, MMP1, MMP3, MMP7, MMP9–MMP12, and MMP14 were consistently upregulated in public dataset and our samples. Whereas, MMP28 was consistently downregulated in public dataset and our samples. In the clinicopathological analyses, upregulated MMP11, MMP14, MMP16, MMP17, MMP19, and MMP23B were significantly associated with a higher tumor stage. In the survival analyses, upregulated MMP11, MMP14, MMP17, and MMP19 were significantly associated with a shorter progression-free survival (PFS) time and a shorter relapse-free (RFS) time.

Discussion: This study implied that MMP11, MMP14, MMP17, and MMP19 are potential targets of precision therapy for patients with colorectal cancer.

Keywords: colorectal cancer, MMPs, prognosis, expression, tumor stage

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Weifeng Ding, Nantong University, China
Di Peng, Huazhong University of Science and Technology, China

***Correspondence:**
Haitao Chen
chenht56@mail.sysu.edu.cn
Ping Lan
lanping@mail.sysu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *05 September 2021*
Accepted: *13 October 2021*
Published: *05 November 2021*

Citation:
Yu J, He Z, He X, Luo Z, Lian L, Wu B, Lan P and Chen H (2021) Comprehensive Analysis of the Expression and Prognosis for MMPs in Human Colorectal Cancer. Front. Oncol. 11:771099. doi: 10.3389/fonc.2021.771099

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of worldwide cancer mortality. It accounts for 9.2% of all cancer deaths according to the Global Cancer Statistics 2020 (1). In the USA, according to the SEER database, those with CRC have an overall 5-year survival rate of ~64%, primarily dependent on pathological stage at diagnosis. CRC patients diagnosed with disease limited to the colon have greater than 90% 5-year survival rate. Five-year survival decreases to ~70% with regional spread, and for patients diagnosed with distant metastases, the 5-year survival rate drops to 12.5% (2). Despite the significant advances in screening and diagnosis, there are limited therapeutic options for patients with advanced disease, which highlight the need for additional tumor molecular markers and prognostic predictors (3).

The human matrix metalloproteinases (MMPs) family belongs to the metzincin superfamily. The main function of MMPs is catalyzing the proteolytic activities and aiding breakdown of the extracellular matrix (ECM) (4). By degrading connective tissue between cells and in the lining of blood vessels, they enable tumor cells to escape from their original location and seed metastases (5). A large body of experimental and clinical evidence has implicated MMPs in tumor invasion, neoangiogenesis, and metastasis (6). Also, from the 1990s to early 2000s, inhibitors of MMPs (MMPI) were studied in various cancer types. However, despite strongly promising preclinical data, all trials failed due to lack of efficacy and severe side effects (7–9). One important reason to explain the failure is that some MMPs have antitumor effects, while the broad-spectrum MMPIs used in the initial trials might block these MMPs and result in tumor progression (10). Recently, with growing knowledge of MMPs in tumor invasion and metastasis and broader roles in cancer biology, narrow-spectrum MMPIs which were safer and more selective were currently being developed (11).

MMPs play complex and distinct roles in CRC. To date, 24 MMPs (MMP1, MMP2, MMP3, MMP4, MMP7, MMP8, MMP9, MMP10, MMP11, MMP12, MMP13, MMP14, MMP15, MMP16, MMP17, MMP19, MMP20, MMP21, MMP23a/MMP23b, MMP24, MMP25, MMP26, MMP27, and MMP28) were identified. For MMP1, Sunami et al. found that the expression of MMP1 was significantly correlated with hematogenous metastasis of colorectal cancer, which were further supported by research made by Shiozawa et al. and Bendardaf et al. (12–14) MMP2 and MMP9 comprise the gelatinase subfamily of MMPs. Marcus et al. found that the concentrations of MMP2 protein expression in tumor tissue were significantly higher than that in tumor-free tissue. In addition, the lymph node status was correlated with the expression of MMP2 in plasma, that is, the expression of MMP2 was significantly increased in patients with lymph node metastasis compared with those without (15). MMP7, also known as matrilysin, is frequently overexpressed in human cancer tissues. Adachi et al. found that the expression of MMP7 correlated significantly with the presence of nodal or distant metastases (16, 17). Another member of the gelatinase subfamily, MMP9, was expressed at significantly higher ratios in the sera of persons with CRC compared with normal controls. Overexpression of p38 gamma MAPK was shown to increase MMP9 transcription, enhancing cell invasion (18). Whereas, TGF-β receptor kinase inhibitors can reduce expression of MMP9 and block CRC metastasis to the liver (19, 20). However, for colitis-associated colon cancer, MMP9 has a protective role and acts as a tumor suppressor (21). MMP12, also called metalloelastase, was reported to be associated with both reduced tumor growth and increased overall survival (22). MMP13, sharing structural homology with MMP1, was reported to be associated with advanced cancer stage, and its overexpression can increase the risk of postoperative relapse (23). In addition to the MMPs mentioned above, MMP3, MMP11, and MMP14 were also found to be highly expressed in malignant tumors as compared with normal tissue (24–26).

As previously described, the relationship between MMPs and the prognosis of human CRC was only partly reported. By integrating state-of-art databases, we conducted a systematical analysis for all 24 human MMPs. Differential expression analyses were implemented in public database and our samples. Prognosis analyses were evaluated in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets. Pathway and network analyses were further used to investigate the mechanisms underlying them. To the best of our knowledge, this is among the first bioinformatic analyses to comprehensively evaluate all 24 MMPs in CRC.

METHODS

Ethics Statement

This study was approved by the Academic Committee of Sun Yat-Sen University, and it was conducted according to the principles expressed in the Declaration of Helsinki.

Differential Expression Analyses by Oncomine

Oncomine is an online cancer microarray database (https://www.oncomine.org/resource/login.html). Gene expression array datasets from Oncomine were used to analyze the transcription levels of MMPs in different cancers. Differential gene expression analyses of all MMPs were implemented between cancer samples and normal controls. *p*-value was calculated using Student's *t*-test. Cutoffs of *p*-value and fold change were 0.01 and 1.5, respectively.

Differential Expression Analyses by GEPIA

Gene Expression Profiling Interactive Analysis (GEPIA) is an interactive web server which was developed by Tang et al. (27) By using a standard processing pipeline, they analyzed the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples. GEPIA provides customizable tumor/normal differential expression analysis, profiling according to cancer types. Cutoff of *p*-value and fold change used in GEPIA were 0.01 and 2, respectively.

Validation by Quantitative Real-Time Polymerase Chain Reaction

All fresh frozen tissues were archived from The Sixth Affiliated Hospital of Sun Yat-Sen University. The related protocol of human sample usage and the informed consent was approved by the Ethical Review Board of the The Sixth Affiliated Hospital of Sun Yat-Sen University.

Total RNA was extracted from the tumor and normal tissues of 12 patients using Total RNA Kit (Vazyme, China) according to the manufacturer's instruction. Detailed information of these 12 patients can be found in **Supplementary Table S1**. For cDNA synthesis, 1 μg total RNA was reverse-transcribed into cDNA by Hiscript@ III RT Super Mix with gDNA wiper (Vazyme, Nanjing, China). Quantitative PCR reaction was then performed using 2×SYBR mix (Vazyme, China) and the reaction was run on Applied Biosystems 7500 Real-time PCR system. The Ct values obtained from different samples were compared using the 2-ΔΔCt method. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) served as internal reference genes. Sequence information of all used primers is listed in **Supplementary Table S2**.

Protein Levels in UALCAN

UALCAN is a comprehensive, user-friendly, and interactive web resource for analyzing cancer OMICS data. It is built on PERL-CGI with high-quality graphics using JavaScript and CSS (http://ualcan.path.uab.edu/index.html) (28). Using UALCAN, we evaluated the protein level of MMPs in cancer tissue and normal tissue of colorectal cancer patients.

Protein Level of MMPs in Our Samples

For preparation of protein extracts, 12 pairs of cancer and adjacent normal tissues were crushed with a mortar under ice cold conditions and lysed with RIPA lysis buffer together with protease inhibitors. Cells were collected and lysed with RIPA lysis buffer together with protease inhibitors. After centrifugation at 12,000 rpm at 4°C for 20 min, supernatants were collected and protein concentration was determined using the Pierce™ BCA protein assay (Thermo, Waltham, MA, USA). Proteins were separated by electrophoresis on a 10% SDS-polyacrylamide gel, electroblotted onto a PVDF membrane, and blocked by 5% nonfat dry milk for 1 h. Membranes were then washed in TBST three times for 5 min and then incubated with anti-MMP1 (Abcam, Cambridge, MA, USA), anti-MMP2 (Abcam), anti-MMP3 (Abcam, USA), anti-MMP7 (Abcam, USA), anti-MMP8 (Abcam, USA), anti-MMP9 (Invitrogen, Waltham, MA, USA), anti-MMP11 (Bioss, Beijing, China), anti-MMP12 (Abcam, USA), anti-MMP14 (Abcam, USA), anti-MMP17 (Abcam, USA), anti-MMP19 (Bioss, China), anti-MMP28 (Abcam, USA), anti-Collagen (Abcam), anti-TIMP2 (Bioss, China), or anti-GAPDH (Abcam). Subsequently, the membranes were washed with PBST and incubated for 1 h with goat anti-rabbit IgG (Abcam). Finally, membranes were washed three times and immunoreactivity was determined by using a Chemi DOC™ XRS + system (BioRad Laboratories, Hercules, CA, USA).

Clinicopathological and Survival Analyses

By integrating TCGA dataset and standardized survival endpoints defined by Liu et al. recently, we performed clinicopathological and survival analyses (29). Nonparametric Kruskal-Wallis test was used to evaluate the association of American Joint Committee on Cancer (AJCC) stage of colorectal cancer (stage I, stage II, stage III, and stage IV) with the expression of MMPs. Four kinds of survival analyses were implemented, including overall survival (OS), disease-specific survival (DSS), disease-free survival (DFS) also called disease-free interval (DFI), and progression-free survival (PFS) also called progression-free interval (PFI). Disease-free survival is a concept used to describe the period after a successful treatment during which there are no signs and symptoms of the disease that was treated. In addition, by using the GEO dataset GSE39582, we did a relapse-free survival (RFS) analyses (30). As MMP4, MMP23A/MMP23B were not included in the GSE39582 dataset, only 22 MMPs were analyzed in the RFS analyses. Samples were split into two groups by median expression (high vs. low expression), and Kaplan-Meier plot were depicted (denoted with log rank *p*-value). Hazard ratio (HR) and 85% confidence intervals (CIs) were calculated by multivariate Cox regression adjusting the effect of age at diagnosis and sex.

TCGA Data and cBioPortal

TCGA collected many types of data for each of over 20,000 tumor and normal samples (31). The colorectal cancer dataset, including data from 640 cases with pathology reports, was selected for further analyses of MMPs using cBioPortal (http://www.cbioportal.org/). The genomic profiles included mutations, putative copy number alterations (CNAs) from genomic identification of significant targets in cancer (GISTIC), mRNA expression Z scores (RNA-seq v.2 RSEM), and protein expression Z scores (reversed-phase protein array (RPPA)). Coexpression and network were calculated according to the cBioPortal's online instructions. By using the expression data in TCGA, we also calculated the correlation of MMPs with each other and several cancer-associated genes, including MYC, TP53, cyclin-D, as well as CDK4/6. The correlation coefficient was calculated using Spearman's method.

siRNA Transfection

HCT116 were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, USA), penicillin (100 U/ml), and streptomycin (100 μg/ml) at 37°C in a humidified CO_2 (5%) atmosphere. MMP11, MMP14, MMP17, MMP19, small interfering RNA (siRNA), and nontargeting siRNA (si-control) were purchased from Ribobio (Guangzhou, China) and used at 20 mM. Opti-MEM transfection media and Lipo3000

(Invitrogen) were used to transfect the cells once they reached 50% confluency. Knockdown was assessed by Western blotting after 48 h of transfection. Sequence information of all used primers is listed in **Supplementary Table S3**.

RESULTS

Transcriptional Levels of MMPs in Patients With Colorectal Cancer

By using the Oncomine database, we did a Pan-cancer differential gene expression analyses for all MMPs. As shown in **Figure 1**, MMP1–MMP4, MMP7–MMP14, and MMP24 were significantly upregulated in colorectal cancer samples, while MMP15, MMP17, MMP19, and MMP24–MMP28 were significantly downregulated in colorectal cancer samples. Detailed performance of each MMP in Oncomine database can be found in **Supplementary Tables S4**, **S5**.

We then used GEPIA to compare the expression level of all MMPs between colorectal tumor tissue and normal tissue. As shown in **Figure 2**, we found that MMP1, MMP3, MMP7, MMP9–MMP12, and MMP14 were significantly upregulated in tumor tissue, while MMP28 was significantly downregulated in tumor tissue.

We further validated the expression level of MMPs in 12 colorectal cancer patients which were recruited from our hospital (including seven patients with colon cancer and five patients with rectal cancer, detailed information can be found in **Supplementary Table S1**) and measured the expression level of 24 MMPs in their tumor tissue and adjacent normal tissue by quantitative real-time polymerase chain reaction (qRT-PCR). As shown in **Figure 3**, we found that MMP1, MMP3, MMP7, MMP9-MMP12, and MMP14 were significantly upregulated in tumor tissue, while MMP15–MMP17, MMP19–MMP21, MMP23A, MMP23B, and MMP25–MMP28 were significantly downregulated in tumor tissue.

In summary, MMP1, MMP3, MMP7, MMP9–MMP12, and MMP14 were consistently upregulated in Oncomine, GEPIA, and our samples. Thus, MMP28 was consistently downregulated in Oncomine, GEPIA, and our samples.

Protein Levels of MMPs in Patients With Colorectal Cancer

By using the UALCAN database, we further evaluated the protein levels of MMPs in patients with colorectal cancer. As some proteins were not included in UALCAN, we can only do the analyses for MMP1, MMP2, MMP3, MMP7, MMP8, MMP9, MMP12, MMP14, and MMP28. As shown in **Figure 4A**, the protein level of MMP1, MMP2, MMP3, MMP7, MMP8, MMP9, MMP12, and MMP14 in colorectal tumor tissue were significantly higher than that in normal tissue, while the protein level of MMP28 in tumor was significantly lower than that in normal tissue.

We also evaluated the protein level of MMPs in our patients and measured the expression level of MMP1–MMP3, MMP7–MMP9, MMP11, MMP12, MMP14, MMP17, MMP19, and MMP28 in their tumor tissue and adjacent normal tissue by Western blot. As shown in **Figure 4B**, we found that the protein level of MMP2, MMP7, MMP9, MMP12, and MMP14 in the tumor tissue were basically higher than that in the normal tissue.

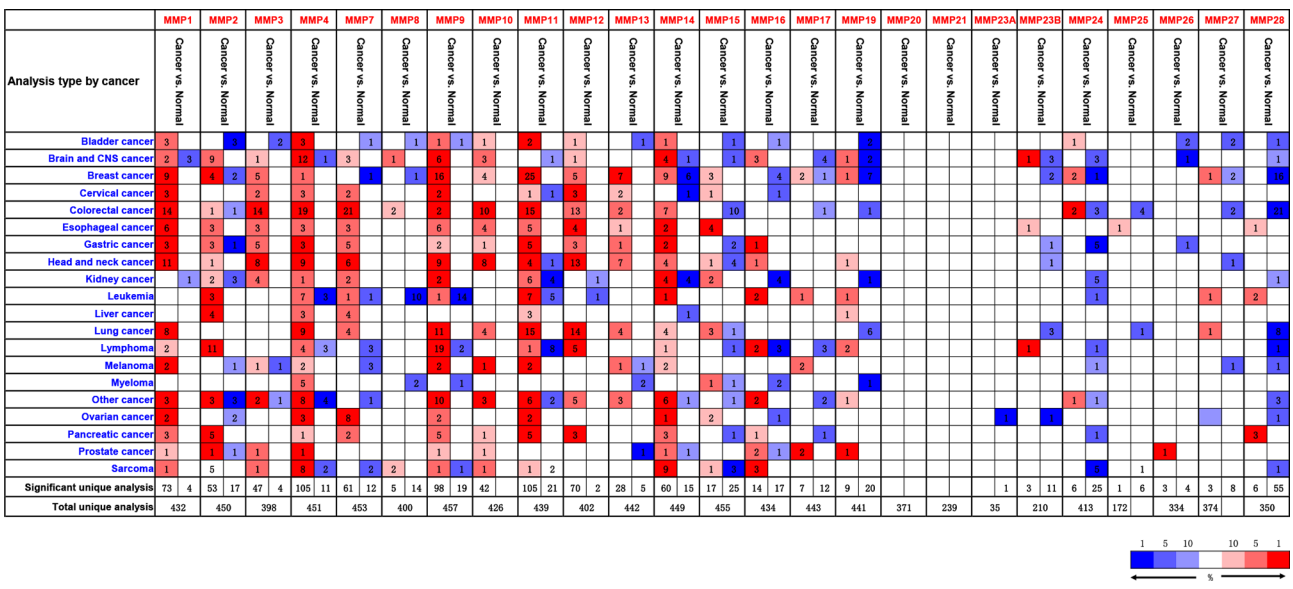


FIGURE 1 | The transcription levels of MMPs in different types of cancers (Oncomine). Upregulated records are highlighted in red, while downregulated records are highlighted in blue. The number in each block means the number of unique analyses, which were fully described in **Supplementary Tables S4**, **S5**.

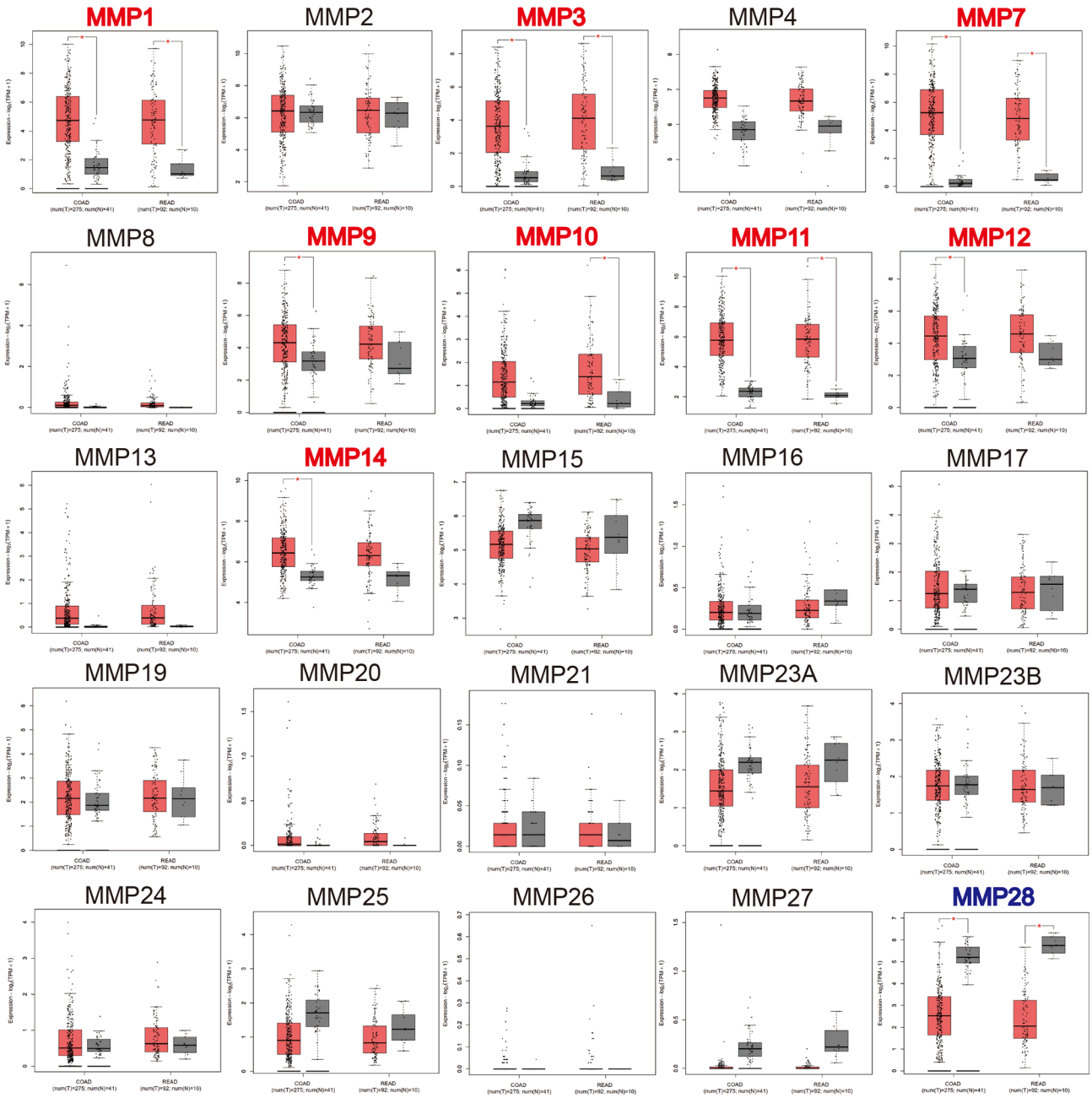


FIGURE 2 | The transcription levels of MMPs in colorectal cancer tissue and normal tissue (GEPIA). Significant records are denoted by the red asterisk on top of the boxplot. The titles of upregulated records are highlighted in red, while the titles of downregulated records are highlighted in blue.

Relationship Between the mRNA Levels of MMPs and the Clinicopathological Parameters of Patients With Colorectal Cancer

By using the TCGA dataset, we analyzed the association of MMP expression with the AJCC stage of colorectal cancer. As shown in **Figure 5**, MMP11, MMP14, MMP16, MMP17, MMP19, and MMP23b were positively correlated with the tumor stage, that is, the mRNA levels of MMPs in patients with higher tumor stage were always high. Detailed information can be seen in **Supplementary Table S6**. Take MMP14 as an example, the mean expression level (log2(normalized count of MMP14)) were 8.43, 8.56, 8.74, and 8.72 for stage I, stage II, stage III, and stage IV patients, respectively (p = 0.007).

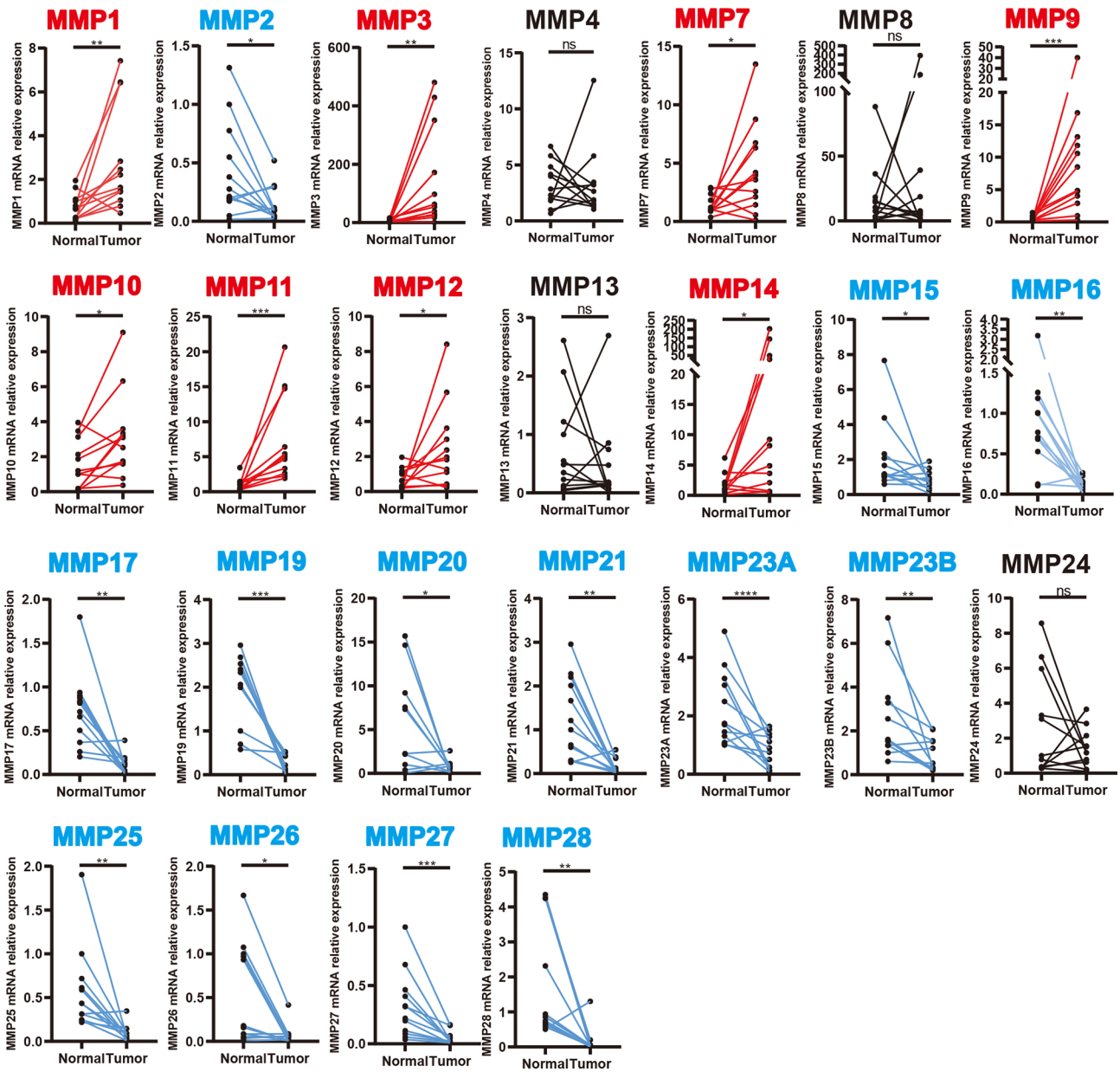


FIGURE 3 | Transcription levels of MMPs in colorectal cancer tissue and adjacent normal tissue (12 colorectal cancer patients in our hospital). Significant records are denoted by the red asterisk on top of the plot ($^{ns}p > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ and $^{****}p < 0.0001$). Upregulated records are highlighted in red, while downregulated records are highlighted in blue.

Association of the mRNA Expression of MMPs With the Prognosis of Patients With Colorectal Cancer

By integrating the TCGA data and four standardized survival endpoints defined by Liu et al. in 2018, we further performed the OS, DSS, DFS, and PFS analyses for all MMPs (**Supplementary Figures S1–S4**; **Table 1**). In the OS analyses, upregulated MMP11, MMP16, MMP17, MMP19, and MMP23B were significantly associated with a shorter overall survival time (**Table 1**; **Supplementary Figure S1**); in the DSS analyses, upregulated MMP14, MMP16, MMP17, MMP19, and MMP23B were significantly associated with a shorter disease-specific survival time (**Table 1**; **Supplementary Figure S2**); in the PFS analyses, upregulated MMP11, MMP14, MMP16, MMP17, MMP19, and MMP23B were significantly associated with a shorter progression-free time (**Table 1**; **Supplementary Figure S3**); and in the DFS analyses, downregulated MMP1, MMP3, MMP9, and MMP12 were significantly associated with a shorter disease-free period

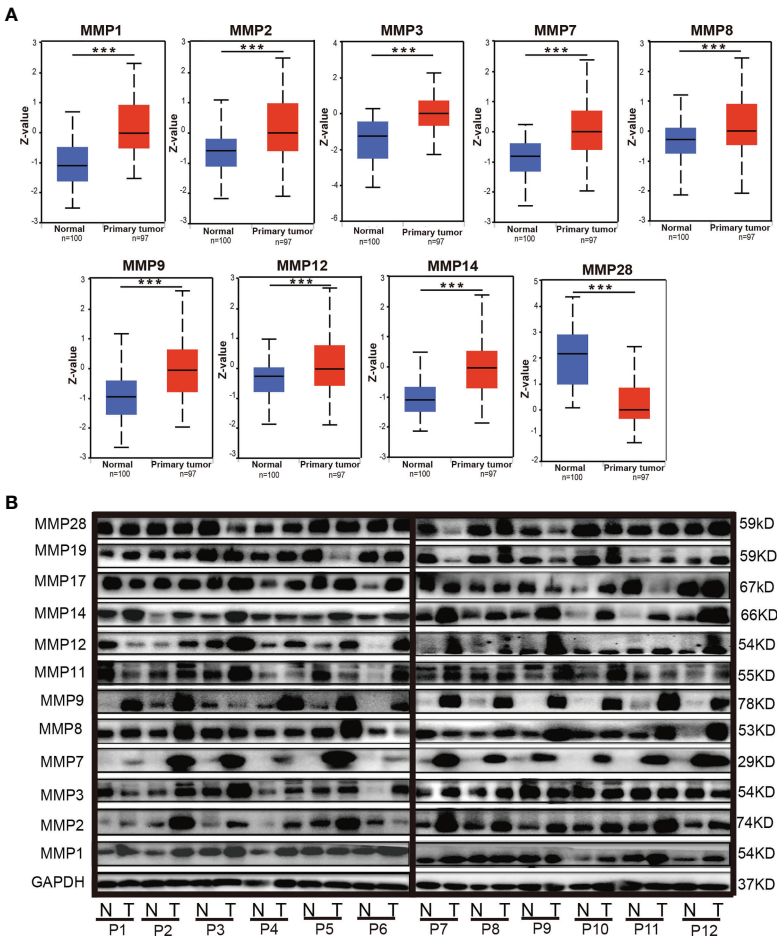


FIGURE 4 | The protein levels of MMPs in colorectal cancer tissue and normal tissue. **(A)** The protein levels of MMPs in UALCAN database; significant records are denoted by the red asterisk on top of the boxplot (***$p < 0.001$). **(B)** The protein levels of MMPs in our 12 samples, which were measured by Western blotting (WB).

(**Table 1**; **Supplementary Figure S4**). By using the GEO dataset, we further performed the RFS analyses. As shown in **Supplementary Figure S5** and **Table 1**, upregulated MMP2, MMP11, MMP14, MMP17, MMP19, MMP24, and MMP28 were significantly associated with a shorter relapse-free time, while the downregulated MMP8, MMP13, MMP16, MMP20, and MMP27 was significantly associated with a shorter relapse-free survival time.

Prediction Function and Pathways of the Changes in MMPs and Their Frequently Altered Neighbor Genes in Patients With Colorectal Cancer

We analyzed the MMP alterations and networks by using the cBioPortal online tool for colorectal cancer. As shown in **Figure 6**, of these 220 colorectal cancer patients, MMPs were altered in more than 30% of them (**Figure 6A**). The top 5 altered genes were MMP24 (10%), MMP9 (9%), and MMP16 (5%) (**Figure 6B**). As shown in **Supplementary Figure S6**, we also calculated the correlation of MMPs with each other and several cancer-associated genes, including MYC, TP53, cyclin-D, as well as CDK4/6. We found that multiple MMPs including MMP1, MMP3, MMP4, MMP7, MMP8, and MMP10–MMP14 were positively correlated with the expression of MYC, CCND1, and CDK4/6. We then constructed the network for MMPs and the 80 most frequently altered neighbor genes (**Figure 6C**). The results showed that collagen-related genes (for example, COL1A1) and metalloproteinase inhibitor-related genes (for example, TIMP2) were closely associated with MMP alterations. The functions of MMPs and the genes significantly associated with MMP alterations were predicted by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) in the Database for Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/summary.jsp).

GO enrichment analyses predicted the functional roles of target host genes on the basis of three aspects, including biological processes, cellular components, and molecular functions. For biological processes, the top 3 pathways were collagen catabolic

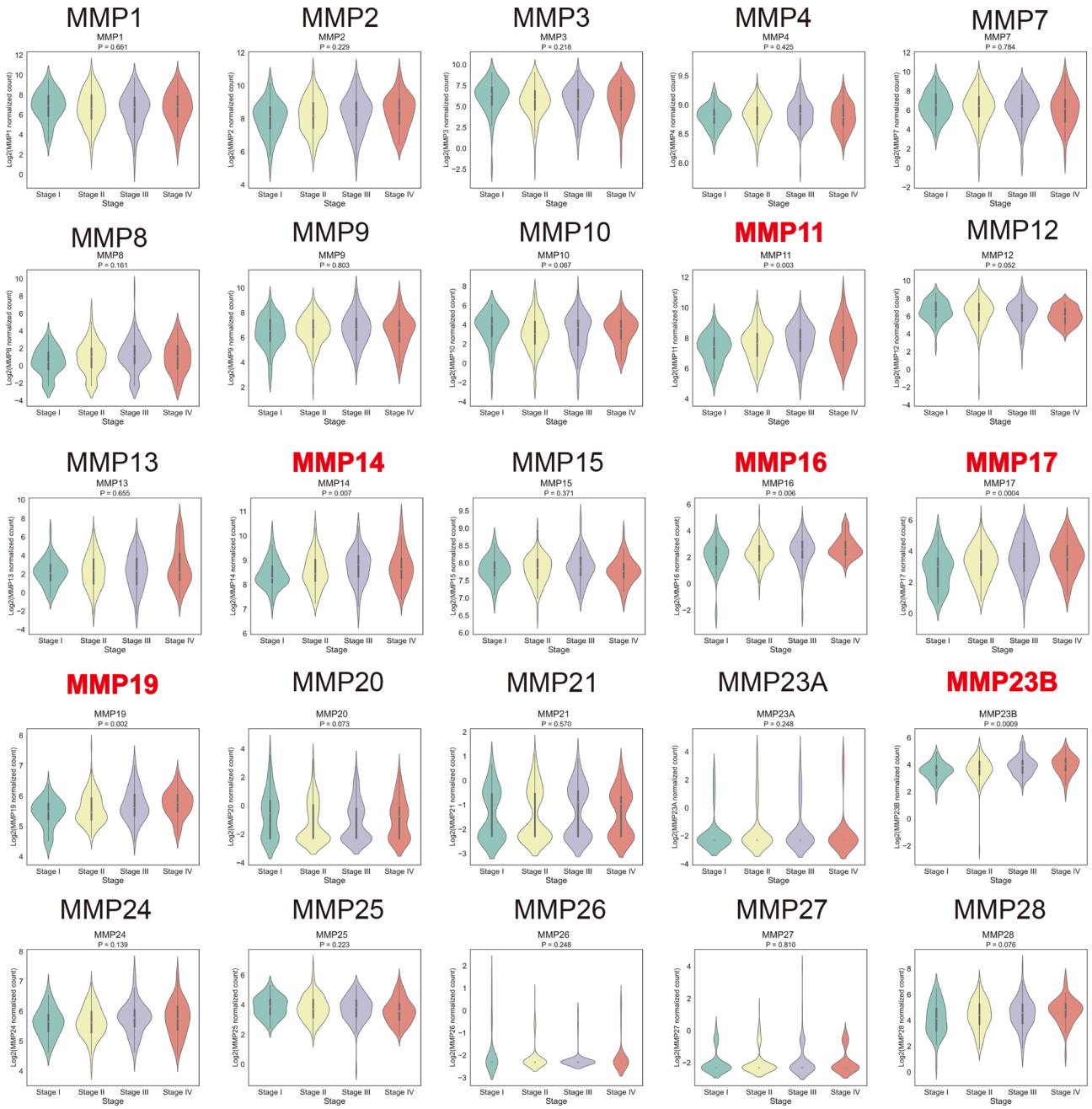


FIGURE 5 | Association of the mRNA levels of MMPs with tumor stage of patients with colorectal cancer. The titles of upregulated records are highlighted in red, while the titles of downregulated records are highlighted in blue.

process, extracellular matrix disassembly, and proteolysis, respectively. For cellular components, the top 3 pathways were extracellular matrix, proteinaceous extracellular matrix, and extracellular region, respectively, and for the molecular functions, the top 3 pathways were metalloendopeptidase activity, calcium ion binding, and serine-type endopeptidase activity, respectively (**Figure 6D**). In the KEGG enrichment analyses, the top 3 pathways were ECM-receptor interaction pathway, protein digestion and absorption pathway, and focal adhesion pathway, respectively (**Figure 6E**). Finally, by knocking down the expression of MMP11, MMP14, MMP17, and MMP19, we found that the expression of TIMP2 were significantly downregulated (**Figure 6F**). Similar trends were found for collagen-I (COL1A1) but not so obvious as TIMP2.

TABLE 1 | Association of the mRNA expression of MMPs with the prognosis of patients with colorectal cancer.

MMPs	TCGA								GEO	
	Overall survival (OS)		Disease-specific survival (DSS)		Progression-free survival (PFS)		Disease-free survival (DFS)		Relapse-free survival (RFS)	
	HR (95% CI)	p^a	HR (95% CI)	p^a	HR (95% CI)	p^a	HR (95% CI)	p^a	HR (95% CI)	p^a
MMP1	0.93 (0.83–1.05)	2.73E–01	0.93 (0.77–1.11)	4.15E–01	0.94 (0.84–1.05)	2.74E–01	**0.69 (0.51–0.92)**	**1.35E–02**	1.03 (0.98–1.09)	2.29E–01
MMP2	1.12 (0.92–1.36)	2.46E–01	1.24 (0.92–1.65)	1.53E–01	1.14 (0.95–1.37)	1.56E–01	0.86 (0.57–1.31)	4.90E–01	**1.17 (1.06–1.29)**	**1.38E–03**
MMP3	0.95 (0.85–1.06)	3.74E–01	0.96 (0.82–1.13)	6.44E–01	0.95 (0.85–1.05)	2.76E–01	**0.72 (0.56–0.93)**	**1.30E–02**	1 (0.94–1.06)	9.62E–01
MMP4	0.94 (0.35–2.49)	8.99E–01	0.87 (0.23–3.22)	8.30E–01	0.63 (0.27–1.49)	2.95E–01	1.07 (0.18–6.43)	9.39E–01	–	–
MMP7	1.11 (0.98–1.27)	9.83E–02	1.17 (0.97–1.41)	1.10E–01	1.06 (0.94–1.18)	3.63E–01	0.97 (0.75–1.25)	8.16E–01	1.09 (1–1.18)	5.36E–02
MMP8	1.13 (0.98–1.31)	1.04E–01	1.09 (0.89–1.33)	4.07E–01	1.12 (0.99–1.27)	7.52E–02	0.97 (0.71–1.33)	8.55E–01	**0.77 (0.64–0.92)**	**4.84E–03**
MMP9	1.05 (0.89–1.24)	5.66E–01	1.05 (0.82–1.33)	7.11E–01	0.97 (0.84–1.14)	7.42E–01	**0.58 (0.37–0.91)**	**1.71E–02**	1.03 (0.93–1.14)	5.35E–01
MMP10	0.91 (0.8–1.03)	1.43E–01	0.86 (0.72–1.04)	1.30E–01	0.93 (0.83–1.05)	2.21E–01	0.9 (0.69–1.17)	4.28E–01	1.02 (0.91–1.14)	7.44E–01
MMP11	**1.23 (1.01–1.49)**	**3.71E–02**	1.27 (0.96–1.69)	9.37E–02	**1.33 (1.11–1.59)**	**2.15E–03**	1.27 (0.83–1.94)	2.63E–01	**1.24 (1.09–1.42)**	**1.71E–03**
MMP12	0.96 (0.83–1.11)	5.72E–01	0.91 (0.74–1.13)	3.97E–01	0.94 (0.82–1.07)	3.59E–01	**0.67 (0.48–0.94)**	**2.08E–02**	1.05 (0.99–1.1)	9.10E–02
MMP13	1.04 (0.92–1.17)	5.72E–01	1.08 (0.91–1.28)	3.79E–01	1.03 (0.92–1.15)	6.25E–01	0.76 (0.57–1.01)	5.43E–02	**0.84 (0.73–0.97)**	**1.40E–02**
MMP14	1.36 (0.99–1.86)	5.43E–02	**1.73 (1.11–2.68)**	**1.44E–02**	**1.38 (1.04–1.82)**	**2.47E–02**	1.01 (0.54–1.9)	9.69E–01	**1.47 (1.15–1.89)**	**2.48E–03**
MMP15	0.89 (0.53–1.49)	6.53E–01	0.92 (0.44–1.9)	8.14E–01	1.04 (0.66–1.66)	8.56E–01	1.63 (0.59–4.55)	3.49E–01	0.99 (0.85–1.16)	8.95E–01
MMP16	**1.28 (1.02–1.62)**	**3.66E–02**	**1.49 (1.06–2.1)**	**2.08E–02**	**1.25 (1.01–1.56)**	**4.03E–02**	1 (0.59–1.69)	9.94E–01	**0.88 (0.81–0.95)**	**1.97E–03**
MMP17	**1.24 (1.04–1.49)**	**1.97E–02**	**1.41 (1.06–1.87)**	**1.84E–02**	**1.19 (1.01–1.41)**	**3.91E–02**	1.36 (0.9–2.06)	1.49E–01	**1.48 (1.14–1.92)**	**3.01E–03**
MMP19	**1.92 (1.28–2.88)**	**1.51E–03**	**1.9 (1.08–3.32)**	**2.51E–02**	**1.48 (1.03–2.13)**	**3.47E–02**	1.15 (0.56–2.38)	6.97E–01	**1.31 (1.16–1.49)**	**2.35E–05**
MMP20	1 (0.74–1.37)	9.83E–01	0.87 (0.55–1.38)	5.61E–01	0.94 (0.72–1.23)	6.47E–01	0.65 (0.34–1.23)	1.83E–01	**0.67 (0.51–0.9)**	**6.96E–03**
MMP21	1.44 (0.8–2.58)	2.25E–01	1.19 (0.45–3.12)	7.29E–01	0.9 (0.49–1.66)	7.47E–01	1.37 (0.4–4.71)	6.17E–01	0.54 (0.26–1.12)	9.95E–02
MMP23a	1.12 (0.75–1.69)	5.76E–01	0.86 (0.48–1.56)	6.23E–01	1.01 (0.69–1.48)	9.71E–01	1.59 (0.62–4.07)	3.29E–01	–	–
MMP23b	**1.72 (1.26–2.35)**	**6.96E–04**	**1.64 (1.03–2.59)**	**3.52E–02**	**1.4 (1.06–1.86)**	**1.91E–02**	1.12 (0.59–2.12)	7.24E–01	–	–
MMP24	1.11 (0.72–1.69)	6.43E–01	0.57 (0.31–1.05)	7.28E–02	0.85 (0.58–1.25)	4.19E–01	1.24 (0.46–3.34)	6.75E–01	**1.67 (1.24–2.25)**	**8.48E–04**
MMP25	0.87 (0.68–1.12)	2.91E–01	0.95 (0.66–1.39)	8.08E–01	0.82 (0.65–1.03)	8.39E–02	0.61 (0.34–1.09)	9.51E–02	1.25 (0.98–1.59)	6.65E–02
MMP26	1.4 (0.54–3.67)	4.91E–01	1176.9 (0–0)	1.00E+00	0.61 (0.17–2.14)	4.40E–01	0.02 (0–17787.09)	5.71E–01	1.25 (0.77–2.02)	3.73E–01
MMP27	0.65 (0.2–2.18)	4.89E–01	0.6 (0.08–4.71)	6.29E–01	1.19 (0.71–2)	5.00E–01	0.02 (0–42.44)	3.28E–01	**0.48 (0.3–0.77)**	**2.43E–03**
MMP28	0.97 (0.81–1.16)	7.65E–01	1.05 (0.8–1.37)	7.24E–01	1.09 (0.92–1.28)	3.30E–01	0.94 (0.65–1.37)	7.62E–01	**1.57 (1.19–2.07)**	**1.52E–03**

[a]*Age at diagnosis and sex were adjusted by using multivariate Cox regression. Records with a p<0.05 were bolded. MMP11, MMP14, MMP17 and MMP19 were bolded because they were positively associated with CRC prognosis both in the PFS and in the RFS analyses.*

DISCUSSION

MMPs were reported to be associated with the progression of colorectal cancer; however, a comprehensive bioinformatic analysis for all MMPs has yet to be performed. In this study, we systematically explored the mRNA expression level of all 24 MMPs and their prognosis value in colorectal cancer. We found that, the transcriptional level of MMP1, MMP3, MMP7, MMP9–MMP12, and MMP14 in tumor were significantly upregulated, both in public database and in our samples. Also, in the clinicopathological and prognosis analyses, upregulated MMP11, MMP14, MMP17, and MMP19 were significantly associated with a higher tumor stage and a worse prognosis.

In this study, five survival endpoints were used in the survival analyses. OS is an important endpoint and is easy to define (the patient is either alive or dead). However, using OS as an endpoint may weaken a clinical study as deaths because of noncancer causes that do not necessarily reflect tumor biology. DSS can overcome the shortage of OS as DSS only considers the people who have not died from a specific disease in a defined period of time. However, both OS and DSS demand longer follow-up times; thus, in many clinical trials, DFS or PFS are preferred. PFS is defined as the time to disease progression or death from any cause. Whereas, DFS is used to describe the period after a successful treatment during which there are no signs and symptoms of the disease that was treated. The above four endpoints of TCGA dataset were standardized by Liu et al. in 2018 (29). Another survival endpoint, RFS which was used by the GEO database (GSE39582), was defined as the time from surgery to the first relapse and was censored at 5 years (30).

MMP11 also named stromelysin-3 is a member of the stromelysin subgroup belonging to the MMP superfamily. In this study, MMP11 was significantly upregulated in tumor, both in public database (Oncomine and GEPIA) and in our samples. The protein levels of MMP11 in our 12 pair samples were upregulated in tumor tissue for patients 2, 3, 6, 9, and 10 but not for other patients. Also, in the clinicopathological and survival analyses, upregulated MMP11 was significantly associated with a higher tumor stage (p = 0.003), a shorter OS (HR = 1.23, $p = 3.71 \times 10^{-2}$), a shorter PFS time (HR = 1.33, $p = 2.15 \times 10^{-3}$), and a shorter RFS time (HR = 1.24, $p = 1.71 \times 10^{-3}$). In the DSS and DFS analyses, although the association did not reach a significant level, a similar trend was found (HR >1). In a previous study, Li et al. measured the serum levels of MMP11 in 92 colon cancer patients and 92 healthy individuals using ELISA. They found that the serum levels of MMP11 were substantially higher in colon cancer patients than in healthy controls and was an independent predictor of the OS and DFS of colon cancer (32). MMP11 also played an important role in the tumorigenesis, proliferation, and invasion process of

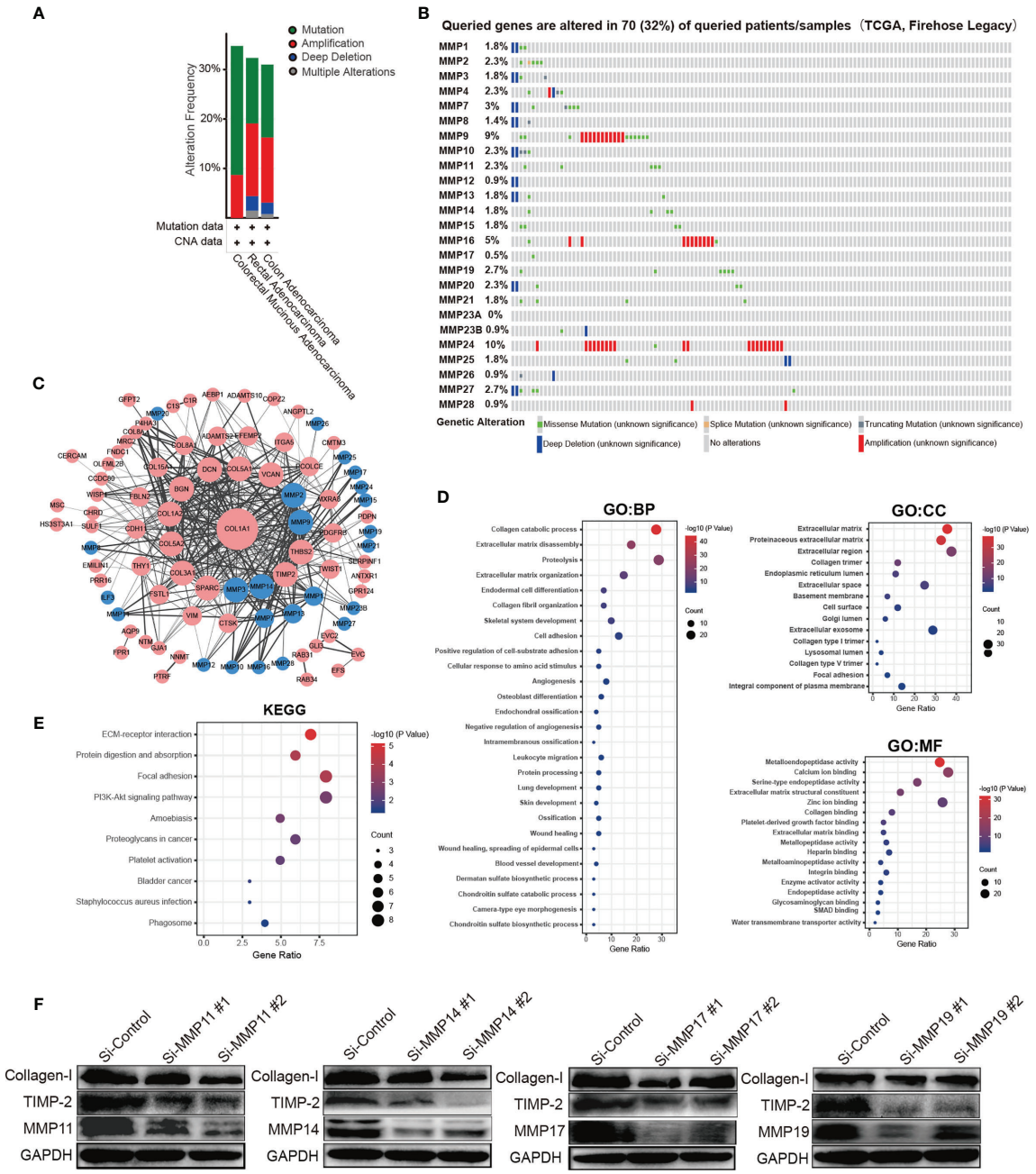


FIGURE 6 | Prediction function and pathways of the changes in MMPs and their frequently altered neighbor genes in patients with colorectal cancer. **(A)** Overview of mutation and copy number validation of MMPs in different types of colorectal cancer, **(B)** detailed alteration proportion and types of each MMP gene, **(C)** network analyses for MMPs and their 50 most frequently altered neighbor genes. **(D)** GO pathway analyses for MMPs and the genes significantly associated with MMP alterations. **(E)** KEGG pathway analyses for MMPs and the genes significantly associated with MMP alterations. **(F)** the expression levels of collagen-I (COL1A1) and TIMP2 after knocking down of MMP11, MMP14, MMP17, and MMP19 in HCT116 cell line by using siRNAs.

other cancers (33, 34). The mechanism behind it, may by inhibiting apoptosis as well as enhancing migration and invasion of cancer cells (35).

MMP14 plays an important role in extracellular matrix remodeling during aging. It has been reported to interact with TIMP2 (36). In our network analyses (**Figure 6C**), TIMP2 was indeed the closest gene of MMP14. Thus, by knocking down MMP14, the expression of TIMP2 was significantly downregulated (**Figure 6F**). In the transcriptional level, MMP14 was significantly upregulated in tumor tissue both in the public database and our own subjects. In the protein level, MMP14 was significantly upregulated in tumor tissue both in the

UALCAN database and in patients 2, 3, 7, 9, 10, 11, and 12 of our own subjects. Furthermore, in the clinicopathological and prognosis analyses, upregulated MMP14 was significantly associated with a higher tumor stage ($p = 0.007$), a shorter DSS survival time (HR = 1.73, $p = 0.01$), a shorter PFS time (HR = 1.38, $p = 0.02$), and a significantly shorter RFS time (HR = 1.47, $p = 2.48 \times 10^{-3}$). The association of MMP14 with the prognosis of colorectal cancer was also reported by Cui et al. in 2019. In addition, Cui et al. found that patient with upregulated MMP14 was significantly associated with a lower 5-year DFS and OS (37). Recently, Ragusa and coworkers found that upregulated MMP14 levels correlated with blood vessel dysfunction and a lack of cytotoxic T cells (38).

MMP17 and MMP19 were another two MMPs. In this study, we found that upregulated MMP17 and MMP19 were significantly associated with a higher tumor stage ($p = 4 \times 10^{-4}$ and $p = 2 \times 10^{-3}$ for MMP17 and MMP19, respectively), a shorter OS time (HR = 1.24, $p = 0.02$ for MMP17 and HR = 1.92, $p = 1.51 \times 10^{-3}$ for MMP19), a shorter DSS time (HR = 1.41, $p = 0.02$ for MMP17 and HR = 1.9, $p = 0.03$ for MMP19), a shorter PFS (HR = 1.19, $p = 0.04$ for MMP17 and HR = 1.48, $p = 0.03$ for MMP19), and a shorter RFS (HR = 1.48, $p = 3.01 \times 10^{-3}$ for MMP17 and HR = 1.31, $p = 2.35 \times 10^{-5}$ for MMP19). However, in the transcriptional analyses, MMP17 and MMP19 were significantly upregulated in Oncomine and our 12 samples but not in the GEPIA. In the protein analyses of our samples, MMP17 was upregulated in tumor tissue for patients 4, 6, and 10. Recently, by detecting MMP19 mRNA expression in 198 CRC cancer tissues and paired normal controls, Chen et al. found that MMP19 expression was significantly upregulated in cancer tissues than in normal controls. In addition, by using immunohistochemistry to detect the expression of MMP19 protein in 42 patients, they further found that MMP19 mRNA expression is highly correlated with their protein levels. In their prognosis analyses, significant association between upregulated MMP19 expression and worse prognosis was also found (39). The transcriptional level of MMP17 and MMP19 in colorectal cancer tissue and normal tissue may need to be further confirmed.

Recently, a Pan-cancer analysis for MMPs in TCGA was implemented by Emily et al. (40) Different from our study, they focus on the overall performance of MMPs in several cancers. In their study, they only used the colon cancer patients (COPD) of TCGA (without rectal cancer) and there is no validation dataset. In addition, the only survival endpoint used in their analyses was OS. As we described above, due to shorter follow-up time in TCGA, the accuracy of OS may not be as good as PFS. Finally, without adjusting the effect of age at diagnosis and sex, the log-rank test used in their study may bias the final result.

In summary, our study was among the first study to systematically evaluate the performance of MMPs in colorectal cancer. This study will deepen our understanding of the prognosis mechanism of colorectal cancer. Also, MMP11, MMP14, MMP17, and MMP19 are potential targets of precision therapy for patients with colorectal cancer.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Academic Committee of Sun Yat-Sen University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JY: acquisition of data, statistical analysis and technical interpretation of data, drafting of the manuscript, and critical revision of the manuscript for important intellectual content. ZH: acquisition of data and material, technical, and administrative support. XH, ZL, and LL: acquisition of data or material support. PL: study concept and design, acquisition of data, material support, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, and administrative support. HC: study concept and design, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, and administrative support. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Key R&D Program of China (PL, 2017YFC1308800); National Natural Science Foundation of China (PL, 81970452; ZH, 81902938); Science and Technology Program of Shenzhen, China (PL, JCYJ20190807161807867); Natural Science Foundation of Guangdong Province, China (ZH, 2020A1515011248); Fundamental Research Funds for the Central Universities (ZH, 19ykzd02); Guangdong Science and Technology Department (HC and BW, 2020B1212060018); the National Natural Science Foundation of China (HC, 81802833); and the 100 Top Talent Programs of Sun Yat-Sen University (HC, 58000-18841290).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.771099/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Hankey BF, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results Program: A National Resource. *Cancer Epidemiol Biomarkers Prev* (1999) 8(12):1117–21 https://cebp.aacrjournals.org/content/8/12/1117.long.
3. Lech G, Slotwinski R, Slodkowski M, Krasnodebski IW. Colorectal Cancer Tumour Markers and Biomarkers: Recent Therapeutic Advances. *World J Gastroenterol* (2016) 22(5):1745–55. doi: 10.3748/wjg.v22.i5.1745
4. Vandenbroucke RE, Libert C. Is There New Hope for Therapeutic Matrix Metalloproteinase Inhibition? *Nat Rev Drug Discov* (2014) 13(12):904–27. doi: 10.1038/nrd4390
5. Gialeli C, Theocharis AD, Karamanos NK. Roles of Matrix Metalloproteinases in Cancer Progression and Their Pharmacological Targeting. *FEBS J* (2011) 278(1):16–27. doi: 10.1111/j.1742-4658.2010.07919.x
6. Coussens LM, Fingleton B, Matrisian LM. Matrix Metalloproteinase Inhibitors and Cancer: Trials and Tribulations. *Science* (2002) 295 (5564):2387–92. doi: 10.1126/science.1067100
7. Bramhall SR, Schulz J, Nemunaitis J, Brown PD, Baillet M, Buckels JA. A Double-Blind Placebo-Controlled, Randomised Study Comparing Gemcitabine and Marimastat With Gemcitabine and Placebo as First Line Therapy in Patients With Advanced Pancreatic Cancer. *Br J Cancer* (2002) 87 (2):161–7. doi: 10.1038/sj.bjc.6600446
8. Sparano JA, Bernardo P, Stephenson P, Gradishar WJ, Ingle JN, Zucker S, et al. Randomized Phase III Trial of Marimastat Versus Placebo in Patients With Metastatic Breast Cancer Who Have Responding or Stable Disease After First-Line Chemotherapy: Eastern Cooperative Oncology Group Trial E2196. *J Clin Oncol* (2004) 22(23):4683–90. doi: 10.1200/JCO.2004.08.054
9. Hirte H, Vergote IB, Jeffrey JR, Grimshaw RN, Coppieters S, Schwartz B, et al. A Phase III Randomized Trial of BAY 12-9566 (Tanomastat) as Maintenance Therapy in Patients With Advanced Ovarian Cancer Responsive to Primary Surgery and Paclitaxel/Platinum Containing Chemotherapy: A National Cancer Institute of Canada Clinical Trials Group Study. *Gynecol Oncol* (2006) 102(2):300–8. doi: 10.1016/j.ygyno.2005.12.020
10. Winer A, Adams S, Mignatti P. Matrix Metalloproteinase Inhibitors in Cancer Therapy: Turning Past Failures Into Future Successes. *Mol Cancer Ther* (2018) 17(6):1147–55. doi: 10.1158/1535-7163.MCT-17-0646
11. Shah MA, Starodub A, Sharma S, Berlin J, Patel M, Wainberg ZA, et al. Andecaliximab/GS-5745 Alone and Combined With Mfolfox6 in Advanced Gastric and Gastroesophageal Junction Adenocarcinoma: Results From a Phase I Study. *Clin Cancer Res* (2018) 24(16):3829–37. doi: 10.1158/1078-0432.CCR-17-2469
12. Shiozawa J, Ito M, Nakayama T, Nakashima M, Kohno S, Sekine I. Expression of Matrix Metalloproteinase-1 in Human Colorectal Carcinoma. *Mod Pathol* (2000) 13(9):925–33. doi: 10.1038/modpathol.3880169
13. Sunami E, Tsuno N, Osada T, Saito S, Kitayama J, Tomozawa S, et al. MMP-1 Is a Prognostic Marker for Hematogenous Metastasis of Colorectal Cancer. *Oncologist* (2000) 5(2):108–14. doi: 10.1634/theoncologist.5-2-108
14. Bendardaf R, Buhmeida A, Ristamaki R, Syrjanen K, Pyrhonen S. MMP-1 (Collagenase-1) Expression in Primary Colorectal Cancer and Its Metastases. *Scand J Gastroenterol* (2007) 42(12):1473–8. doi: 10.1080/00365520701485449
15. Langenskiold M, Holmdahl L, Falk P, Ivarsson ML. Increased Plasma MMP-2 Protein Expression in Lymph Node-Positive Patients With Colorectal Cancer. *Int J Colorectal Dis* (2005) 20(3):245–52. doi: 10.1007/s00384-004-0667-4
16. Adachi Y, Yamamoto H, Itoh F, Hinoda Y, Okada Y, Imai K. Contribution of Matrilysin (MMP-7) to the Metastatic Pathway of Human Colorectal Cancers. *Gut* (1999) 45(2):252–8. doi: 10.1136/gut.45.2.252
17. Ii M, Yamamoto H, Adachi Y, Maruyama Y, Shinomura Y. Role of Matrix Metalloproteinase-7 (Matrilysin) in Human Cancer Invasion, Apoptosis, Growth, and Angiogenesis. *Exp Biol Med (Maywood)* (2006) 231(1):20–7. doi: 10.1177/153537020623100103
18. Loesch M, Zhi HY, Hou SW, Qi XM, Li RS, Basir Z, et al. P38gamma MAPK Cooperates With C-Jun in Trans-Activating Matrix Metalloproteinase 9. *J Biol Chem* (2010) 285(20):15149–58. doi: 10.1074/jbc.M110.105429
19. Zhang B, Halder SK, Kashikar ND, Cho YJ, Datta A, Gorden DL, et al. Antimetastatic Role of Smad4 Signaling in Colorectal Cancer. *Gastroenterology* (2010) 138(3):969–80.e1-3. doi: 10.1053/j.gastro.2009.11.004
20. Papageorgis P, Cheng K, Ozturk S, Gong Y, Lambert AW, Abdolmaleky HM, et al. Smad4 Inactivation Promotes Malignancy and Drug Resistance of Colon Cancer. *Cancer Res* (2011) 71(3):998–1008. doi: 10.1158/0008-5472.CAN-09-3269
21. Garg P, Sarma D, Jeppsson S, Patel NR, Gewirtz AT, Merlin D, et al. Matrix Metalloproteinase-9 Functions as a Tumor Suppressor in Colitis-Associated Cancer. *Cancer Res* (2010) 70(2):792–801. doi: 10.1158/0008-5472.CAN-09-3166
22. Yang W, Arii S, Gorrin-Rivas MJ, Mori A, Onodera H, Imamura M. Human Macrophage Metalloelastase Gene Expression in Colorectal Carcinoma and Its Clinicopathologic Significance. *Cancer* (2001) 91(7):1277–83. doi: 10.1002/1097-0142(20010401)91:7<1277::AID-CNCR1129>3.0.CO;2-H
23. Huang MY, Chang HJ, Chung FY, Yang MJ, Yang YH, Wang JY, et al. MMP13 Is a Potential Prognostic Marker for Colorectal Cancer. *Oncol Rep* (2010) 24(5):1241–7. doi: 10.3892/or_00000978
24. Porte H, Chastre E, Prevot S, Nordlinger B, Empereur S, Basset P, et al. Neoplastic Progression of Human Colorectal Cancer Is Associated With Overexpression of the Stromelysin-3 and BM-40/SPARC Genes. *Int J Cancer* (1995) 64(1):70–5. doi: 10.1002/ijc.2910640114
25. Baker EA, Bergin FG, Leaper DJ. Matrix Metalloproteinases, Their Tissue Inhibitors and Colorectal Cancer Staging. *Br J Surg* (2000) 87(9):1215–21. doi: 10.1046/j.1365-2168.2000.01531.x
26. Kikuchi R, Noguchi T, Takeno S, Kubo N, Uchida Y. Immunohistochemical Detection of Membrane-Type-1-Matrix Metalloproteinase in Colorectal Carcinoma. *Br J Cancer* (2000) 83(2):215–8. doi: 10.1054/bjoc.2000.1195
27. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res* (2017) 45(W1):W98–W102. doi: 10.1093/nar/gkx247
28. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* (2017) 19 (8):649–58. doi: 10.1016/j.neo.2017.05.002
29. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* (2018) 173(2):400–16.e11. doi: 10.1016/j.cell.2018.02.052
30. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene Expression Classification of Colon Cancer Into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PloS Med* (2013) 10(5): e1001453. doi: 10.1371/journal.pmed.1001453
31. Cancer Genome Atlas N. Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature* (2012) 487(7407):330–7. doi: 10.1038/nature11252
32. Pang L, Wang DW, Zhang N, Xu DH, Meng XW. Elevated Serum Levels of MMP-11 Correlate With Poor Prognosis in Colon Cancer Patients. *Cancer Biomark* (2016) 16(4):599–607. doi: 10.3233/CBM-160601
33. Deng H, Guo RF, Li WM, Zhao M, Lu YY. Matrix Metalloproteinase 11 Depletion Inhibits Cell Proliferation in Gastric Cancer Cells. *Biochem Biophys Res Commun* (2005) 326(2):274–81. doi: 10.1016/j.bbrc.2004.11.027
34. Chen C, Liu X, Jiang J, Li S, Wang G, Ju L, et al. Matrix Metalloproteinase 11 Is a Potential Biomarker in Bladder Cancer Diagnosis and Prognosis. *Onco Targets Ther* (2020) 13:9059–69. doi: 10.2147/OTT.S243452
35. Zhang X, Huang S, Guo J, Zhou L, You L, Zhang T, et al. Insights Into the Distinct Roles of MMP-11 in Tumor Biology and Future Therapeutics (Review). *Int J Oncol* (2016) 48(5):1783–93. doi: 10.3892/ijo.2016.3400
36. Zucker S, Drews M, Conner C, Foda HD, DeClerck YA, Langley KE, et al. Tissue Inhibitor of Metalloproteinase-2 (TIMP-2) Binds to the Catalytic Domain of the Cell Surface Receptor, Membrane Type 1-Matrix Metalloproteinase 1 (MT1-MMP). *J Biol Chem* (1998) 273(2):1216–22. doi: 10.1074/jbc.273.2.1216
37. Cui G, Cai F, Ding Z, Gao L. MMP14 Predicts a Poor Prognosis in Patients With Colorectal Cancer. *Hum Pathol* (2019) 83:36–42. doi: 10.1016/j.humpath.2018.03.030
38. Ragusa S, Prat-Luri B, Gonzalez-Loyola A, Nassiri S, Squadrito ML, Guichard A, et al. Antiangiogenic Immunotherapy Suppresses Desmoplastic and

Chemoresistant Intestinal Tumors in Mice. *J Clin Invest* (2020) 130(3):1199–216. doi: 10.1172/JCI129558
39. Chen Z, Wu G, Ye F, Chen G, Fan Q, Dong H, et al. High Expression of MMP19 Is Associated With Poor Prognosis in Patients With Colorectal Cancer. *BMC Cancer* (2019) 19(1):448. doi: 10.1186/s12885-019-5673-6
40. Gobin E, Bagwell K, Wagner J, Mysona D, Sandirasegarane S, Smith N, et al. A Pan-Cancer Perspective of Matrix Metalloproteases (MMP) Gene Expression Profile and Their Diagnostic/Prognostic Potential. *BMC Cancer* (2019) 19 (1):581. doi: 10.1186/s12885-019-5768-0

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

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

frontiers in Oncology

SYSTEMATIC REVIEW
published: 18 November 2021
doi: 10.3389/fonc.2021.754162

The Prognosis of Leptin rs2167270 G > A (G19A) Polymorphism in the Risk of Cancer: A Meta-Analysis

Aiqiao Zhang[1,2†], *Shangren Wang*[3†], *Fujun Zhang*[1,2], *Wei Li*[1,2], *Qian Li*[1,2] *and Xiaoqiang Liu*[3*]

[1] Department of Neonatology, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, China, [2] Department of Neonatology, National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin, China, [3] Department of Urology, Tianjin Medical University General Hospital, Tianjin, China

Background: Although the effect of the LEP G19A (rs2167270) polymorphism on cancers is assumed, the results of its influence have been contradictory. A meta-analysis was conducted to precisely verify the relationships between LEP G19A and the development of digestion-related cancers.

Methods: Investigators systematically searched the literature in PubMed, Embase, and Web of Science and used STATA software 14.0 for the meta–analysis. The odds ratios (ORs) and 95% confidence intervals (CIs) were calculated to evaluate the associations. Subgroup analyses stratified by ethnicity, cancer type, and cancer system were further conducted to assess the relationship between the LEP G19A polymorphism and digestion-related cancers.

Results: In the overall population, we found a significant relationship with overall cancer (allele comparison: OR = 0.921, $p = 0.000$; dominant comparison: OR = 0.923, $p = 0.004$; recessive comparison: OR = 0.842, $p = 0.000$; homozygote model: OR = 0.0843, $p = 0.001$). In a subgroup analysis conducted by ethnicity, we obtained significant results in Asians (Asian allele comparison: OR = 0.885, $p = 0.000$; dominant comparison: OR = 0.862, $p = 0.000$; homozygote model: OR = 0.824, $p = 0.039$; and heterozygote comparison: OR = 0.868, $p = 0.000$) but not in Caucasians. In a subgroup analysis conducted by cancer type and cancer system, we obtained significant results that the LEP G19A polymorphism may decrease the risk of colorectal cancer, esophageal cancer, digestive system cancer, and urinary system cancer.

Conclusions: This meta-analysis revealed that the LEP G19A polymorphism may decrease the risk of cancer.

Keywords: leptin (LEP), cancer, polymorphism, A19G, rs2167270

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Manuel Pires Bicho, University of Lisbon, Portugal
Rajeev K. Singla, Sichuan University, China

***Correspondence:**
Xiaoqiang Liu
1291313638@qq.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *06 August 2021*
Accepted: *29 October 2021*
Published: *18 November 2021*

Citation:
Zhang A, Wang S, Zhang F, Li W, Li Q and Liu X (2021) The Prognosis of Leptin rs2167270 G > A (G19A) Polymorphism in the Risk of Cancer: A Meta-Analysis. Front. Oncol. 11:754162. doi: 10.3389/fonc.2021.754162

INTRODUCTION

It is well known that cancer is one of major causes of death with over 6.1 million projected to die each year, and morbidity rates have increased gradually over the past decade (1, 2), so it has been a public health burden worldwide. The reason for cancer is complicated and the etiology and mechanism of carcinogenesis are not clearly elucidated to date. It was widely accepted that the interplay between

environmental factors, genetics, and lifestyle plays an important role in the carcinogenesis according to epidemiology. There is mounting evidence indicating that many metabolic diseases such as obesity and diabetes may significantly increase the risk of cancer (3–5). The polymorphism of obesity and diabetes gene may be associated with genetic susceptibility of cancer.

Leptin (LEP), a 16-kDa hormone of energy expenditure, is a balancing mediator of homeostasis by regulating acquisition and consumption of energy, which was a basic pathophysiological process in normal cells and cancer cells. Many epidemiological studies have revealed the link between LEP and the development of many kinds of cancers (6–8). Among the pathophysiological mechanisms of cancer, LEP seems relevant to the proliferation of cancer stem cells (9). Some studies also revealed that LEP through its signal pathways regulating energy intake and expenditure [MAPK, PI3K, mTOR, and JAK/STAT (10, 11)] produced an effect in angiogenesis processes that were critical in the genesis and development of cancer (12). Pathophysiological mechanisms of cancer such as inflammation, invasion, and metastasis are also favored by LEP (13–15). So, LEP may be involved in various pathological processes of carcinogenesis.

Single-nucleotide polymorphism can change the functions of genes and the expression of protein. LEP G19A polymorphism, positioning at the 5′-untranslated region of gene, may impact mRNA translation and change the serum level of LEP. With the development of molecular epidemiology, various studies have demonstrated that LEP G19A polymorphism is related to cancer risk (16–19). However, results between G19A polymorphism with cancers have been inconsistent or inconclusive. Therefore, we performed a metaanalysis to verify the correlation between the G19A mutation of the LEP gene and susceptibility to cancers.

In this study, we conducted a meta-analysis to verify whether the G19A polymorphism of the LEP gene affects the risk of cancer.

METHODS

Literature Search

A comprehensive literature search of PubMed, Embase, and Web of Science was performed to search all potential studies that involved the relevance between the G19A polymorphism and cancers prior to June 2021. Our study contained the following terms: ("leptin" OR "LEP" OR "G19A" OR "rs2167270") AND ("polymorphism" OR "variant" OR "mutation") AND ("malignancy" OR "cancer" OR "carcinoma" OR "neoplasm").

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) investigate the association between the LEP G19A (rs2167270) mutation and cancers; (2) meet cohort design or case–control design; (3) abundant data should behave to estimate an odds ratio (OR) and 95% confidence interval; (4) results were reported in English; and (5) include human subjects. We adopted the following exclusion criteria: (1) duplicated studies; (2) studies in which subjects were not human; and (3) studies in which we could not obtain sufficient raw data.

Data Extraction

Investigators extracted genotype data independently, and every data point reached a consensus. The extracted data contained the (1) name of the first author; (2) year of publication; (3) ethnicity of cases and controls; (4) cancer type of studies; and (5) frequency of LEP G19A in genes.

Statistical Analysis

We computed ORs and their 95% CIs to estimate the association between the LEP G19A (rs2167270) mutation and cancers. The pooled ORs and their 95% CIs were computed for genes using the following five models: dominant model (AA + AG vs. GG), recessive model (AA vs. GG + AG), allele model (A vs. G), homozygous model (AA vs. GG), and heterozygote model (AG vs. GG).

The Q test was used to estimate heterogeneity between different studies, and $p < 0.05$ was considered significant for heterogeneity. In addition, inconsistency was quantified by the I^2 statistic. Twenty-five percent and 50% of the I^2 values indicated low and high levels of heterogeneity, respectively. An $I^2 < 50\%$ suggested that no heterogeneity existed. When heterogeneity existed, the fixed effects model (FEM) was utilized; otherwise, the random-effects model (REM) was utilized for calculation.

To evaluate the specific effects of ethnicity, cancer type, and cancer system, investigators performed subgroup analyses by ethnicity, cancer type, and cancer system.

Sensitivity analyses were performed to evaluate the stability of the results. A funnel plot of Egger's or Begg's test was conducted to reveal possible publication bias. We used the Newcastle-Ottawa Scale to assess the including literature quality. All meta-analyses were performed using STATA software (Version 12.0, College Station, TX).

RESULTS

Study Characteristics

Depending on the search strategy, 633 articles were retrieved (**Figure 1**). Among them, 103 articles were eligible after excluding repeated publications. By reviewing the titles and study abstracts, 58 articles were excluded. Of the remaining 45 studies, 26 articles were excluded, including 11 studies that were not focused on the LEP G19A genetic mutation. Five studies were meta-analyses. Eight studies were on other disorders that were not cancer. Two articles did not provide raw data. Finally, 19 studies conformed to our meta-analyses, and **Tables 1** and **2** summarize the extracted data (16–34).

Effect of the LEP G19A Polymorphism on Cancers

We investigated the effect of the LEP G19A mutation on cancer susceptibility in five genetic models. In all models, if the heterogeneity was less than 50%, the authors applied fixed models, whereas if the heterogeneity was greater than 50%, random models were used.

In the overall population, we found a significant relationship with cancer in four models (allele comparison: OR = 0.921, p =

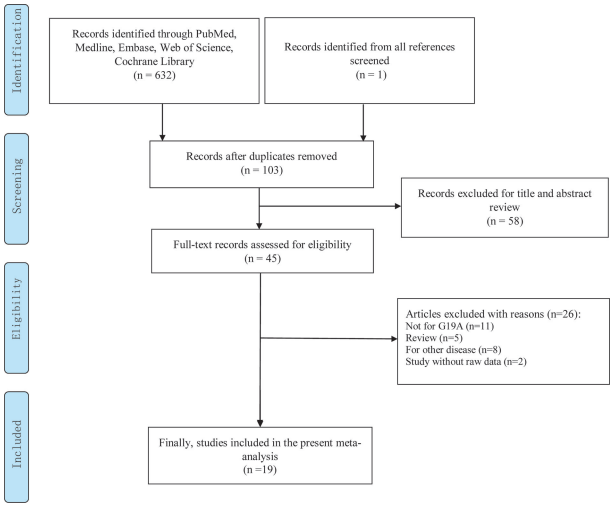


FIGURE 1 | Flowchart of literature selection.

0.000; dominant comparison: OR = 0.923, p = 0.004; recessive comparison: OR = 0.842, p = 0.000; homozygote model: OR = 0.0843, p = 0.001), and no relevance was observed in the heterozygote model (OR = 0.944, p = 0.05) (**Table 3** and **Figure 2**).

In a subgroup analysis conducted by ethnicity, we obtained significant results in Asians in four models (allele comparison: OR = 0.885, p = 0.000; dominant comparison: OR = 0.862, p = 0.000; homozygote model: OR = 0.824, p = 0.039; and heterozygote comparison: OR = 0.868, p = 0.000); we also obtained significant results in the mixed recessive model: OR = 0.785, p = 0.1006. We obtained no significant results in the Caucasian population in five models (**Table 3** and **Figure 2**).

In a subgroup analysis conducted by cancer type, we obtained significant results that the LEP G19A polymorphism decreased the risk of colorectal cancer in one model (recessive model: OR = 0.816, p = 0.010); decreased the risk of esophageal cancer in two models (allele model: OR = 0.888, p = 0.014; dominant comparison: OR = 0.874, p = 0.022); and decreased the risk of other types of cancer in three models (allele comparison: OR = 0.866, p = 0.010; dominant comparison: OR = 0.842, p = 0.010; heterozygote comparison: OR = 0.849, p = 0.018) (**Table 3** and **Figure 2**).

In a subgroup analysis conducted by cancer system, we obtained significant results that the LEP G19A polymorphism decreased the risk of digestive system cancer in three models (allele comparison: OR = 0.937, p = 0.016; recessive comparison: OR = 0.838, p = 0.005; homozygote comparison: OR = 0.863, p = 0.028); we also obtained significant results that the LEP G19A polymorphism decreased the risk of urinary system cancer in three models (allele comparison: OR = 0.881, p = 0.022; dominant comparison: OR = 0.842, p = 0.019; heterozygote comparison: OR = 0.855, p = 0.043) (**Table 3** and **Figure 2**).

Sensitivity Analysis and Publication Bias

We used Begg's and Egger's tests to evaluate publication bias in all models. All results of Begg's and Egger's tests were >0.05 in all models and funnel plots, revealing that publication bias may not exist among our studies (**Table 2**, **Figures 3** and **4**). We conducted a sensitivity analysis, and pooled ORs and the corresponding 95% CIs were computed. The results did not show a significant change even though one study was deleted each time, which suggested that the results were statistically stable (**Figure 5**).

DISCUSSION

It has been confirmed that the occurrence of cancer is a complex, multistep, and multifactorial event that contains various genetic,

TABLE 1 | The characters of included studies in the meta-analysis.

Author	Year	Country	Ethnicity	Cancer type	Cancer system	Genotype	
						Case	Control
Skibola et al. (16)	2004	USA	Caucasians	Non-Hodgkin lymphoma	Hematopoietic malignancy	373	805
Willett et al. (17)	2005	UK	Caucasians	Non-Hodgkin lymphoma	Hematopoietic malignancy	590	754
Slattery et al. (18)	2008	USA	Mixed	Colorectal cancer	Digestive system cancer	1,567	1,965
Doecke et al. (20)	2008	Australia	Caucasians	Esophageal cancer	Digestive system cancer	774	1,352
Tsilidis et al. (19)	2009	USA	Mixed	Colorectal cancer	Digestive system cancer	204	362
Wang et al. (21)	2009	USA	Caucasians	Prostate cancer	Urinary system cancer	253	257
Moore et al. (22)	2009	Finland	Caucasians	Prostate cancer	Urinary system cancer	945	840
Partida-Perez et al. (23)	2010	Mexico	Caucasians	Colorectal cancer	Digestive system cancer	68	102
Kim et al. (24)	2012	Korea	Asian	Breast cancer	Others	400	452
Zhang et al. (25)	2012	China	Asian	Non-Hodgkin lymphoma	Hematopoietic malignancy	514	557
Qiu et al. (26)	2017	China	Asian	Esophageal cancer	Digestive system cancer	502	1,496
Zhang et al. (27)	2018	China	Asian	Hepatocellular carcinoma	Digestive system cancer	584	923
Huang et al. (28)	2018	USA	Mixed	Colorectal cancer	Digestive system cancer	134	259
Yang et al. (29)	2019	China	Asian	Esophageal cancer	Digestive system cancer	1,063	1,677
Lin et al. (30)	2020	China	Asian	Colorectal cancer	Digestive system cancer	1,003	1,303
Ma (31)	2020	China	Asian	Gastric cancer	Digestive system cancer	379	463
Al-Khatib et al. (32)	2020	Jordan	Caucasians	Large B-Cell lymphoma	Hematopoietic malignancy	118	228
Mao et al. (33)	2020	China	Asian	Bladder cancer	Urinary system cancer	353	433
Mhaidat et al. (34)	2021	Jordan	Caucasians	Colorectal cancer	Digestive system cancer	54	23

TABLE 2 | Distribution of LEP G19A polymorphism genotype and allele.

Author	Year	Genotype distribution										HWE
		Case					Control					
		AA	AG	GG	A	G	AA	AG	GG	A	G	
Skibola et al. (16)	2004	36	169	168	241	505	119	335	351	573	1,037	0.009
Willett et al. (17)	2005	79	276	235	434	746	122	357	275	601	907	0.734
Slattery et al. (18)	2008	190	766	611	1,146	1,988	304	867	794	1,475	2,455	0.009
Doecke et al. (20)	2008	34	130	94	198	318	176	622	541	974	1,704	0.633
Tsilidis et al. (19)	2009	33	91	80	157	251	61	170	131	292	432	0.940
Wang et al. (21)	2009	39	122	92	200	306	38	119	100	195	319	0.789
Moore et al. (22)	2009	113	404	428	630	1,260	107	387	346	601	1,079	0.644
Partida-Perez et al. (23)	2010	7	44	17	58	78	25	53	24	103	101	0.691
Kim et al. (24)	2012	12	110	269	134	648	18	147	284	183	715	0.851
Zhang et al. (25)	2012	26	166	322	218	810	29	190	338	248	866	0.733
Qiu et al. (26)	2017	19	165	318	203	801	67	528	894	662	2,316	0.764
Zhang et al. (27)	2018	34	198	343	266	884	36	321	564	393	1,449	0.448
Huang et al. (28)	2018	13	71	50	97	171	29	119	111	177	341	0.089
Yang et al. (29)	2019	29	334	678	392	1,690	73	603	998	749	2,599	0.109
Lin et al. (30)	2020	51	340	589	442	1,518	59	474	767	592	2,008	0.832
Ma (31)	2020	14	120	245	148	610	30	170	263	230	696	0.883
Al-Khatib et al. (32)	2020	19	50	49	88	148	19	102	107	140	316	0.307
Mao et al. (33)	2020	11	114	228	136	570	29	162	242	220	646	0.473
Mhaidat et al. (34)	2021	10	23	21	43	65	4	11	8	19	27	0.414

HWE, Hardy–Weinberg equilibrium.

TABLE 3 | The association between LEP G19A and cancer susceptibility.

G19A	No	A vs. G			AA+AG vs. GG			AA vs. AG+GG			AA vs. GG			AG vs. GG		
		p^a	OR (95% CI)	I^2	p^a	OR (95% CI)	I^2	p^a	OR (95% CI)	I^2	p^a	OR(95% CI)	I^2	p^a	OR (95% CI)	I^2
Overall	19	**0.000**	**0.921 (0.883–0.961)**	43.60%	**0.004**	**0.923 (0.874–0.975)**	33.70%	**0.000**	**0.842 (0.765–0.927)**	44.10%	**0.001**	**0.843 (0.762–0.933)**	43.30%	0.050	0.944 (0.890–1.000)	29.30%
Begg's test[b]	19		0.649			0.600			0.972			0.916			0.382	
Egger's test[c]	19		0.963			0.802			0.460			0.587			0.907	
Ethnicity																
Caucasians	8	0.106	0.941 (0.873–1.013)	36.00%	0.337	0.951 (0.858–1.054)	0.00%	0.066	0.869 (0.749–1.009)	53.30%	0.078	0.866 (0.737–1.016)	46.70%	0.674	0.977 (0.876–1.089)	0.00%
Mixed	3	0.423	0.965 (0.885–1.053)	0.00%	0.411	1.052 (0.932–1.188)	0.00%	**0.006**	**0.785 (0.660–0.934)**	0.00%	0.056	0.833 (0.690–1.004)	0.00%	0.069	1.126 (0.991–1.280)	12.20%
Asians	8	**0.000**	0.885 (0.830–0.944)	59.80%	**0.000**	**0.862 (0.799–0.931)**	36.00%	0.119	0.866 (0.722–1.038)	54.10%	**0.039**	**0.824 (0.686–0.990)**	60.40%	**0.000**	**0.868 (0.802–0.939)**	0.00%
Cancer type																
NHL	4	0.103	0.921 (0.835–1.017)	48.70%	0.299	0.932 (0.817–1.062)	0.00%	0.082	0.832 (0.676–1.024)	69.40%	0.079	0.82 (0.658–1.023)	66.90%	0.579	0.962 (0.837–1.104)	0.00%
CRC	6	0.312	0.963 (0.896–1.036)	0.00%	0.771	1.015 (0.921–1.118)	0.00%	**0.010**	**0.816 (0.700–0.952)**	35.00%	0.081	0.863 (0.732–1.018)	0.00%	0.306	1.055 (0.952–1.168)	8.20%
EC	3	**0.014**	**0.888 (0.808–0.976)**	68.00%	**0.022**	**0.874 (0.779–0.980)**	67.30%	0.113	0.813 (0.630–1.050)	17.80%	0.100	0.801 (0.615–1.043)	52.50%	**0.060**	**0.892 (0.791–1.005)**	61.90%
PC	2	0.275	0.935 (0.828–1.055)	29.70%	0.204	0.898 (0.760–1.060)	43.90%	0.740	0.959 (0.752–1.225)	0.00%	0.485	0.911 (0.702–1.183)	0.00%	0.221	0.896 (0.752–1.068)	38.20%
Others	4	**0.010**	**0.866 (0.777–0.966)**	76.40%	**0.010**	**0.842 (0.739–0.959)**	62.50%	0.247	0.837 (0.619–1.132)	72.40%	0.132	0.791 (0.583–1.073)	76.40%	**0.018**	**0.849 (0.742–0.973)**	24.40%
System of cancer																
Hematopoietic malignancy	4	0.103	0.921 (0.835–1.017)	48.70%	0.290	0.932 (0.817–1.062)	0.00%	0.082	0.832 (0.676–1.024)	69.40%	0.079	0.82 (0.658–1.023)	66.90%	0.579	0.962 (0.837–1.104)	0.00%
Digestive system	11	**0.016**	**0.937 (0.889–0.988)**	44.90%	0.127	0.948 (0.886–1.015)	43.20%	**0.005**	**0.838 (0.740–0.949)**	44.50%	**0.028**	**0.863 (0.757–0.984)**	41.40%	0.407	0.907 (0.904–1.042)	44.70%
Urinary system	3	**0.022**	**0.881 (0.791–0.982)**	65.50%	**0.019**	**0.842 (0.729–0.973)**	50.70%	0.262	0.877 (0.698–1.103)	51.80%	0.109	0.82 (0.643–1.045)	61.20%	**0.043**	**0.855 (0.735–0.995)**	25.00%
Others	1	0.090	0.808 (0.631–1.034)	–	0.091	0.781 (0.586–1.040)	–	0.465	0.758 (0.361–1.594)	–	0.358	0.704 (0.333–1.489)	–	0.121	0.79 (0.586–1.064)	–

NO, number of study; NHL, non-Hodgkin lymphoma; CRC, colorectal cancer; EC, esophageal cancer; HWE, Hardy–Weinberg equilibrium; _, no available.

The meaning of bold values is statistically significant (P<0.05).

[a]P value of Q test for heterogeneity test; [b]P value of Begg rank for testing publication bias; [c]P value of Egger rank for testing publication bias.

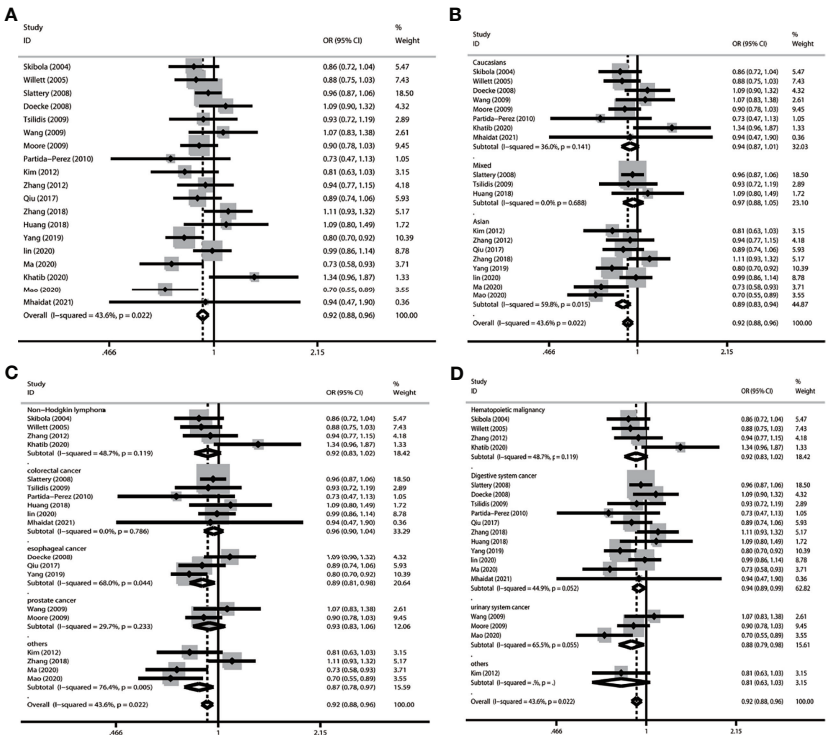


FIGURE 2 | Forest plot of subgroup analysis of LEP G19A and cancer risk in the allele model (A vs. G) **(A)** LEP G19A polymorphism and overall cancer risk; **(B)** LEP G19A polymorphism and cancer risk on ethnicity; **(C)** LEP G19A polymorphism and risk of cancer type; **(D)** LEP G19A polymorphism and the risk of cancer system).

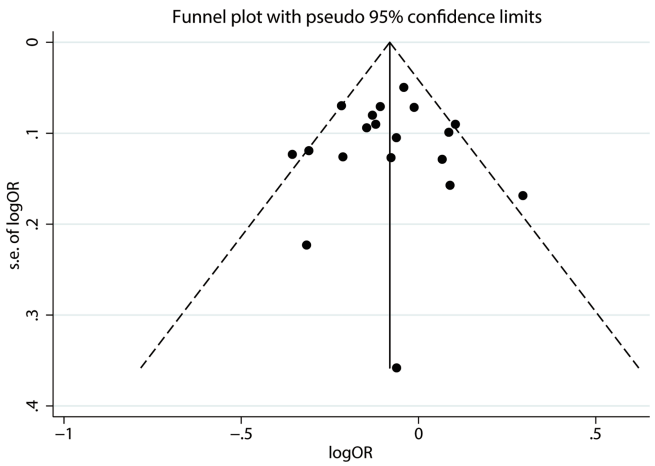


FIGURE 3 | Funnel plot of publication bias on the relationship between LEP G19A polymorphism and the risk of digestion-related cancer in allele model (A vs. G).

environmental, and lifestyle factors, such as smoking, drinking, obesity, and genetic factor. Multiple studies have revealed that metabolic-related factors are associated with the risk of cancer (35–37). The LEP, metabolic-related factors regulating balancing by regulating acquisition and energy consumption, was confirmed relevant to cancer (38–40). The LEP G19A polymorphism may alter the transcription of mRNA and the level of LEP was confirmed to be associated with any kind of cancer (21, 22, 24–26). However, the conclusions of those studies were inconsistent. Two meta-analyses were researched by Liu et al. (41), including 10 studies, and Yang et al. (29), including 13 studies, generating conflicting results in subgroup analysis and

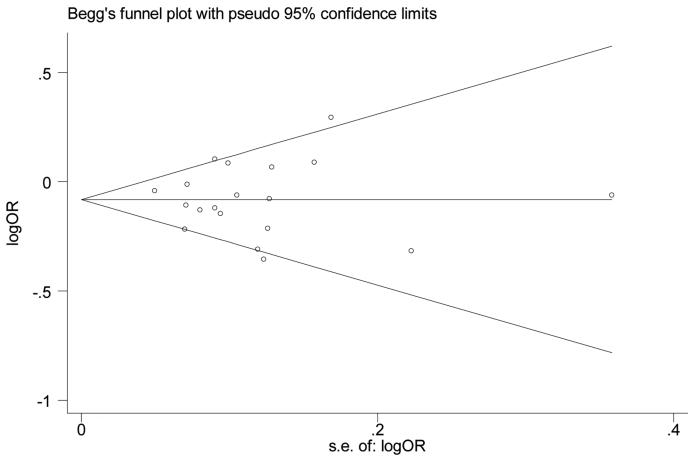


FIGURE 4 | Begg's funnel plot of meta–analysis in the allele model (A vs. G).

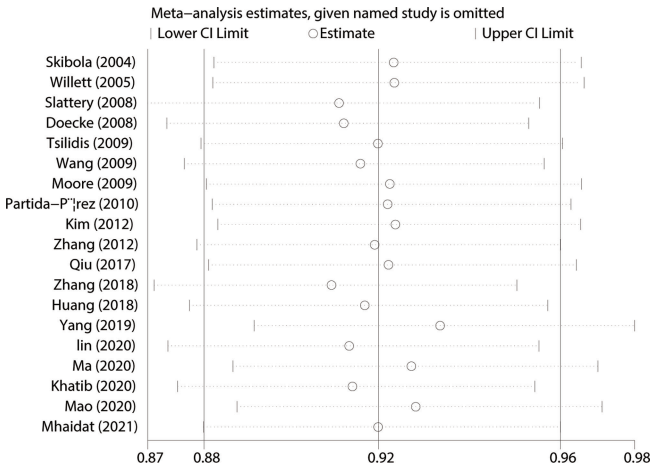


FIGURE 5 | Sensitivity analysis of the influence of A vs. G comparison.

lacking subgroup analysis of the cancer system. Meanwhile, an expanding body of literature on the relationship between LEP G19A polymorphism and cancer risk has been published. Therefore, we conducted this meta-analysis to address this relevance between the LEP G19A polymorphism and cancer risk.

Our current meta-analysis contained 19 studies of cancers containing 9,878 patients, and 14,251 controls were pooled, which contained more participants and cancer types than the previous meta-analysis. Overall, we found a significant correlation between the LEP G19A mutation and susceptibility to cancers under four models (allele model, dominant model, recessive model, and homozygote model), which means that this mutation may decrease the risk of overall cancer. This result was confirmed in a meta-analysis conducted by Liu et al. (41) and Yang et al. (29). Studies (42, 43) confirmed that the LEP G19A mutation might reduce mRNA translation with a lower serum level of LEP, which may attenuate the cancer risk as a protective factor.

Obesity was defined as an imbalance between caloric consumption and energy expenditure. Meanwhile, the LEP is a metabolic-related factor regulating balancing by regulating acquisition and consumption of energy. So it seems that obesity has a positive correlation with LEP polymorphism. However, some studies showed that there was no association between LEP polymorphism and obesity (44, 45). Mizuta et al. (46) study showed that LEP G19A was not associated with obesity. The study by Nesrine et al. (47) even showed that different polymorphisms of the LEP gene have distinct correlations with obesity. Our study showed that LEP G19A polymorphism decreases cancer risk, but the exact mechanism is unknown and mounting evidence indicates that obesity may greatly increase the risk of cancer (3–5). This provides us with a

hint that LEP G19A polymorphism may not lead to cancer by gaining weight. Further studies are needed to elucidate the mechanism of action of LEP G19A polymorphism and cancer.

When stratified by ethnicity, we found a significant correlation between this mutation and Asians and no significant in Caucasians, which means that this mutation may decrease the risk of Asian people not Caucasians. This difference might be caused by a discrepancy in the interplay between genes and the environment. Moreover, the frequency of the A allele in Caucasians (68%) and Asians (44%) might be the reason for contributing to the discrepancy in the non-significant results in Caucasians. When stratified by cancer type and cancer system, it was first to describe the association between LEP G19A mutation and the cancer system. We found a significant correlation between this mutation and colorectal cancer, esophageal cancer, digestive system cancer, and urinary system cancer, which means that this mutation may decrease the risk of colorectal cancer, esophageal cancer, digestive system cancer, and urinary system cancer, but we found no correlation between this mutation and the other cancer system; the reason for this difference in risk with different tumors is as yet unknown, possibly due to LEP and its receptors playing various roles in the mediation of physiological reactions and carcinogenesis in different pathological types of cancer.

Heterogeneity may exist in our meta-analysis of cancer in the overall analysis. Stratified analyses indicated that heterogeneity was significant in some subgroups (e.g., Asians, esophageal cancer, and urinary system cancer). These factors may cause heterogeneity in our study. We checked the stability of our pooled results by sensitivity analyses. The trend of relevance was not significantly changed in the sensitivity analyses, which meant that the pooled results in our meta-analysis were statistically stable. We used Begg's and Egger's tests to evaluate publication bias. Begg's and Egger's tests' p-values > 0.05 in all models, so that publication bias may exist in this meta-analysis.

The following limitations should be mentioned: (1) The number of studies focused on the relationship between LEP G19A and cancer was relatively small, so little information about stratified analyses of ethnicity, cancer type, and cancer system was available; therefore, further studies are required to determine the actual relationship in all populations. (2) Our study had no access to other potential factors influencing the results, such as other lifestyles, environments, and ages.

CONCLUSION

In conclusion, this meta-analysis suggests that the LEP G19A mutation may decrease the risk of overall cancer, colorectal cancer, esophageal cancer, digestive system cancer, and urinary system cancer. In the future, more comprehensive objects containing genetic environmental interaction are warranted to discover the correlation between LEP G19A mutation and the risk of cancer.

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.

AUTHOR CONTRIBUTIONS

AZ, SW, FZ, WL, QL, and XL conceived the study. FZ, WL, and QL contributed to data acquisition, data interpretation, and statistical analysis. AZ, SW, and XL contributed to the study design, statistical analysis, writing, and revising of the manuscript critically. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Funds of China (82171594) and Zhao Yi-Cheng Medical Science Foundation (ZYYFY2018031).

ACKNOWLEDGMENTS

We thank all authors whose studies were included in our meta-analysis.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590
3. Kolb R, Sutterwala FS, Zhang W. Obesity and Cancer: Inflammation Bridges the Two. *Curr Opin Pharmacol* (2016) 29:77–89. doi: 10.1016/j.coph.2016.07.005
4. Dalamaga M, Christodoulatos GS, Mantzoros CS. The Role of Extracellular and Intracellular Nicotinamide Phosphoribosyl-Transferase In Cancer: Diagnostic and Therapeutic Perspectives and Challenges. *Metabolism* (2018) 82:72–87. doi: 10.1016/j.metabol.2018.01.001
5. Garg SK, Maurer H, Reed K, Selagamsetty R. Diabetes and Cancer: Two Diseases With Obesity as a Common Risk Factor. *Diabetes Obes Metab* (2014) 16:97–110. doi: 10.1111/dom.12124
6. Bieńkiewicz J, Romanowicz H, Wilczyński M, Jabłoński G, Stepowicz A, Obłękowska A, et al. Association of Single Nucleotide Polymorphism LEP-R C.668A>G (P.Gln223Arg, Rs1137101) of Leptin Receptor Gene With Endometrial Cancer. *BMC Cancer* (2021) 21:925. doi: 10.1186/s12885-021-08620-y
7. Hołysz H, Paszel-Jaworska A, Romaniuk-Drapała A, Grodecka-Gazdecka S, Rubiś B. LEP (-2548g>A LEP) and LEPR (223gln>Arg, 109lys>Arg) Polymorphisms as Breast Cancer Risk Factors in the Polish Female Population. *Mol Biol Rep* (2021) 48:3237–44. doi: 10.1007/s11033-021-06328-7
8. Tayel SI, Alhanafy AM, Ahmed SM, Eltorgoman AA, Elsayed IE. Biochemical Study on Modifying Role of Variants of Leptin Gene and its Receptor on Serum Leptin Levels in Breast Cancer. *Mol Biol Rep* (2020) 47:3807–20. doi: 10.1007/s11033-020-05436-0
9. Delort L, Bougaret L, Cholet J, Vermerie M, Billard H, Decombat C, et al. Hormonal Therapy Resistance and Breast Cancer: Involvement of Adipocytes and Leptin. *Nutrients* (2019) 11(12):2839–55. doi: 10.3390/nu11122839
10. Ghasemi A, Saeidi J, Azimi-Nejad M, Hashemy SI. Leptin-Induced Signaling Pathways in Cancer Cell Migration and Invasion. *Cell Oncol (Dordr)* (2019) 42:243–60. doi: 10.1007/s13402-019-00428-0

11. Olea-Flores M, Juárez-Cruz JC, Mendoza-Catalán MA, Padilla-Benavides T, Navarro-Tito N. Signaling Pathways Induced by Leptin During Epithelial–Mesenchymal Transition in Breast Cancer. *Int J Mol Sci* (2018) 19(11):3493–513. doi: 10.3390/ijms19113493
12. Llanos A, Yao S, Singh A, Aremu JB, Khiabanian H, Lin Y, et al. Gene Expression of Adipokines and Adipokine Receptors in the Tumor Microenvironment: Associations of Lower Expression With More Aggressive Breast Tumor Features. *Breast Cancer Res Treat* (2021) 185:785–98. doi: 10.1007/s10549-020-05972-0
13. Nwadozi E, Ng A, Strömberg A, Liu HY, Olsson K, Gustafsson T, et al. Leptin is a Physiological Regulator of Skeletal Muscle Angiogenesis and is Locally Produced by PDGFRα and PDGFRβ Expressing Perivascular Cells. *Angiogenesis* (2019) 22:103–15. doi: 10.1007/s10456-018-9641-6
14. Clements VK, Long T, Long R, Figley C, Smith D, Ostrand-Rosenberg S. Frontline Science: High Fat Diet and Leptin Promote Tumor Progression by Inducing Myeloid-Derived Suppressor Cells. *J Leukoc Biol* (2018) 103:395–407. doi: 10.1002/JLB.4HI0517-210R
15. Ray A, Cleary MP. The Potential Role of Leptin in Tumor Invasion and Metastasis. *Cytokine Growth Factor Rev* (2017) 38:80–97. doi: 10.1016/j.cytogfr.2017.11.002
16. Skibola CF, Holly EA, Forrest MS, Hubbard A, Bracci PM, Skibola DR, et al. Body Mass Index, Leptin and Leptin Receptor Polymorphisms, and non-Hodgkin Lymphoma. *Cancer Epidemiol Biomarkers Prev* (2004) 13:779–86.
17. Willett EV, Skibola CF, Adamson P, Skibola DR, Morgan GJ, Smith MT, et al. Non-Hodgkin's Lymphoma, Obesity and Energy Homeostasis Polymorphisms. *Br J Cancer* (2005) 93:811–6. doi: 10.1038/sj.bjc.6602762
18. Slattery ML, Wolff RK, Herrick J, Caan BJ, Potter JD. Leptin and Leptin Receptor Genotypes and Colon Cancer: Gene-Gene and Gene-Lifestyle Interactions. *Int J Cancer* (2008) 122:1611–7. doi: 10.1002/ijc.23135
19. Tsilidis KK, Helzlsouer KJ, Smith MW, Grinberg V, Hoffman-Bolton J, Clipp SL, et al. Association of Common Polymorphisms in IL10, and in Other Genes Related to Inflammatory Response and Obesity With Colorectal Cancer. *Cancer Causes Control* (2009) 20:1739–51. doi: 10.1007/s10552-009-9427-7
20. Doecke JD, Zhao ZZ, Stark MS, Green AC, Hayward NK, Montgomery GW, et al. Single Nucleotide Polymorphisms in Obesity-Related Genes and the Risk of Esophageal Cancers. *Cancer Epidemiol Biomarkers Prev* (2008) 17:1007–12. doi: 10.1158/1055-9965.EPI-08-0023
21. Wang MH, Helzlsouer KJ, Smith MW, Hoffman-Bolton JA, Clipp SL, Grinberg V, et al. Association of IL10 and Other Immune Response- and Obesity-Related Genes With Prostate Cancer in CLUE II. *Prostate* (2009) 69:874–85. doi: 10.1002/pros.20933
22. Moore SC, Leitzmann MF, Albanes D, Weinstein SJ, Snyder K, Virtamo J, et al. Adipokine Genes and Prostate Cancer Risk. *Int J Cancer* (2009) 124:869–76. doi: 10.1002/ijc.24043
23. Partida-Pérez M, de la Luz AM, Peregrina-Sandoval J, Macías-Gómez N, Moreno-Ortiz J, Leal-Ugarte E, et al. Association of LEP and ADIPOQ Common Variants With Colorectal Cancer in Mexican Patients. *Cancer Biomark* (2010) 7:117–21. doi: 10.3233/CBM-2010-0154
24. Kim KZ, Shin A, Lee YS, Kim SY, Kim Y, Lee ES. Polymorphisms in Adiposity-Related Genes are Associated With Age at Menarche and Menopause in Breast Cancer Patients and Healthy Women. *Hum Reprod* (2012) 27:2193–200. doi: 10.1093/humrep/des147
25. Zhang Y, Wang MY, He J, Wang JC, Yang YJ, Jin L, et al. Tumor Necrosis Factor-α Induced Protein 8 Polymorphism and Risk of non-Hodgkin's Lymphoma in a Chinese Population: A Case-Control Study. *PloS One* (2012) 7:e37846. doi: 10.1371/journal.pone.0037846
26. Qiu H, Lin X, Tang W, Liu C, Chen Y, Ding H, et al. Investigation of TCF7L2, LEP and LEPR Polymorphisms With Esophageal Squamous Cell Carcinomas. *Oncotarget* (2017) 8:109107–19. doi: 10.18632/oncotarget.22619
27. Zhang S, Jiang J, Chen Z, Wang Y, Tang W, Liu C, et al. Investigation of LEP and LEPR Polymorphisms With the Risk of Hepatocellular Carcinoma: A Case-Control Study in Eastern Chinese Han Population. *Onco Targets Ther* (2018) 11:2083–9. doi: 10.2147/OTT.S153931
28. Huang BZ, Tsilidis KK, Smith MW, Hoffman-Bolton J, Visvanathan K, Platz EA, et al. Polymorphisms in Genes Related to Inflammation and Obesity and Colorectal Adenoma Risk. *Mol Carcinog* (2018) 57:1278–88. doi: 10.1002/mc.22842
29. Yang J, Zhong Z, Tang W, Chen J. Leptin Rs2167270 G > A (G19A) Polymorphism may Decrease the Risk of Cancer: A Case-Control Study and Meta-Analysis Involving 19 989 Subjects. *J Cell Biochem* (2019) 120:10998–1007. doi: 10.1002/jcb.28378
30. Lin J, Xie Z, Lan B, Guo Z, Tang WF, Liu C, et al. Investigation of Leptin and its Receptor (LEPR) for Single Nucleotide Polymorphisms In Colorectal Cancer: A Case-Control Study Involving 2,306 Subjects. *Am J Transl Res* (2020) 12:3613–28. doi: 10.1200/JCO.2020.38.15_suppl.e16100
31. Ma R, He Q. A Variant of Leptin Gene Decreases the Risk of Gastric Cancer in Chinese Individuals: Evidence From a Case-Control Study. *Pharmgenomics Pers Med* (2020) 13:397–404. doi: 10.2147/PGPM.S258672
32. Al-Khatib SM, Abdo N, Al-Eitan LN, Al-Mistarehi AW, Zahran DJ, Kewan TZ. LTA, LEP, and TNF-A Gene Polymorphisms are Associated With Susceptibility and Overall Survival of Diffuse Large B-Cell Lymphoma in an Arab Population: A Case-Control Study. *Asian Pac J Cancer Prev* (2020) 21:2783–91. doi: 10.31557/APJCP.2020.21.9.2783
33. Mao F, Niu XB, Gu S, Ji L, Wei BJ, Wang HB. Investigation of Leptin G19A Polymorphism With Bladder Cancer Risk: A Case-Control Study. *J Clin Lab Anal* (2020) 34:e23351. doi: 10.1002/jcla.23351
34. Mhaidat NM, Alzoubi KH, Kubas MA, Banihani MN, Hamdan N, Al-Jaberi TM. High Levels of Leptin and non-High Molecular Weight-Adiponectin in Patients With Colorectal Cancer: Association With Chemotherapy and Common Genetic Polymorphisms. *BioMed Rep* (2021) 14:13. doi: 10.3892/br.2020.1389
35. Vander HM, DeBerardinis RJ. Understanding the Intersections Between Metabolism and Cancer Biology. *Cell* (2017) 168:657–69. doi: 10.1016/j.cell.2016.12.039
36. Xu H, Zhou S, Tang Q, Xia H, Bi F. Cholesterol Metabolism: New Functions and Therapeutic Approaches in Cancer. *Biochim Biophys Acta Rev Cancer* (2020) 1874:188394. doi: 10.1016/j.bbcan.2020.188394
37. Li Z, Zhang H. Reprogramming of Glucose, Fatty Acid and Amino Acid Metabolism for Cancer Progression. *Cell Mol Life Sci* (2016) 73:377–92. doi: 10.1007/s00018-015-2070-4
38. Lee KN, Choi HS, Yang SY, Park HK, Lee YY, Lee OY, et al. The Role of Leptin in Gastric Cancer: Clinicopathologic Features and Molecular Mechanisms. *Biochem Biophys Res Commun* (2014) 446:822–9. doi: 10.1016/j.bbrc.2014.02.072
39. Dong Z, Fu S, Xu X, Yang Y, Du L, Li W, et al. Leptin-Mediated Regulation of ICAM-1 is Rho/ROCK Dependent and Enhances Gastric Cancer Cell Migration. *Br J Cancer* (2014) 110:1801–10. doi: 10.1038/bjc.2014.70
40. Aleksandrova K, Boeing H, Jenab M, Bueno-de-Mesquita HB, Jansen E, van Duijnhoven FJ, et al. Leptin and Soluble Leptin Receptor in Risk of Colorectal Cancer in the European Prospective Investigation Into Cancer and Nutrition Cohort. *Cancer Res* (2012) 72:5328–37. doi: 10.1158/0008-5472.CAN-12-0465
41. Liu P, Shi H, Huang C, Shu H, Liu R, Yang Y, et al. Association of LEP A19G Polymorphism With Cancer Risk: A Systematic Review and Pooled Analysis. *Tumour Biol* (2014) 35:8133–41. doi: 10.1007/s13277-014-2088-5
42. Pan H, Deng LL, Cui JQ, Shi L, Yang YC, Luo JH, et al. Association Between Serum Leptin Levels and Breast Cancer Risk: An Updated Systematic Review and Meta-Analysis. *Medicine (Baltimore)* (2018) 97:e11345. doi: 10.1097/MD.0000000000011345
43. Savino F, Rossi L, Di Stasio L, Galliano I, Montanari P, Bergallo M. Mismatch Amplification Mutation Assay Real-Time PCR Analysis of the Leptin Gene G2548A and A19G Polymorphisms and Serum Leptin in Infancy: A Preliminary Investigation. *Horm Res Paediatr* (2016) 85:318–24. doi: 10.1159/000444484
44. Fairbrother U, Kidd E, Malagamuwa T, Walley A. Genetics of Severe Obesity. *Curr Diabetes Rep* (2018) 18:85. doi: 10.1007/s11892-018-1053-x
45. Zhang L, Yuan LH, Xiao Y, Lu MY, Zhang LJ, Wang Y. Association of Leptin Gene -2548 G/A Polymorphism With Obesity: A Meta-Analysis. *Ann Nutr Metab* (2014) 64:127–36. doi: 10.1159/000363392
46. Mizuta E, Kokubo Y, Yamanaka I, Miyamoto Y, Okayama A, Yoshimasa Y, et al. Leptin Gene and Leptin Receptor Gene Polymorphisms are Associated With Sweet Preference and Obesity. *Hypertens Res* (2008) 31:1069–77. doi: 10.1291/hypres.31.1069
47. Nesrine Z, Haithem H, Imen B, Fadoua N, Asma O, Fadhel NM, et al. Leptin and Leptin Receptor Polymorphisms, Plasma Leptin Levels and Obesity in

Tunisian Volunteers. *Int J Exp Pathol* (2018) 99:121–30. doi: 10.1111/iep.12271

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fonc.2021.771312

Identification of Recessively Inherited Genetic Variants Potentially Linked to Pancreatic Cancer Risk

Ye Lu[1,2], Manuel Gentiluomo[3], Angelica Macauda[1], Domenica Gioffreda[4], Maria Gazouli[5], Maria C. Petrone[6], Dezső Kelemen[7], Laura Ginocchi[8], Luca Morelli[9], Konstantinos Papiris[10], William Greenhalf[11], Jakob R. Izbicki[12], Vytautas Kiudelis[13], Beatrice Mohelníková-Duchoňová[14], Bas Bueno-de-Mesquita[15], Pavel Vodicka[16,17,18], Hermann Brenner[19,20,21], Markus K. Diener[22], Raffaele Pezzilli[23], Audrius Ivanauskas[13], Roberto Salvia[24], Andrea Szentesi[25,26], Mateus Nóbrega Aoki[27], Balázs C. Németh[28], Cosimo Sperti[29], Krzysztof Jamroziak[30], Roger Chammas[31,32], Martin Oliverius[33], Livia Archibugi[6,34,35], Stefano Ermini[36], János Novák[37], Juozas Kupcinskas[13], Ondřej Strouhal[14,38], Pavel Souček[39], Giulia M. Cavestro[40], Anna C. Milanetto[29], Giuseppe Vanella[6,34,35], John P. Neoptolemos[22], George E. Theodoropoulos[41], Hanneke W. M. van Laarhoven[42], Andrea Mambrini[8], Stefania Moz[43], Zdenek Kala[44], Martin Loveček[45], Daniela Basso[43], Faik G. Uzunoglu[12], Thilo Hackert[22], Sabrina G. G. Testoni[6], Viktor Hlaváč[39], Angelo Andriulli[4], Maurizio Lucchesi[8], Francesca Tavano[4], Silvia Carrara[46], Péter Hegyi[25,26], Paolo G. Arcidiacono[6], Olivier R. Busch[47], Rita T. Lawlor[48], Marta Puzzono[40], Ugo Boggi[49], Feng Guo[19], Ewa Małecka-Panas[50], Gabriele Capurso[6,34,35], Stefano Landi[3], Renata Talar-Wojnarowska[50], Oliver Strobel[22], Xin Gao[19], Yogesh Vashist[51], Daniele Campa[3†] and Federico Canzian[1†]*

[1] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [2] Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany, [3] Department of Biology, University of Pisa, Pisa, Italy, [4] Division of Gastroenterology and Research Laboratory, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, [5] Department of Basic Medical Sciences, Laboratory of Biology, Medical School, National and Kapodistrian University of Athens, Athens, Greece, [6] Pancreato-Biliary Endoscopy and Endosonography Division, Pancreas Translational and Clinical Research Center, San Raffaele Scientific Institute Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy, [7] Department of Surgery, Medical School, University of Pécs, Pécs, Hungary, [8] Oncological Department, Oncological Unit of Massa Carrara, Azienda Unità Sanitaria Locale (USL) Toscana Nord Ovest, Carrara, Italy, [9] General Surgery, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy, [10] Endoscopic Surgery Department, Hippocratio General Hospital of Athens, Athens, Greece, [11] Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, United Kingdom, [12] Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [13] Department of Gastroenterology, Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania, [14] Department of Oncology, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc, Olomouc, Czechia, [15] Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands, [16] Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czechia, [17] Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czechia, [18] First Faculty of Medicine, Institute of Biology and Medical Genetics, Prague, Czechia, [19] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [20] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [21] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [22] Department of General Surgery, University of Heidelberg, Heidelberg, Germany, [23] Department of Internal Medicine, San Carlo Hospital, Potenza, Italy, [24] Department of General and Pancreatic Surgery, The Pancreas Institute, University and Hospital Trust of Verona, Verona, Italy, [25] Institute for Translational Medicine, Medical School, University of Pécs, Pécs, Hungary, [26] Centre for Translational Medicine, Department of Medicine, University of Szeged, Szeged, Hungary, [27] Laboratory for Applied Science and Technology in Health, Carlos Chagas Institute, Curitiba, Brazil, [28] First Department of Medicine, University of Szeged, Szeged, Hungary, [29] Department of Surgery-Dipartimento di Scienze Chirurgiche Oncologiche e Gastroenterologiche (DiSCOG), Padua University Hospital, Padua, Italy, [30] Department of

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Phuong Cao Thi Ngoc, Lund University, Sweden
Florentia Fostira, National Centre of Scientific Research Demokritos, Greece

***Correspondence:**
Federico Canzian
f.canzian@dkfz.de

[†]These authors share last authorship

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: 06 September 2021
Accepted: 16 November 2021
Published: 03 December 2021

Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, Warsaw, Poland, [31] Department of Radiology and Oncology, Institute of Cancer of São Paulo (ICESP), São Paulo, Brazil, [32] Faculty of Medicine, University of São Paulo, São Paulo, Brazil, [33] Department of Surgery, Faculty Hospital Kralovske Vinohrady and Third Faculty of Medicine, Charles University, Prague, Czechia, [34] Digestive and Liver Disease Unit, Sant' Andrea Hospital, Rome, Italy, [35] Faculty of Medicine and Psychology, Sapienza University of Rome, Rome, Italy, [36] Blood Transfusion Service, Children's Hospital, Azienda Ospedaliero-Universitaria Meyer, Florence, Italy, [37] Pándy Kálmán Hospital of Békés County, Gyula, Hungary, [38] Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czechia, [39] Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czechia, [40] Division of Experimental Oncology, Gastroenterology and Gastrointestinal Endoscopy Unit, Vita-Salute San Raffaele University, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Raffaele Scientific Institute, Milan, Italy, [41] First Propaedeutic University Surgery Clinic, Hippocratio General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece, [42] Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [43] Department of Medicine (DIMED), Padua University Hospital, Padua, Italy, [44] Department of Surgery, University Hospital Brno Bohunice, Faculty of Medicine, Masaryk University, Brno, Czechia, [45] Department of Surgery I, Faculty of Medicine and Dentistry, Palacky University Olomouc and University Hospital Olomouc, Olomouc, Czechia, [46] Division of Gastroenterology and Digestive Endoscopy, Department of Gastroenterology, Humanitas Clinical and Research Center Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy, [47] Department of Surgery, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [48] Applied Research on Cancer (ARC)-Net Research Center, University and Hospital Trust of Verona, Verona, Italy, [49] Division of General and Transplant Surgery, Pisa University Hospital, Pisa, Italy, [50] Department of Digestive Tract Diseases, Medical University of Lodz, Lodz, Poland, [51] Centre for Surgical Oncology, Medias Klinikum Burghausen, Burghausen, Germany

Although 21 pancreatic cancer susceptibility loci have been identified in individuals of European ancestry through genome-wide association studies (GWASs), much of the heritability of pancreatic cancer risk remains unidentified. A recessive genetic model could be a powerful tool for identifying additional risk variants. To discover recessively inherited pancreatic cancer risk loci, we performed a re-analysis of the largest pancreatic cancer GWAS, the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case-Control Consortium (PanC4), including 8,769 cases and 7,055 controls of European ancestry. Six single nucleotide polymorphisms (SNPs) showed associations with pancreatic cancer risk according to a recessive model of inheritance. We replicated these variants in 3,212 cases and 3,470 controls collected from the PANcreatic Disease ReseArch (PANDoRA) consortium. The results of the meta-analyses confirmed that rs4626538 (7q32.2), rs7008921 (8p23.2) and rs147904962 (17q21.31) showed specific recessive effects ($p<10^{-5}$) compared with the additive effects ($p>10^{-3}$), although none of the six SNPs reached the conventional threshold for genome-wide significance ($p < 5\times10^{-8}$). Additional bioinformatic analysis explored the functional annotations of the SNPs and indicated a possible relationship between rs36018702 and expression of the *BCL2L11* and *BUB1* genes, which are known to be involved in pancreatic biology. Our findings, while not conclusive, indicate the importance of considering non-additive genetic models when performing GWAS analysis. The SNPs associated with pancreatic cancer in this study could be used for further meta-analysis for recessive association of SNPs and pancreatic cancer risk and might be a useful addiction to improve the performance of polygenic risk scores.

Keywords: pancreatic cancer, susceptibility, genome-wide association study, recessive model, genetic polymorphisms

INTRODUCTION

Pancreatic cancer ranks fourth for cancer-related deaths in western countries and is projected to become the second by 2030 (1, 2). It is a very deadly disease with the mortality rate closely approaching to the incidence rate. The median survival is less than 18 months, and the 5-year survival rate remains as low as 3 ~ 15% (3–5). The poor prognosis is mainly due to the late onset of symptoms, diagnosis at an advanced stage and subsequent rapid progression. A comprehensive identification

of the risk factors can be instrumental to a better understanding of the disease etiology and to the development of methods for risk stratification, that in turn could facilitate early detection, which at the moment remains elusive.

Genetic factors play an important role in the etiology of pancreatic cancer (6). Genome-wide association studies (GWAS) have identified various frequent genetic variants associated with pancreatic cancer risk. The two largest pancreatic cancer GWAS done in European populations are the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case-Control Consortium (PanC4), and a total of 21 susceptibility loci associated at genome-wide significance level have been discovered, and studied individually and in combination (7–15). However, the identified SNPs explain only 4.1% of the total phenotypic variance of pancreatic cancer, which do not fully account for the overall 21.2% estimated genetic heritability (16). This can be explained by the relatively small effect sizes of the individual risk loci, and by the strict multiple testing correction required for GWAS (typically $p < 5x10^{-8}$), which is likely to result in a large number of false negatives.

Over the past decade, GWAS have achieved substantial success in discovering many common variants underlying the genetic architecture of complex diseases (17), including pancreatic cancer. Standard models for implying specific relationships between genotypes and phenotypes include additive, recessive and dominant models (18). The association of biallelic single nucleotide polymorphisms (SNPs) having alleles A/a with a given endpoint (e.g. disease risk) is typically analyzed with a logistic regression model $\text{logit}(P) = \alpha + \beta (X)$, where in an additive model X = 0, 1 or 2 depending on the genotype (homozygotes A/A, heterozygotes A/a and homozygotes a/a, respectively), thus the risk of disease is increased exp(β)-fold for subjects with genotype A/a and exp (2β)-fold for subjects with genotype a/a. A recessive model compares rare homozygotes a/a (who will have X=1) versus the rest (combining heterozygotes A/a and common homozygotes A/A, who will have X=0); a dominant model compares A/A (X=0) versus A/a + a/a (X=1). As most GWAS studies assume that allelic effects are additive, most of the associations reported in GWAS consider only the additive model of inheritance. But for variants which do not follow an intermediate model of inheritance, the recessive or the dominant genetic model can have more power to detect associations. Reanalysis of GWAS data with the recessive model of inheritance, considering homozygotes for the minor allele as the only "exposed" category could help to identify additional risk loci for non-negligible subsets of SNPs (19).

To discover novel recessively inherited pancreatic cancer risk loci, we performed a secondary analysis using genotyping data from all published pancreatic cancer GWAS conducted in subjects of European origin, and then replicated the most promising variants in cases and controls collected from the PANcreatic Disease ReseArch (PANDoRA) consortium. Better understanding the genetic background of the disease could be an invaluable tool to stratify the population by individual risk and increase our chances of early detection.

MATERIALS AND METHODS

Study Populations

The following publicly available GWAS datasets on pancreatic cancer risk were used for this study: the Pancreatic Cancer Cohort Consortium (PanScan, comprising of PanScan I, PanScan II, and PanScan III) and the Pancreatic Cancer Case Control Consortium (PanC4). We obtained the genotype data from the NCBI database of genotypes and phenotypes (dbGaP) (study accession numbers phs000206.v5.p3 and phs000648.v1.p1; project reference #12644). We performed standard quality control and genotype imputation for the four datasets separately, using the Michigan Imputation Server (https://imputationserver.sph.umich.edu) (20) and the Haplotype Reference Consortium (HRC, V.r1.1) reference panel (21). Before imputation, we implemented individual- and SNP-level quality control steps as follows: individual and SNP missingness (call rate<0.9); sex discrepancy; heterozygosity (>3 SD from the mean); relatedness (PI_HAT>0.2, i.e., subjects related up to the second degree); ethnic outliers (population structure was captured by principal component analysis to remove non-European ancestry individuals); minor allele frequency (MAF) <0.005; and Hardy-Weinberg equilibrium (HWE) ($p<1\times10^{-6}$). After imputation, we removed SNPs with low imputation quality (INFO score $r^2<0.7$, MAF<0.05 or call rate<0.9). Then, we merged the four imputed datasets and rechecked for the relatedness in the pooled dataset. At the end, a total of 5,056,279 SNPs in 8,769 cases and 7,055 controls (8,600 males and 7,224 females) remained for further analysis.

Additional samples belonging to the PANDoRA consortium, mostly from European populations, were selected for genotyping. Cases were diagnosed with pancreatic ductal adenocarcinoma (PDAC) and were all collected from the PANDoRA consortium (22). Controls were from the same geographical regions as the cases. A subset of the German controls (N=932) derived from ESTHER, a prospective cohort with 9,953 participants recruited in the Saarland region of Germany during a general health check-up in the period of July 2000 and December 2002. British and Dutch controls were collected from the European Prospective Investigation on Cancer (EPIC, http://epic.iarc.fr/), a prospective cohort study consisting of general population healthy volunteers from ten European countries (23). All subjects provided written informed consent. Approval for the PANDoRA study protocol (including for controls from ESTHER and EPIC cohorts) was received from the Ethics Commission of the Medical Faculty of the University of Heidelberg.

SNP Selection

We performed the association analysis on the pooled imputed PanScan+PanC4 GWAS data using both additive and recessive models. Association statistics (odds ratios (OR) and 95% confidence intervals (CI)) on PDAC risk were obtained with logistic regression adjusting for age, sex and the top ten principal components using PLINK version 1.9 (24). There were 268 SNPs that showed an association with p-value lower than 10^{-5},

according to a recessive model (**Supplementary Table 1**). Most of them overlapped with previously reported pancreatic cancer risk loci (1q32.1, 2p14, 3q28, 5p15.33, 7p14.1, 7q32.3, 9q34, 13q12.2 and 16q23.1) from additive analyses. Among remaining SNPs which were over 1 Mb away from the closest known locus and showed no linkage disequilibrium (LD) with known loci ($r^2 < 0.01$), ten SNPs at six loci, showed large differences in p-values using the two models ($p < 10^{-5}$ using the recessive model, and $p > 10^{-3}$ using the additive model). After filtering SNPs in LD ($r^2>0.8$, N=3) and removing SNPs that showed p≥0.05 for association with PDAC risk in either PanScan or PanC4 (N=1), the top six promising SNPs were moved forward to genotyping.

Genotyping

DNA of PANDoRA samples was isolated from whole blood using QIAamp DNA extraction kit (Qiagen) and distributed in 384-well plates for genotyping. For quality control, 8% of the samples was randomly duplicated throughout the plates and no template controls (NTC) were used in each genotyping plate. Genotyping was performed using TaqMan (ABI, Applied Biosystems, Foster City, CA, USA) and KASP (KBioscence, Hoddesdon, UK) probes on the Real-Time PCR system. Since the genotyping assay for rs147904962 failed to work, rs12943205 was genotyped as a proxy SNP, in high LD ($r^2 = 0.99$). Detection was done with a Viia7 instrument and Viia7 software (Applied Biosystems, Foster City, CA). After calling all the genotypes, samples with a call rate < 83.3% (i.e., missing more than one genotype) were removed. Duplicated samples with low concordance rate (>1 discordant genotype) were excluded. Discordance from HWE distribution was checked in controls, in the overall population and by country, and all the genotyped SNPs were in HWE ($p>10^{-3}$). Dutch and British controls were genotyped in the context of a GWAS using the Human 660W-Quad BeadChip array (Illumina, San Diego, CA). Quality control steps were performed after TaqMan genotyping. Finally, 3,212 PDAC cases and 3,470 controls were included for further analysis. The characteristics of the study population are summarized in **Table 1**.

TABLE 1 | Characteristics of genotyped samples from PANDoRA after quality control.

	Cases	Controls
Male, %	55.0	51.6
Median age, (25th-75th percentile)	66 (58-73)	60 (51-68)
Country, N		
Czech Republic	430	173
Germany	683	1018
Greece	109	16
Hungary	290	413
Italy	1298	1280
Lithuania	102	179
Poland	90	195
Netherlands	117	62
United Kingdom	93	134
Total	3212	3470

Statistical Analysis

To investigate the effect of the genotyped SNPs (rare allele vs. common allele; rare homozygous genotype vs. heterozygous plus common homozygous genotypes) in PANDoRA samples on the PDAC risk, we performed unconditional logistic regression adjusting for sex, age and country. Then we performed meta-analyses using R package "meta" by fixed-effects model (or random-effects model when $p < 0.05$ in the heterogeneity test) between phase one (reanalysis of the pancreatic cancer GWASs, PanScan and PanC4) and phase two (replication in samples collected from PANDoRA), with a final sample size of 11,981 PDAC cases and 10,525 controls. For the analysis with the genotyped SNPs in phase two, age, sex and genotypes had missing rates between 1% to 5%. Considering that missing data can have a significant effect on the conclusion, we applied multiple imputation which is a missing data method that provides valid statistical inferences under the missing at random condition (25). The R package "mice", which imputes incomplete multivariate data by chained equations (26), was used to impute five times the variables involved in analysis, to analyze each of the imputed datasets separately based on the logistic regression model, then to automatically combine all the results together. Since the Brazilian population is known to be ethnically admixed, we performed additional statistical analyses with the PANDoRA Brazilian cases and controls. Meta-analyses were performed after multiple imputation as well. Analyses were carried out with R V3.6.

In addition, we performed gene-based analysis using MAGMA v1.08 to test the associations between all coding genes and PDAC risk based on the p-values under additive and recessive models respectively (27).

Bioinformatic Tools

We used the following tools/databases to explore the possible function of candidate SNPs: the Genotype-Tissue Expression (GTEx, 8th version) project portal (https://www.gtexportal.org, accessed on 30 June 2020), HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and RegulomedB (https://www.regulomedb.org/regulome-search/) (28–30). The Gene Expression Profiling Interactive Analysis (GEPIA2) database (http://gepia2.cancer-pku.cn) was applied to verify the expression levels and evaluate the prognostic value of genes of interest in pancreas tumor and normal tissues (31). Three-Dimensional-genome Interaction Viewer (3DIV, http://3div.kr), which collected all publicly available high-throughput chromatin conformation capture (Hi-C) data from human cell/tissue types, was used to explore the locus regulatory effects of the 3D genome (32). SNPnexus (https://www.snp-nexus.org/) and OpenTargets Genetics (https://genetics.opentargets.org) summarize the results of many different functional annotations (33, 34). The Functional Mapping and Annotation of Genome-Wide Association Studies platform (FUMA, https://fuma.ctglab.nl) was used to annotate the results of the recessive model GWAS (35).

RESULTS

In the first phase, which was conducted at a genome-wide scale, we re-analyzed the data from the PanScan+PanC4 GWAS dataset according to a recessive model of inheritance, and we observed six SNPs that showed specific recessive associations with PDAC risk with $p<10^{-5}$ while $p>10^{-3}$ using the additive model (**Figure 1**, **Supplementary Table 2**). The correlated SNPs in these regions ($r^2>0.8$ in LD) did not show evidence of stronger association under an additive model (**Supplementary Table 3**, **Supplementary Figure 1**). The associations of the genotyped SNPs with PDAC risk under the additive and recessive genetic models are shown in **Figure 1** (**Supplementary Table 2**). In the validation phase in PANDoRA, no statistically significant associations ($p<0.05$) were observed, using the recessive model, except for rs2066357. However, this SNP showed high heterogeneity, with an opposite effect compared to the discovery phase under the recessive genetic model.

In meta-analyses, none of the six SNPs reached the conventional genome-wide significance threshold ($p<5\times10^{-8}$). However, *MIR96* rs4626538 (OR=0.93; $p=4.42\times10^{-6}$), *RP5-991O23.1* rs7008921 (OR=1.42; $p=9.68\times10^{-6}$) and *ARHGAP27* rs147904962 (OR=0.75; $p=4.08\times10^{-6}$) maintained a specific recessive effect compared to the additively inherited effects ($p=1.06\times10^{-3}$, p=0.05 and $p=3.97\times10^{-3}$, respectively), and the p-values of rs4626538 and rs147904962 in the meta-analysis were slightly lower in comparison with those observed in the first phase. The results after multiple imputation were generally consistent with those without multiple imputation (**Supplementary Table 2**). Results did not change when we added the PANDoRA cases and controls from Brazil, who are ethnically admixed (**Supplementary Tables 4, 5**).

We used data from the GTEx consortium to investigate associations between genetic variants and RNA expression. We observed that the rs147904962-A allele was associated with increased *LRRC37A4P* RNA expression in adipose tissue ($p=8.1\times10^{-6}$). An expanded list of linked SNPs (in LD with our six candidate SNPs, $r^2>0.6$) was also considered for the GTEx analysis; we found that the T allele of rs590097 (in LD with the A allele of rs36018702, $r^2 = 0.74$, D'=1) was associated with higher expression of *BCL2L11* in pancreas ($p=5.64\times10^{-6}$). No expression quantitative trait loci (eQTL) associations in pancreas were found for the other SNPs. Haploreg and RegulomeDB did not show evidence for functional effect for these variants.

Using a threshold of >2 for distance-normalized chromatin interaction frequency, 3DIV predicted *C1orf21* and *APOBEC4* to be interaction genes for rs1339571, *BUB1* for rs36018702, *MIR4423* for rs2066357, *SPPL2C*, *SLC4A1*, *RUNDC3A*, *LOC100133991*, *TEX34*, *ITGA2B*, and *C17orf57* for rs147904962, respectively.

Additional analyses with SNPnexus and OpenTargets Genetics did not suggest any clear functional link between our candidate SNPs and pancreatic physiology or pathology. Likewise, when we reanalyzed with FUMA the results of the GWAS analysis according to the recessive model, we did not

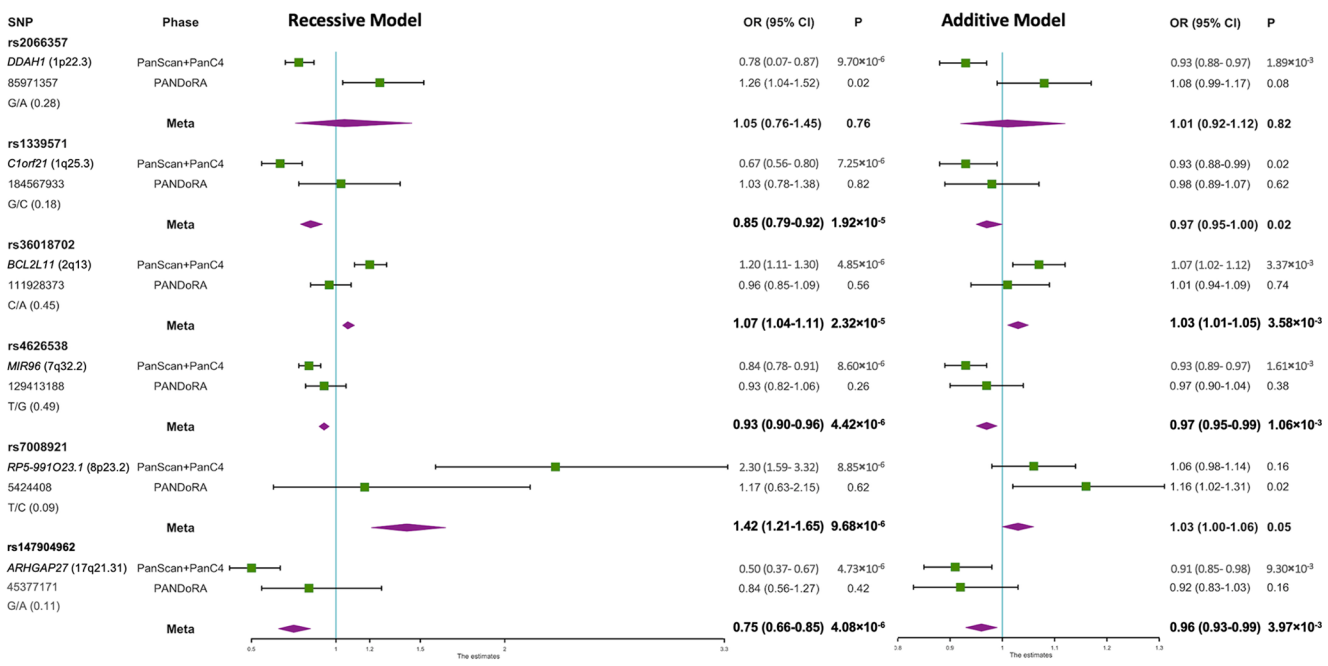


FIGURE 1 | Forest plot of the associations of the 6 SNPs with PDAC risk under recessive and additive genetic models. A forest plot for the 6 SNPs and risk of PDAC is shown by two genetic models using data from discovery and replication analyses combined. Population specific odds ratios (OR) and 95% confidence intervals (CI) are denoted by green boxes and black lines. The combined OR estimates are represented by purple diamonds, where diamond width corresponds to 95% CI bounds. The position information (hg38) and minor allele frequency (MAF) for each SNP are shown on the left.

observe any noteworthy signal in the regions of the six candidate SNPs.

The gene-based analysis using MAGMA based on the p-values of the recessive model revealed that 14 genes were associated with PDAC risk at $p < 0.001$ (**Supplementary Table 6**). Two of these genes showed evidence for association at $p < 0.001$ under the recessive model (*CTSG* 14q12, $p = 2.53 \times 10^{-4}$; *LEPROTL1* 8p12, $p = 4.34 \times 10^{-4}$), but not with the additive one ($p = 0.20$ and $p = 0.10$, respectively). Then we verified the expression level of the two genes in pancreatic cancer patients using GEPIA2. We found that *LEPROTL1* has increased expression in pancreatic cancer tissues compared to adjacent normal pancreatic tissues of the same patients (**Figure 2**).

DISCUSSION

GWAS data are usually analyzed according to an additive genetic model, which is generally considered to be a good surrogate for other genetic models, except for the recessive one (19). Researchers have reported risk variants that showed specifically stronger evidence under a recessive model than an additive model, for type 2 diabetes (36), schizophrenia (37), high triglycerides (38), and other traits (39, 40), but not for PDAC yet. To identify recessive susceptibility loci for PDAC risk, we performed a secondary analysis with the largest currently available pancreatic cancer GWAS datasets (PanScan and PanC4) of European ancestry and attempted the replication of the six most promising variants in additional samples collected from the PANDoRA consortium, with a combined sample size of 11,981 PDAC cases and 10,525 controls. In this study, none of our results reached genome-wide statistical significance ($p<5 \times 10^{-8}$) in either phase, or in the meta-analyses, therefore our results are not conclusive. However, for five of the six selected SNPs the results of the meta-analysis do not exclude the possible recessive association with pancreatic cancer risk. In particular, rs4626538 (7q32.2), rs7008921 (8p23.2) and rs147904962 (17q21.31) maintained a large difference in significance between recessive effects compared with the additively inherited effects.

None of the previous studies indicated a link between these loci and pancreatic cancer risk. No variants in high LD ($r^2>0.8$) have been previously associated with any trait or disease in GWAS, although variants in low to moderate LD ($r^2 = 0.14$~0.60, D'=0.88~1 in Europeans) with rs147904962 have been reported to be associated with waist-to-hip ratio and with risk of developing allergic diseases. The minor G allele of rs7214661 ($r^2 = 0.19$, D'=0.98) was associated with higher risk of allergic disease (41) while the corresponding A allele of rs147904962 was associated with lower risk of pancreatic cancer in our study. It is consistent with the protective effect of allergy for pancreatic cancer in epidemiologic studies (42).

Additionally, GTEx showed that rs590097 regulates *BCL2L11* expression in pancreas tissue. *BCL2L11* is a member of the BCL2 family and plays a role in neuronal and lymphocyte apoptosis. There is evidence shown that *BCL2L11* is one of the major genes contributing to apoptosis, known to be important for pancreatic biology (http://www.genome.jp/kegg/pathway.html) (43). Moreover, the observed association that rs36018702-A (correlated with rs590097-T) showed increased risk of PDAC is consistent with the higher expression of *BCL2L11* in pancreatic cancer tissues than in normal pancreas tissues found through GEPIA2.

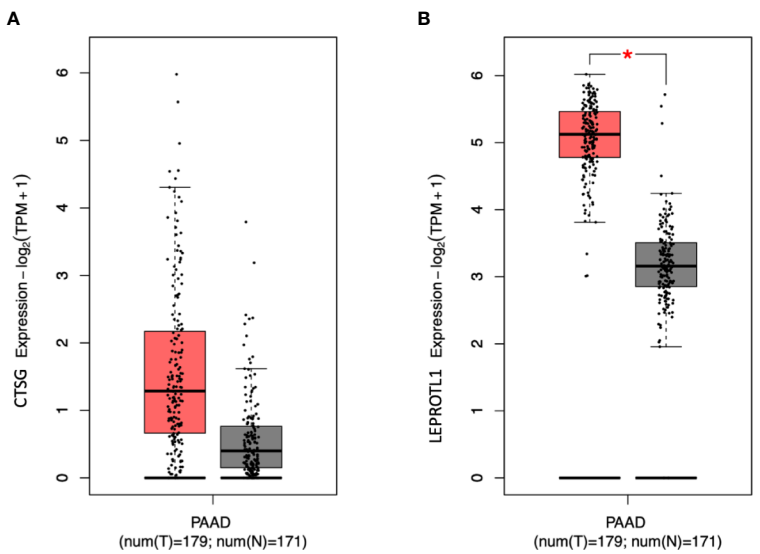


FIGURE 2 | The expression level of *CTSG* and *LEPROTL1* in PAAD patients. GEPIA2 generates box plot for comparing gene expression in pancreatic cancer and paired normal tissues (TCGA tumor versus TCGA normal + GTEx normal). **(A, B)** differential expression analysis. Peach and grey clusters represent tumor and normal samples; * genes with higher |log2FC| values (>1) and lower Q-values (<0.01) were considered differentially expressed genes.

BUB1 is the interaction gene of rs36018702 predicted by 3DIV. There is evidence that *BUB1* is overexpressed in PDAC tissues, suggesting a role of *BUB1* in PDAC progression, and therefore corroborating the association of rs36018702 and PDAC risk (44).

The lowest p-value we observed in the meta-analysis is $4.08x10^{-6}$ for the association of rs147904962 (17q21.31) with the risk of PDAC. rs147904962 is situated 17kb at the 3' end of Rho GTPase Activating Protein 27 (*ARHGAP27*). This gene encodes a member of a large family of proteins that activate Rho-type guanosine triphosphate (GTP) metabolizing enzymes and are involved in cancer through the dysregulation of this mechanism. As *ARHGAP27* mRNA is expressed in pancreatic cancer, we speculate that rs147904962 mediates regulation of cancer-associated *ARHGAP27*, promoting carcinogenesis through dysregulation of Rho/Rac/Cdc42-like GTPases (45). However, it has to be acknowledged that this SNP is not known to be located in a regulatory region of *ARHGAP27*.

Gene-based analyses based on the PanScan and PanC4 datasets (we were not able to replicate these analyses in PANDoRA, which does not have GWAS data) showed that SNPs in *LEPROTL1* and *CTSG* were associated with PDAC risk according to the recessive, but not to the additive model. The bioinformatic analysis identified that *LEPROTL1* was highly expressed in pancreatic cancer compared to matched normal pancreatic tissue of the same patients, suggesting a potential involvement in the etiopathology of PDAC. The leptin receptor overlapping transcript-like 1 gene (*LEPROTL1*) encodes a membrane protein, and may play a role in liver resistance by suppressing the growth hormone activity (46, 47), while the pancreatic cancer-related functions of *LEPROTL1* remain unknown. The cathepsin G gene (*CTSG*) encodes a neutrophil serine protease of the chymotrypsin family, which was shown to affect neutrophil infiltration into the pancreas in a mouse model of pancreatitis (48). Based on this circumstantial evidence it is tempting to speculate a role for this gene and its polymorphisms in modulation of inflammation in the pancreas, which plays a role in the etiology of PDAC. However, to the best of our knowledge, a role for CTSG in pancreatic cancer has not been reported in the literature.

The lack of direct functional evidence for the SNPs of interest from bioinformatic analyses may at least in part reflect the fact that also bioinformatic tools/databases have not been designed to address effects of real recessive alleles. *Ad hoc* tools are needed to better understand the genetic architecture of complex genetic diseases.

It is hard to reach sufficient statistical power to detect variants with recessive effects, unless they are very frequent or have very large effects. Given the effective combined sample size of 11,981 PDAC cases and 10,525 controls, disease prevalence of 1.6%, and a significance cut-off of $p<5\times10^{-8}$, we had at least 80% power to detect a association with ORs equal to those observed in the discovery phase for the rare homozygote genotype for SNPs rs7008921 and rs147904962, whereas for the other SNPs power ranged between 54% to 69%. Thus, our study, in spite of the large sample size, lacked statistical power to confirm the risk with recessive model for some of the SNPs. It is worth noting that between PanScan, PanC4 and PANDoRA we have used the largest available resources for genetics of pancreatic cancer in populations of European origin. Our hypothesis that some variants may be associated with pancreatic cancer risk with a recessive model of inheritance was not disproved, but to prove it convincingly will require even larger datasets that will become available as more GWAS on pancreatic cancer risk are performed.

Identifying high-risk groups could contribute to focus surveillance and invasive screening measures, thereby improving the chance of early detection. Polygenic risk scores (PRS) approaches which could combine modest effect from each risk SNPs have demonstrated accuracies between 59% and 63% for predicting the risk of PDAC when including both non-genetic and genetic factors (14, 49–51). The accuracy of the existing PRS is not ready yet to be used in the clinical practice. It is necessary to expand the PRS with additional risk factors to improve its predictive power. For example, PRS including more SNPs that are not genome-wide significant but having noteworthy effects such as the ones we highlighted in this work may provide an additive contribution to the overall performance.

CONCLUSIONS

In conclusion, we propose some candidate SNPs as recessively inherited genetic variants for pancreatic cancer risk in European populations, which should be further confirmed by better powered investigations and/or meta-analysis of our results with those of other studies. Although none of the SNPs reached the genome-wide statistical significance, it is still worth to include these relevant SNPs into the PRS approach for risk stratification. A risk stratification approach with high predictive power could be used to identify subgroups at particularly increased risk of pancreatic cancer, either in the general population or in groups that are already known to have an elevated risk, such as diabetics.

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/*Materials and Methods, Study Populations* section.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Commission of the Medical Faculty of the University of Heidelberg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DC and FC conceived and designed the study. YL performed the lab work. YL performed data curation and analysis. YL drafted the manuscript. DC, MGe, and FC reviewed and edited the manuscript. All other authors provided samples and data.

All authors critically read, commented and approved the final manuscript.

FUNDING

This work was supported by intramural funding of German Cancer Research Center (DKFZ); and by Fondazione Tizzi (www.fondazionetizzi.it); Fondazione Arpa (www.fondazionearpa.it); the Economic Development and Innovation Operative Programme Grant (Grant number GINOP-2.3.2-15-2016-00048); the Human Resources Development Operational Programme Grant (Grant number EFOP 3.6.2-16-2017-0006); Associazione Italiana Ricerca Cancro (Grant number 5x1000, 12182, and IG 17177); Fondazione Italiana Malattie Pancreas – Ministero Salute (Grant number FIMPCUP_J38D19000690001); Fondazione Cariverona: Oncology Biobank Project "Antonio Schiavi" (Grant number 203885/2017); the Ministry of Health of the Czech Republic (Grant number NV19-03-00097, FNOL-00098892); student grant from Palacky University (Grant number IGA_LF_2020_005). The ESTHER study was funded by the Baden-Württemberg State Ministry of Science, Research and Arts (Stuttgart, Germany).

ACKNOWLEDGMENTS

We would like to thank Mrs. Angelika Stein for help with the lab work. The research used the genotyping data provided by the EPIC, we would like to thank the contributors from the UK and the Netherlands. The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research. The authors would also thank all the participants of the PanScan GWAS Study and Pancreatic Cancer Case Control Association Study, and dbGaP for providing cancer genotyping dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.771312/full#supplementary-material

REFERENCES

1. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer Incidence and Mortality Patterns in Europe: Estimates for 40 Countries and 25 Major Cancers in 2018. *Eur J Cancer* (2018) 103:356–87. doi: 10.1016/j.ejca.2018.07.005
2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res* (2014) 74:2913–21. doi: 10.1158/0008-5472.CAN-14-0155
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
4. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JWW. Recent Trends of Cancer in Europe: A Combined Approach of Incidence, Survival and Mortality for 17 Cancer Sites Since the 1990s. *Eur J Cancer* (2008) 44:1345–89. doi: 10.1016/j.ejca.2007.12.015
5. Arnold M, Rutherford MJ, Bardot A, Ferlay J, Andersson TML, Myklebust TÅ, et al. Progress in Cancer Survival, Mortality, and Incidence in Seven High-Income Countries 1995–2014 (ICBP SURVMARK-2): A Population-Based Study. *Lancet Oncol* (2019) 20:1493–505. doi: 10.1016/S1470-2045(19)30456-5
6. Gentiluomo M, Canzian F, Nicolini A, Gemignani F, Landi S, Campa D. Germline Genetic Variability in Pancreatic Cancer Risk and Prognosis. *Semin Cancer Biol* (2020) S1044-579X(20)30174-7. doi: 10.1016/j.semcancer.2020.08.003
7. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-Wide Association Study Identifies Variants in the ABO Locus Associated With Susceptibility to Pancreatic Cancer. *Nat Genet* (2009) 41:986–90. doi: 10.1038/ng.429
8. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A Genome-Wide Association Study Identifies Pancreatic Cancer Susceptibility Loci on Chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet* (2010) 42:224–8. doi: 10.1038/ng.522
9. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common Variation at 2p13.3, 3q29, 7p13 and 17q25.1 Associated With Susceptibility to Pancreatic Cancer. *Nat Genet* (2015) 47:911–6. doi: 10.1038/ng.3341
10. Zhang M, Wang Z, Obazee O, Jia J, Childs EJ, Hoskins J, et al. Three New Pancreatic Cancer Susceptibility Signals Identified on Chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget* (2016) 7:66328–43. doi: 10.18632/oncotarget.11041
11. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, et al. Genome-Wide Association Study Identifies Multiple Susceptibility Loci for Pancreatic Cancer. *Nat Genet* (2014) 46:994–1000. doi: 10.1038/ng.3052
12. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-Wide Meta-Analysis Identifies Five New Susceptibility Loci for Pancreatic Cancer. *Nat Commun* (2018) 9:1–11. doi: 10.1038/s41467-018-02942-5
13. Campa D, Gentiluomo M, Obazee O, Ballerini A, Vodickova L, Hegyi P, et al. Genome-Wide Association Study Identifies an Early Onset Pancreatic Cancer Risk Locus. *Int J Cancer* (2020) 147:2065–74. doi: 10.1002/ijc.33004
14. Galeotti AA, Gentiluomo M, Rizzato C, Obazee O, Neoptolemos JP, Pasquali C, et al. Polygenic and Multifactorial Scores for Pancreatic Ductal Adenocarcinoma Risk Prediction. *J Med Genet* (2021) 58:369–77. doi: 10.1136/jmedgenet-2020-106961
15. Kim J, Yuan C, Babic A, Bao Y, Clish CB, Pollak MN, et al. Genetic and Circulating Biomarker Data Improve Risk Prediction for Pancreatic Cancer in the General Population. *Cancer Epidemiol Biomarkers Prev* (2020) 29:999–1008. doi: 10.1158/1055-9965.EPI-19-1389
16. Chen F, Childs EJ, Mocci E, Bracci P, Gallinger S, Li D, et al. Analysis of Heritability and Genetic Architecture of Pancreatic Cancer: A PANC4 Study. *Cancer Epidemiol Biomarkers Prev* (2019) 28:1238–45. doi: 10.1158/1055-9965.EPI-18-1235
17. Sud A, Kinnersley B, Houlston RS. Genome-Wide Association Studies of Cancer: Current Insights and Future Perspectives. *Nat Rev Cancer* (2017) 17:692–704. doi: 10.1038/nrc.2017.82
18. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic Statistical Analysis in Genetic Case-Control Studies. *Nat Protoc* (2011) 6:121–33. doi: 10.1038/nprot.2010.182
19. Dizier MH, Demenais F, Mathieu F. Gain of Power of the General Regression Model Compared to Cochran-Armitage Trend Tests: Simulation Study and Application to Bipolar Disorder. *BMC Genet* (2017) 18:24. doi: 10.1186/s12863-017-0486-6
20. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-Generation Genotype Imputation Service and Methods. *Nat Genet* (2016) 48:1284–7. doi: 10.1038/ng.3656

21. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A Reference Panel of 64,976 Haplotypes for Genotype Imputation. *Nat Genet* (2016) 48:1279–83. doi: 10.1038/ng.3643
22. Campa D, Rizzato C, Capurso G, Giese N, Funel N, Greenhalf W, et al. Genetic Susceptibility to Pancreatic Cancer and Its Functional Characterisation: The PANcreatic Disease ReseArch (PANDoRA) Consortium. *Dig Liver Dis* (2013) 45:95–9. doi: 10.1016/j.dld.2012.09.014
23. Riboli E, Hunt K, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation Into Cancer and Nutrition (EPIC): Study Populations and Data Collection. *Public Health Nutr* (2002) 5:1113–24. doi: 10.1079/phn2002394
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* (2007) 81:559–75. doi: 10.1086/519795
25. Graham JW. Missing Data Analysis: Making It Work in the Real World. *Annu Rev Psychol* (2009) 60:549–76. doi: 10.1146/annurev.psych.58.110405.085530
26. van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* (2011) 45:1–67. doi: 10.18637/jss.v045.i03
27. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PloS Comput Biol* (2015) 11:1–19. doi: 10.1371/journal.pcbi.1004219
28. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) Project. *Nat Genet* (2013) 45:580–5. doi: 10.1038/ng.2653
29. Ward LD, Kellis M. HaploReg V4: Systematic Mining of Putative Causal Variants, Cell Types, Regulators and Target Genes for Human Complex Traits and Disease. *Nucleic Acids Res* (2016) 44:D877–81. doi: 10.1093/nar/gkv1340
30. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of Functional Variation in Personal Genomes Using RegulomeDB. *Genome Res* (2012) 22:1790–7. doi: 10.1101/gr.137323.112
31. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res* (2019) 47:W556–60. doi: 10.1093/nar/gkz430
32. Kim K, Jang I, Kim M, Choi J, Kim MS, Lee B, et al. 3DIV Update for 2021: A Comprehensive Resource of 3D Genome and 3D Cancer Genome. *Nucleic Acids Res* (2021) 49:D38–46. doi: 10.1093/nar/gkaa1078
33. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: A Web Server for Functional Annotation of Human Genome Sequence Variation (2020 Update). *Nucleic Acids Res* (2020) 48:W185–92. doi: 10.1093/NAR/GKAA420
34. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: Systematic Identification of Trait-Associated Genes Using Large-Scale Genetics and Functional Genomics. *Nucleic Acids Res* (2021) 49:D1311–20. doi: 10.1093/nar/gkaa840
35. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional Mapping and Annotation of Genetic Associations With FUMA. *Nat Commun* (2017) 8:1826. doi: 10.1038/s41467-017-01261-5
36. Grarup N, Moltke I, Andersen MK, Bjerregaard P, Larsen CVL, Dahl-Petersen IK, et al. Identification of Novel High-Impact Recessively Inherited Type 2 Diabetes Risk Variants in the Greenlandic Population. *Diabetologia* (2018) 61:2005–15. doi: 10.1007/s00125-018-4659-2
37. Zhu D, Yin J, Liang C, Luo X, Lv D, Dai Z, et al. CACNA1C (Rs1006737) may be a Susceptibility Gene for Schizophrenia: An Updated Meta-Analysis. *Brain Behav* (2019) 9:e01292. doi: 10.1002/brb3.1292
38. Hebbar P, Nizam R, Melhem M, Alkayal F, Elkum N, Elsa John S, et al. Genome-Wide Association Study Identifies Novel Recessive Genetic Variants for High TGs in an Arab Population. *J Lipid Res* (2018) 59:1951–96. doi: 10.1194/jlr.P080218
39. Power RA, Keller MC, Ripke S, Abdellaoui A, Wray NR, Sullivan PF, et al. A Recessive Genetic Model and Runs of Homozygosity in Major Depressive Disorder. *Am J Med Genet Part B Neuropsychiatr Genet* (2014) 165:157–66. doi: 10.1002/ajmg.b.32217
40. Camargo M, Rivera D, Moreno L, Lidral AC, Harper U, Jones M, et al. GWAS Reveals New Recessive Loci Associated With Non-Syndromic Facial Clefting. *Eur J Med Genet* (2012) 55:510–4. doi: 10.1016/j.ejmg.2012.06.005
41. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared Genetic Origin of Asthma, Hay Fever and Eczema Elucidates Allergic Disease Biology. *Nat Genet* (2017) 49:1752–7. doi: 10.1038/ng.3985
42. Midha S, Chawla S, Garg PK. Modifiable and Non-Modifiable Risk Factors for Pancreatic Cancer: A Review. *Cancer Lett* (2016) 381:269–77. doi: 10.1016/j.canlet.2016.07.022
43. Li D, Duell EJ, Yu K, Risch HA, Olson SH, Kooperberg C, et al. Pathway Analysis of Genome-Wide Association Study Data Highlights Pancreatic Development Genes as Susceptibility Factors for Pancreatic Cancer. *Carcinogenesis* (2012) 33:1384–90. doi: 10.1093/carcin/bgs151
44. Piao J, Zhu L, Sun J, Li N, Dong B, Yang Y, et al. High Expression of CDK1 and BUB1 Predicts Poor Prognosis of Pancreatic Ductal Adenocarcinoma. *Gene* (2019) 701:15–22. doi: 10.1016/j.gene.2019.02.081
45. Katoh Y, Katoh M. Identification and Characterization of ARHGAP27 Gene in Silico. *Int J Mol Med* (2004) 14:943–7. doi: 10.3892/ijmm.14.5.943
46. Huang Y, Ying K, Xie Y, Zhou Z, Wang W, Tang R, et al. Cloning and Characterization of a Novel Human Leptin Receptor Overlapping Transcript-Like 1 Gene (LEPROTL1). *Biochim Biophys Acta Gene Struct Expr* (2001) 1517:327–31. doi: 10.1016/S0167-4781(00)00266-9
47. Touvier T, Conte-Auriol F, Briand O, Cudejko C, Paumelle R, Caron S, et al. LEPROT and LEPROTL1 Cooperatively Decrease Hepatic Growth Hormone Action in Mice. *J Clin Invest* (2009) 119:3830–8. doi: 10.1172/JCI34997
48. Aghdassi AA, John DS, Sendler M, Storck C, van den Brandt C, Krüger B, et al. Absence of the Neutrophil Serine Protease Cathepsin G Decreases Neutrophil Granulocyte Infiltration But Does Not Change the Severity of Acute Pancreatitis. *Sci Rep* (2019) 9:16774. doi: 10.1038/s41598-019-53293-0
49. Zhang YD, Hurson AN, Zhang H, Choudhury PP, Easton DF, Milne RL, et al. Assessment of Polygenic Architecture and Risk Prediction Based on Common Variants Across Fourteen Cancers. *Nat Commun* (2020) 11:1–13. doi: 10.1038/s41467-020-16483-3
50. Torkamani A, Wineinger NE, Topol EJ. The Personal and Clinical Utility of Polygenic Risk Scores. *Nat Rev Genet* (2018) 19:581–90. doi: 10.1038/s41576-018-0018-x
51. Klein AP, Lindström S, Mendelsohn JB, Steplowski E, Arslan AA, Bueno-de-Mesquita HB, et al. An Absolute Risk Model to Identify Individuals at Elevated Risk for Pancreatic Cancer in the General Population. *PloS One* (2013) 8:e72311. doi: 10.1371/journal.pone.0072311

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.

Citation: *Lu Y, Gentiluomo M, Macauda A, Gioffreda D, Gazouli M, Petrone MC, Kelemen D, Ginocchi L, Morelli L, Papiris K, Greenhalf W, Izbicki JR, Kiudelis V, Mohelníková-Duchoňová B, Bueno-de-Mesquita B, Vodicka P, Brenner H, Diener MK, Pezzilli R, Ivanauskas A, Salvia R, Szentesi A, Aoki MN, Németh BC, Sperti C, Jamroziak K, Chammas R, Oliverius M, Archibugi L, Ermini S, Novák J, Kupcinskas J, Strouhal O, Souček P, Cavestro GM, Milanetto AC, Vanella G, Neoptolemos JP, Theodoropoulos GE, van Laarhoven HWM, Mambrini A, Moz S, Kala Z, Loveček M, Basso D, Uzunoglu FG, Hackert T, Testoni SGG, Hlaváč V, Andriulli A, Lucchesi M, Tavano F, Carrara S, Hegyi P, Arcidiacono PG, Busch OR, Lawlor RT, Puzzono M, Boggi U, Guo F, Małecka-Panas E, Capurso G, Landi S, Talar-Wojnarowska R, Strobel O, Gao X, Vashist Y, Campa D and Canzian F (2021) Identification of Recessively Inherited Genetic Variants Potentially Linked to Pancreatic Cancer Risk.*
Front. Oncol. 11:771312. doi: 10.3389/fonc.2021.771312

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fonc.2021.796774

The Association of *MEG3* Gene rs7158663 Polymorphism With Cancer Susceptibility

Xueren Gao[1], Xianyang Li[1], Shulong Zhang[2] and Xiaoting Wang[3]*

[1] School of Pharmacy, Yancheng Teachers' University, Yancheng, China, [2] Department of General Surgery, Xuhui District Central Hospital, Shanghai, China, [3] Physical Examination Centre, Xuhui District Central Hospital, Shanghai, China

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Macrina Beatriz Silva Cázares, Autonomous University of San Luis Potosí, Mexico
Qing Chun Zhao, Shenyang Pharmaceutical University, China

***Correspondence:**
Xueren Gao
gaoxr@yctu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *17 October 2021*
Accepted: *23 November 2021*
Published: *09 December 2021*

Citation:
Gao X, Li X, Zhang S and Wang X (2021) The Association of MEG3 Gene rs7158663 Polymorphism With Cancer Susceptibility. Front. Oncol. 11:796774. doi: 10.3389/fonc.2021.796774

Although the association of *MEG3* gene rs7158663 polymorphism with cancer susceptibility has been investigated, the findings are inconsistent. The aim of this study was to analyze the association between the rs7158663 polymorphism and cancer susceptibility through a case-control study and meta-analysis. In a case-control study with 430 colorectal cancer (CRC) cases and 445 healthy controls, the rs7158663 polymorphism was genotyped by direct sequencing. STATA software was used to calculate the pooled odds ratio and 95% confidence interval in a meta-analysis including 4,649 cancer cases and 5,590 controls. Both the case-control study and meta-analysis showed that the rs7158663 polymorphism was associated with increased susceptibility to CRC. Individuals carrying the AA or GA genotype were more likely to develop CRC than those carrying the rs7158663 GG genotype. Interestingly, MEG3 expression was significantly lower in colorectal tissues of the AA or GA genotype compared to those of the rs7158663 GG genotype. In addition, the meta-analysis suggested that the rs7158663 polymorphism was also associated with increased susceptibility to breast cancer and gastric cancer. Bioinformatics analysis showed that the rs7158663 A allele contributed to the binding of hsa-miR-4307 and hsa-miR-1265 to MEG3. In conclusion, the current findings suggest that the *MEG3* gene rs7158663 polymorphism may serve as a genetic marker for predicting the risk of cancers, such as breast cancer, gastric cancer and CRC. However, the sample size of the current study is still insufficient, especially in the subgroup analysis. Therefore large and well-designed studies are needed to validate our findings.

Keywords: MEG3, polymorphism, rs7158663, cancer, susceptibility

INTRODUCTION

Cancer is one of the most serious public health issues in the world, with approximately 18.1 million new cancer diagnoses and 9.6 million cancer deaths in 2018 (1). Although the precise processes of cancer development and progression are still largely unclear, a growing body of research suggests that genetic predisposition has a substantial influence on the likelihood of individual cancer development (2, 3).

Long non-coding RNA (lncRNA) is a form of RNA transcript that is longer than 200 nucleotides but does not transcribe into protein in cells. LncRNAs play a role in a variety of cell activities, such as cell proliferation, migration, invasion, and angiogenesis, and their dysregulation has been linked to a variety of cancers (4–7). Maternally expressed 3 (MEG3) is one of the most well-studied lncRNAs and is expressed in multiple organs, such as the liver, brain, pancreas, stomach and ovary. However, MEG3 expression is typically suppressed in a variety of cancer tissues (8). Functional studies showed that this lncRNA could regulate the expression of various tumor suppressor genes and oncogenes (9–12). Zhu et al. found that ectopic expression of MEG3 could significantly inhibit proliferation and induce apoptosis in hepatoma cells. MEG3 could function as a tumor suppressor in hepatoma cells through interacting with p53 protein to activate p53-mediated transcriptional activity and influence the expression of partial p53 target genes (9). Dong et al. found that downregulation of MEG3 expression could promote proliferation, migration, and invasion of hepatocellular carcinoma cells by upregulating TGF-β1 expression (10). Zuo et al. demonstrated that MEG3 activated by vitamin D could inhibit glycolysis in colorectal cancer (CRC) *via* promoting c-Myc degradation (11). Xu et al. found that MEG3 could mediate the miR-149-3p/FOXP3 axis by reducing p53 ubiquitination to exert a suppressive effect on regulatory T cell differentiation and immune escape in esophageal cancer (12). In addition, certain polymorphisms (rs3087918 T>G, rs11160608 A>C, rs4081134 G>A, rs7158663 A>G) within the *MEG3* gene are implicated in cancer susceptibility (13–16). For example, *MEG3* gene rs3087918 was associated with a decreased risk of breast cancer in a Chinese population (14). *MEG3* gene rs11160608 was related to an increased risk of oral squamous cell carcinoma in a Chinese Han population (15). *MEG3* gene rs4081134 was significantly associated with a decreased risk of lung cancer in a Northeast Chinese population (16). *MEG3* gene rs7158663 is the most interesting polymorphic locus located on the MEG3 transcript. Bioinformatic analysis showed that the rs7158663 polymorphism had the potential to change the local RNA folding structure and affect miRNA-lncRNA interactions, which in turn affected the expression level of miRNA and/or MEG3 (17, 18). Several studies have explored the relationship between this potentially functional polymorphism and cancer susceptibility, but the results are inconsistent and need to be further clarified.

In the current study, we first explored the relationship between the rs7158663 polymorphism and CRC susceptibility using a case-control study, and then analyzed its effect on MEG3 expression in colorectal tissues. In addition, a meta-analysis was conducted to systematically evaluated the relationship between this polymorphism and cancer susceptibility, which would help us to better understand the role of the rs7158663 polymorphism in cancer susceptibility.

MATERIALS AND METHODS

Sample Collection

Peripheral blood of 430 CRC patients and 445 healthy controls were collected from Shanghai Xuhui District Central Hospital. All participants were genetically unrelated Han Chinese. Diagnosis of CRC patients was histopathologically confirmed. Healthy controls were cancer-free individuals living in the same residential area and seeking routine physical exams. Furthermore, colorectal tissues were obtained from 40 CRC surgery patients who had not received radiochemotherapy before surgery. Written informed consent was obtained from all participants. The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of Shanghai Xuhui District Central Hospital.

Genotyping

TIANamp genomic DNA Kit (Tiangen) was used to isolate genomic DNA from peripheral blood according to the manufacturer's instructions. Genomic DNA concentration was detected using a NanoDrop spectrophotometer. Direct sequencing was used to detect the genotype of the rs7158663 locus in each individual.

Real-Time Quantitative PCR

Total RNA was isolated from colorectal tissues using the RNAsimple total RNA kit (Tiangen) according to the manufacturer's instructions. ReverTra Ace qPCR RT Kit (TOYOBO) was used to synthesize cDNA. FastStart Universal SYBR Green Master (Roche) was used to conduct real-time quantitative PCR. MEG3 expression was normalized to the internal control GAPDH. The specific primer sequences are presented in **Table S1**.

Bioinformatic Analysis

The lncRNASNP online tool (http://bioinfo.life.hust.edu.cn/lncRNASNP) was used to analyze whether the rs7158663 polymorphism affects miRNA binding (19).

Statistical Analysis

Hardy-Weinberg equilibrium (HWE) for the control group was tested by a goodness-of-fit $\chi 2$ test. The association of *MEG3* gene rs7158663 polymorphism with CRC susceptibility was evaluated using adjusted odds ratios (ORs) with their 95% confidence intervals (CIs). Student's t-test was used to check the differences for age variable between CRC cases and controls. $\chi 2$ test was used to assess the differences in gender variable between CRC cases and controls. The normalized expression levels of MEG3 among

different genotypes were compared using one-way ANOVA. All statistical analyses were performed by SAS 9.4 (SAS Institute, Cary, USA). $P < 0.05$ was defined as the level of significance.

Meta-Analysis

PubMed, CNKI and EMBASE databases were searched based on the following keywords: "Maternally expressed 3 or MEG3", "polymorphism or variant" and "cancer or carcinoma or malignancy". The last literature search was conducted on October 7, 2021. The primary inclusion criterion for previous studies was to have sufficient genotype data. If numerous studies had overlapping or duplicate data, only studies with complete data were included. Data from the included studies were extracted independently by two investigators. Disagreements were settled by conversation. The pooled ORs and their 95% CIs were applied to determine the relationship of the rs7158663 polymorphism with cancer susceptibility. The between-study heterogeneity was assessed using Chi-square-based statistic I^2 test and Cochran's Q-test. When $I^2 > 50\%$ or $P_H < 0.1$, we used the random-effects model to estimate the pooled OR. Otherwise, the fixed-effects model was applied. To assess the quality and consistency of the results, sensitivity analysis was undertaken by removing each study in turn. Begg's and Egger's tests were used to assess potential publication bias. Trial sequential analysis (TSA) was conducted in a selected genetic model to assess the statistical reliability of the meta-analysis. TSA was conducted with a 5% risk of type I error and a 20% risk of type II error (20). The statistical analyses were performed by STATA 12.0 (Stata Corporation, College Station, TX, USA).

RESULTS

The results of the case-control study are shown in **Table 1**. There was no statistical difference in the age and gender distribution between the case and control groups (P>0.05). The genotype frequency distribution of the control group was consistent with HWE (P_{HWE}=0.43). There was a significant association between *MEG3* gene rs7158663 polymorphism and CRC susceptibility [GA vs. GG: OR=1.48, 95%CI= 1.11-1.96, P=0.007; AA vs. GG: OR=1.83, 95%CI=1.11-3.03, P=0.018; (GA+AA) vs. GG: OR=1.53, 95%CI=1.17-2.00, P=0.002; A vs. G: OR=1.41, 95% CI=1.14-1.74, P=0.001].

Genotype-tissue expression results showed that the rs7158663 polymorphism was significantly associated with the expression of MEG3 in colorectal tissues. MEG3 expression was significantly lower in colorectal tissues of the AA or GA genotype compared to those of the rs7158663 GG genotype (**Figure 1**). The results of bioinformatics analysis showed that the rs7158663 A allele contributed to the binding of hsa-miR-4307 and hsa-miR-1265 to MEG3 (**Figure 2**).

Based on database searches, a total of 11 case-control studies exploring the association of the rs7158663 polymorphism with cancer susceptibility were included in the current meta-analysis (**Figure S1**). The relevant studies were published between 2016 and 2021. The current meta-analysis combined our results included 4,649 cancer cases and 5,590 controls (**Table 2**). The overall combined analysis showed that the rs7158663 polymorphism was not associated with cancer susceptibility (**Table 3**). However, the country-based stratified analysis showed that the rs7158663 polymorphism was associated with cancer susceptibility in the Chinese population under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models, and in the Egyptians under (AA+AG) vs. GG model. The stratified analysis based on cancer type showed that the rs7158663 polymorphism was associated with susceptibility to breast cancer under (AA+AG) vs. GG and AG vs. GG models, and gastric cancer under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models, and colorectal cancer under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models.

TABLE 1 | Characteristics of age, gender and rs7158663 polymorphism in cases and controls.

Variables	Case (%) (N = 430)	Controls (%) (N = 445)	[a]OR (95% CI)	[a]P value
Age, mean ± SD	57.7 ± 5.9	57.7 ± 6.1		0.86
Gender				
Male	251 (58.4)	249 (56.0)		0.47
Female	179 (41.6)	196 (44.0)		
Genotype				
GG	202 (47.0)	256 (57.5)	Reference	
GA	185 (43.0)	159 (35.7)	1.48 (1.11-1.96)	0.007
AA	43 (10.0)	30 (6.7)	1.83 (1.11-3.03)	0.018
P_{trend}				0.002
P_{HWE}				0.43
GG	202 (47.0)	256 (57.5)	Reference	
GA+AA	228 (53.0)	189 (42.5)	1.53 (1.17-2.00)	0.002
GG+GA	387 (90.0)	415 (93.3)	Reference	
AA	43 (10.0)	30 (6.7)	1.55 (0.95-2.52)	0.08
Allele				
G	589 (68.5)	671 (75.4)	Reference	
A	271 (31.5)	219 (24.6)	1.41 (1.14-1.74)	0.001

[a]*Adjusted for age and gender when appropriate.*

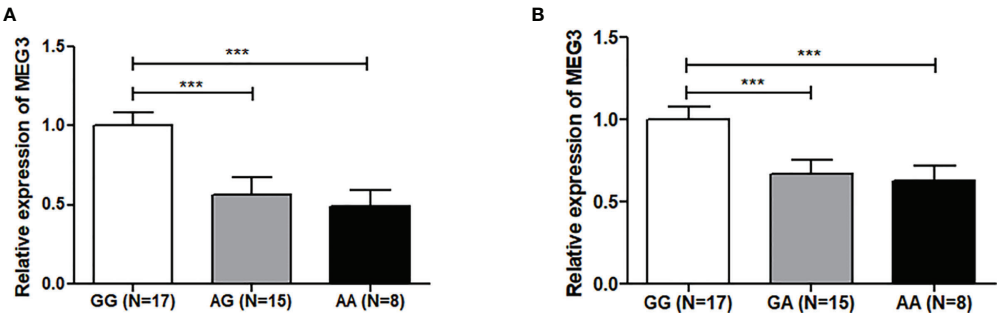


FIGURE 1 | Relationship between rs7158663 genotype and MEG3 expression in CRC tissues **(A)** and normal paracancerous tissues **(B)**. ***$p < 0.001$.

miRNA	hsa-miR-4307	
SNP	rs7158663:100853087(A/G)	
lncRNA	NONHSAT039747.2	
TargetScan	**start**	**end**
	100853082	100853088
miRanda	**start**	**end**
	100853070	100853089
	Score: 156 **Energy:** -10.1kCal/mol **Strand:** +	
	Q:2 to 18 R:9 to 28 Align Len (17) (76.47%) (82.35%)	
	miRNA: 3' ccUUUGUC-CUUUUUUGUAa 5' |: ||| || ||||||| lncRNA:5' aaAGCCAGAGATAAAACATc 3'	

miRNA	hsa-miR-1265	
SNP	rs7158663:100853087(A/G)	
lncRNA	NONHSAT039747.2	
TargetScan	**start**	**end**
	100853085	100853091
miRanda	**start**	**end**
	100853071	100853092
	Score: 140 **Energy:** -12.08kCal/mol **Strand:** +	
	Q:2 to 9 R:10 to 31 Align Len (7) (100.00%) (100.00%)	
	miRNA: 3' uuguugugaacuggUGUAGGAc 5' ||||||| lncRNA:5' aagccagagataaaACATCCTt 3'	

FIGURE 2 | lncRNASNP-based analysis of the effect of rs7158663 polymorphism on miRNA binding.

TABLE 2 | Main characteristics of the case-control studies in the current meta-analysis.

Authors	Year of publication	Country	Cancer type	Genotyping method	Number of cases				Number of controls				P_{HWE}
					GG	GA	AA	Total	GG	GA	AA	Total	
Gao et al.*	2021	China	Colorectal cancer	TaqMan	202	185	43	430	256	159	30	445	0.43
Shaker et al. (21)	2021	Egypt	Breast cancer	TaqMan	63	117		180	93	57		150	–
Kong et al. (22)	2020	China	Gastric cancer	TaqMan	215	198	61	474	290	203	50	543	0.1
Xu et al. (23)	2020	China	Prostate cancer	TaqMan	98	54	13	165	111	78	11	200	0.57
Zheng et al. (14)	2020	China	Breast cancer	MassArray	224	170	33	427	403	250	47	700	0.33
Ali et al. (24)	2020	Egypt	Breast cancer	TaqMan	57	63	30	150	84	63	7	154	0.26
Mazraeh et al. (25)	2020	Iran	Acute myeloid leukemia	PCR-based restriction fragment length polymorphism	43	36	21	100	16	48	36	100	1
Wei (26)	2019	China	Liver cancer	Taqman	717	349	51	1117	795	391	62	1248	0.13
Yang et al. (16)	2018	China	Lung cancer	Taqman	268	219	39	526	289	204	33	526	0.71
Zhuo et al. (27)	2018	China	Neuroblastoma	TaqMan	233	141	18	392	433	296	54	783	0.72
Zhang et al. (28)	2018	China	Gastric cancer	TaqMan	83	74	15	172	138	76	10	224	0.91
Cao et al. (29)	2016	China	Colorectal cancer	TaqMan	264	200	52	516	298	188	31	517	0.85

Current study.

TABLE 3 | Meta-analysis of the association between the rs7158663 polymorphism and cancer risk.

Comparison	*Subgroup	Heterogeneity		Effect model	OR[95%CI]	P
		P_H	I^2			
(AA+AG) vs. GG	Overall	<0.00001	83%	Random	1.22[0.99,1.49]	0.06
	China	0.003	65%	Random	1.18[1.02,1.37]	0.03
	Egypt	0.18	44%	Fixed	2.44[1.77,3.37]	<0.00001
	Breast cancer	0.001	85%	Random	1.89[1.08,3.30]	0.02
	Gastric cancer	0.36	0%	Fixed	1.47[1.19,1.81]	0.0004
	Colorectal cancer	0.38	0%	Fixed	1.40[1.17,1.68]	0.0003
AA vs. (AG+GG)	Overall	0.0003	70%	Random	1.27[0.95,1.70]	0.11
	China	0.11	39%	Fixed	1.23[1.05,1.44]	0.01
	Breast cancer	0.002	89%	Random	2.36[0.54,10.42]	0.26
	Gastric cancer	0.47	0%	Fixed	1.55[1.09,2.22]	0.02
	Colorectal cancer	0.7	0%	Fixed	1.65[1.18,2.31]	0.003
AA vs. GG	Overall	<0.00001	80%	Random	1.31[0.90,1.90]	0.16
	China	0.02	55%	Random	1.33[1.03,1.73]	0.03
	Breast cancer	0.002	90%	Random	2.70[0.55,13.13]	0.22
	Gastric cancer	0.39	0%	Fixed	1.78[1.23,2.58]	0.002
	Colorectal cancer	0.91	0%	Fixed	1.86[1.32,2.62]	0.0004
AG vs. GG	Overall	0.0004	69%	Random	1.11[0.94,1.31]	0.21
	China	0.04	51%	Random	1.15[1.01,1.31]	0.04
	Breast cancer	0.51	0%	Fixed	1.27[1.02,1.60]	0.04
	Gastric cancer	0.41	0%	Fixed	1.39[1.12,1.74]	0.003
	Colorectal cancer	0.29	10%	Fixed	1.32[1.09,1.60]	0.004
A vs. G	Overall	<0.00001	84%	Random	1.14[0.96,1.34]	0.14
	China	0.0007	70%	Random	1.16[1.02,1.32]	0.02
	Breast cancer	0.004	88%	Random	1.53[0.87,2.69]	0.14
	Gastric cancer	0.32	0%	Fixed	1.38[1.17,1.63]	0.0001
	Colorectal cancer	0.61	0%	Fixed	1.36[1.17,1.56]	<0.0001

**Two and more studies were combined for analysis.*

The sensitivity analysis showed that after Mazraeh's study was removed, the overall combined results were significantly altered under all comparison models (**Figure S2**). After Zhuo's study was removed, the overall combined results were significantly altered under (AA+AG) vs. GG, and AA vs. (AG+GG) models. After Xu's study was removed, the overall combined results were significantly altered under (AA+AG) vs. GG model. Begg's and Egger's tests showed no publication bias in the current meta-analysis (**Table 4**). TSA was conducted in the (AA+AG) vs. GG model. The result showed that the cumulative Z-curve (blue line) has crossed the required information sizes (n=8,329) (**Figure 3**), which indicated that the cumulative evidence was adequate in the overall analysis.

DISCUSSION

Recent studies have shown that certain genetic variants on lncRNA genes may be associated with cancer risk (30–32). These genetic variants contain the rs7158663 polymorphism on the *MEG3* gene. For instance, Cao et al. found that rs7158663 AA genotype had significantly higher CRC risk than GG genotype, which was consistent with our results. The further stratified analysis revealed that the elevated risk was strongly associated with people with age ≤ 60 and a family history of cancer. However, there was no link found between the rs7158663 polymorphism and CRC site or stage (29). Both Zhang et al. and Kong et al. found that individuals carrying the rs7158663 AG+AA genotype or A allele had a significantly increased risk of gastric cancer (22, 28). However, some studies suggested that the rs7158663 polymorphism was not associated with cancer risk. For instance, Wei found that the rs7158663 polymorphism was not associated with hepatocarcinogenesis (26). Yang et al. found that the rs7158663 polymorphism was not associated with susceptibility to lung cancer (16). Zhuo et al. found that the rs7158663 polymorphism was not linked with neuroblastoma susceptibility, regardless of whether it was corrected for age and gender (27). These inconsistent results forced us to clarify the relationship between the rs7158663 polymorphism and cancer susceptibility by meta-analysis. By combining two and more studies for analysis, we found that the rs7158663 polymorphism was not associated with overall cancer susceptibility. However, the country-based stratified analysis showed that the rs7158663 polymorphism was associated with cancer susceptibility in the Chinese population under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models, and in the Egyptians under (AA+AG) vs. GG model. The stratified analysis based on

TABLE 4 | Publication bias analysis of included studies.

Comparison	P value	
	Begg's test	Egger's test
(AA+AG) vs. GG	1	0.58
AA vs. (AG+GG)	0.53	0.57
AA vs. GG	0.64	0.39
AG vs. GG	0.76	0.60
A vs. G	0.64	0.39

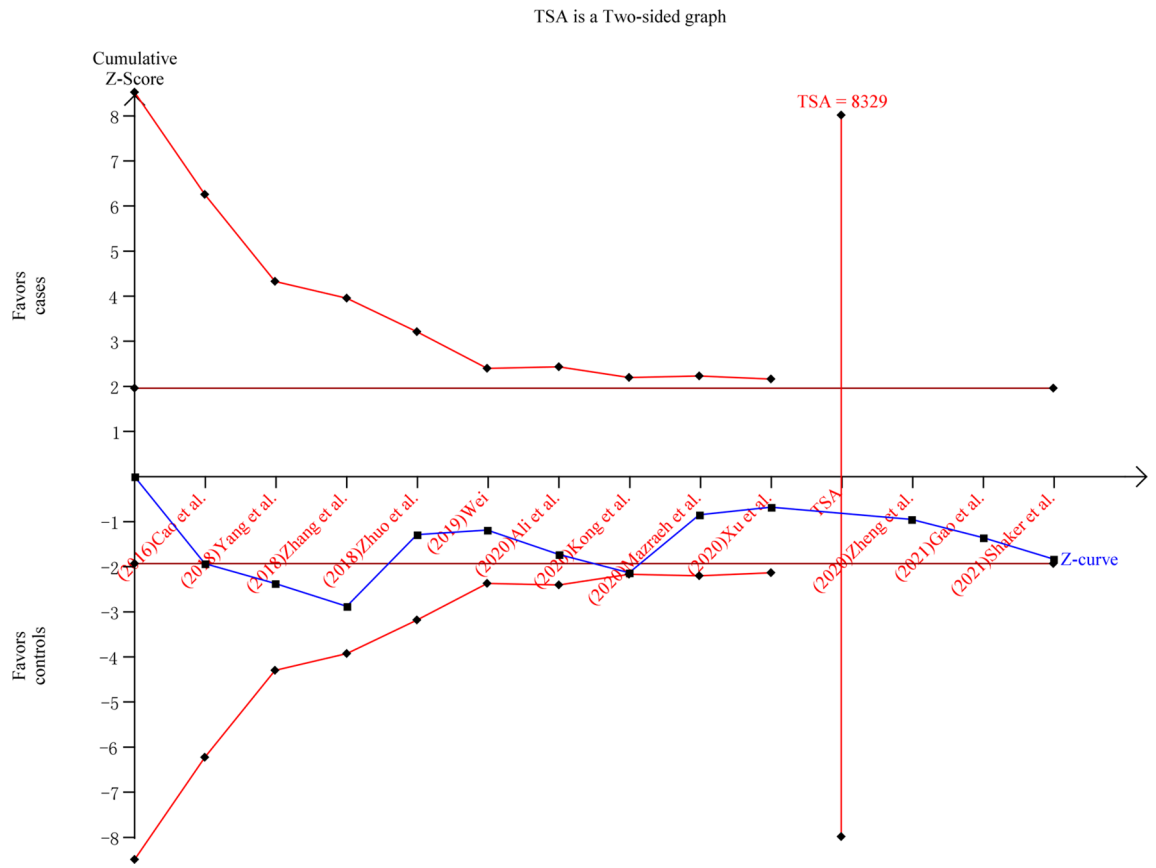


FIGURE 3 | Trial sequential analysis of the current meta-analysis under the (AA+AG) vs. GG model.

cancer type showed that the rs7158663 polymorphism was associated with susceptibility to breast cancer under (AA+AG) vs. GG and AG vs. GG models, and gastric cancer under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models, and colorectal cancer under (AA+AG) vs. GG, AA vs. (AG+GG), AA vs. GG, AG vs. GG, and A vs. G models. There was no publication bias in the current meta-analysis, and TSA suggested that the sample size in the overall combined analysis was adequate. However, the sensitivity analysis results suggested that the current meta-analysis results were not sufficiently stable. Therefore we needed more studies to confirm the current findings.

MEG3 could inhibit the malignant phenotype of many cancers including gastric, breast and colorectal cancers (33–35). The current study found that the rs7158663 polymorphism could affect MEG3 expression in colorectal tissues. MEG3 expression was significantly lower in colorectal tissues of the AA or GA genotype compared to those of the rs7158663 GG genotype. Bioinformatics analysis showed that the rs7158663 A allele contributed to the binding of hsa-miR-4307 and hsa-miR-1265 to MEG3. Therefore, we speculated that the rs7158663 polymorphism may affect an individual's susceptibility to CRC by influencing the regulation of MEG3 expression by miRNAs.

Although the current study has yielded some meaningful results, some shortcomings needed to be pointed out. Due to the insufficient sample size of the current case-control study and the unavailability of some clinical data, we did not further analyze the relationship between the rs7158663 polymorphism and the clinicopathological features. In addition, we did not consider the effect of the rs7158663 polymorphism interaction with environmental factors on cancer susceptibility.

CONCLUSIONS

The current study results suggest that the *MEG3* gene rs7158663 polymorphism is associated with susceptibility to a variety of cancers, such as breast cancer, gastric cancer and CRC. However, large and well-designed studies are still needed to validate our findings.

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 Shanghai Xuhui District Central Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XG carried out the molecular genetic studies, did the literature search and the statistical analysis, and wrote the paper. XL performed biochemistry tests. XW were responsible for the acquisition of data. SZ participated in study design and coordination and helped to draft the manuscript. XG and SZ interpreted the data and were responsible for the manuscript preparation. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We gratefully acknowledge the cooperation of the patients and normal control individuals involved in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.796774/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Gao X, Wang X. HMGA2 Rs968697 T > C Polymorphism Is Associated With the Risk of Colorectal Cancer. *Nucleosides Nucleotides Nucleic Acids* (2021) 40 (8):821–8. doi: 10.1080/15257770.2021.1952596
3. Ni W, Wang X, Sun Y, Gao X. Meta-Analysis of the Association Between MALAT1 Rs619586 a>G Polymorphism and Cancer Risk. *J Int Med Res* (2020) 48:300060520941969. doi: 10.1177/0300060520941969
4. Qi ZY, Wang LL, Qu XL. lncRNA LINC00355 Acts as a Novel Biomarker and Promotes Glioma Biological Activities *via* the Regulation of miR-1225/FNDC3B. *Dis Markers* (2021) 2021:1683129. doi: 10.1155/2021/1683129
5. Feng Y, Wei G, Zhang L, Zhou H, Wang W, Guo P, et al. LncRNA DARS-AS1 Aggravates the Growth and Metastasis of Hepatocellular Carcinoma via Regulating the miR-3200-5p-Cytoskeleton Associated Protein 2 (CKAP2) Axis. *Bioengineered* (2021) 12(1):8217–32. doi: 10.1080/21655979.2021.1982272
6. Fu C, Xin J, Zhang W, Lai J, Huang Z. LINC00992 Exerts Oncogenic Activities in Prostate Cancer via Regulation of SOX4. *Exp Cell Res* (2021) 408(1):112855. doi: 10.1016/j.yexcr.2021.112855
7. Xu L, Liu C, Ye Z, Wu C, Ding Y, Huang J. Overexpressed LINC00467 Promotes the Viability and Proliferation Yet Inhibits Apoptosis of Gastric Cancer Cells *via* Raising ITGB3 Level. *Tissue Cell* (2021) 73:101644. doi: 10.1016/j.tice.2021.101644
8. Ghafouri-Fard S, Taheri M. Maternally Expressed Gene 3 (MEG3): A Tumor Suppressor Long Non Coding RNA. *BioMed Pharmacother* (2019) 118:109129. doi: 10.1016/j.biopha.2019.109129
9. Zhu J, Liu S, Ye F, Shen Y, Tie Y, Zhu J, et al. Long Noncoding RNA MEG3 Interacts With P53 Protein and Regulates Partial P53 Target Genes in Hepatoma Cells. *PloS One* (2015) 10(10):e0139790. doi: 10.1371/journal.pone.0139790
10. Dong H, Zhang Y, Xu Y, Ma R, Liu L, Luo C, et al. Downregulation of Long Non-Coding RNA MEG3 Promotes Proliferation, Migration, and Invasion of Human Hepatocellular Carcinoma Cells by Upregulating TGF-β1. *Acta Biochim Biophys Sin (Shanghai)* (2019) 51(6):645–52. doi: 10.1093/abbs/gmz046
11. Zuo S, Wu L, Wang Y, Yuan X. Long Non-Coding RNA MEG3 Activated by Vitamin D Suppresses Glycolysis in Colorectal Cancer via Promoting C-Myc Degradation. *Front Oncol* (2020) 10:274. doi: 10.3389/fonc.2020.00274
12. Xu QR, Tang J, Liao HY, Yu BT, He XY, Zheng YZ, et al. Long Non-Coding RNA MEG3 Mediates the miR-149-3p/FOXP3 Axis by Reducing P53 Ubiquitination to Exert a Suppressive Effect on Regulatory T Cell Differentiation and Immune Escape in Esophageal Cancer. *J Transl Med* (2021) 19(1):264. doi: 10.1186/s12967-021-02907-1
13. Dong X, Gao W, Lv X, Wang Y, Wu Q, Yang Z, et al. Association Between lncRNA GAS5, MEG3, and PCAT-1 Polymorphisms and Cancer Risk: A Meta-Analysis. *Dis Markers* (2020) 2020:6723487. doi: 10.1155/2020/6723487
14. Zheng Y, Wang M, Wang S, Xu P, Deng Y, Lin S, et al. LncRNA MEG3 Rs3087918 was Associated With a Decreased Breast Cancer Risk in a Chinese Population: A Case-Control Study. *BMC Cancer* (2020) 20(1):659. doi: 10.1186/s12885-020-07145-0
15. Hou Y, Zhang B, Miao L, Ji Y, Yu Y, Zhu L, et al. Association of Long Non-Coding RNA MEG3 Polymorphisms With Oral Squamous Cell Carcinoma Risk. *Oral Dis* (2019) 25(5):1318–24. doi: 10.1111/odi.13103
16. Yang Z, Li H, Li J, Lv X, Gao M, Bi Y, et al. Association Between Long Noncoding RNA MEG3 Polymorphisms and Lung Cancer Susceptibility in Chinese Northeast Population. *DNA Cell Biol* (2018) 37(10):812–20. doi: 10.1089/dna.2018.4277
17. Cao X, Zhuang S, Hu Y, Xi L, Deng L, Sheng H, et al. Associations Between Polymorphisms of Long Non-Coding RNA MEG3 and Risk of Colorectal Cancer in Chinese. *Oncotarget* (2016) 7(14):19054–9. doi: 10.18632/oncotarget.7764
18. Ghaedi H, Zare A, Omrani MD, Doustimotlagh AH, Meshkani R, Alipoor S, et al. Genetic Variants in Long Noncoding RNA H19 and MEG3 Confer Risk of Type 2 Diabetes in an Iranian Population. *Gene* (2018) 675:265–71. doi: 10.1016/j.gene.2018.07.002
19. Gong J, Liu W, Zhang J, Miao X, Guo AY. lncRNASNP: A Database of SNPs in lncRNAs and their Potential Functions in Human and Mouse. *Nucleic Acids Res* (2015) 43:D181–6. doi: 10.1093/nar/gku1000
20. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial Sequential Analysis may Establish When Firm Evidence Is Reached in Cumulative Meta-Analysis. *J Clin Epidemiol* (2008) 61(1):64–75. doi: 10.1016/j.jclinepi.2007.03.013
21. Shaker O, Ayeldeen G, Abdelhamid A. The Impact of Single Nucleotide Polymorphism in the Long Non-Coding MEG3 Gene on microRNA-182 and microRNA-29 Expression Levels in the Development of Breast Cancer in Egyptian Women. *Front Genet* (2021) 12:683809. doi: 10.3389/fgene.2021.683809
22. Kong X, Yang S, Liu C, Tang H, Chen Y, Zhang X, et al. Relationship Between MEG3 Gene Polymorphism and Risk of Gastric Cancer in Chinese Population With High Incidence of Gastric Cancer. *Biosci Rep* (2020) 40(11): BSR20200305. doi: 10.1042/BSR20200305

23. Xu B, Zhang M, Liu C, Wang C, You Z, Wang Y, et al. Association of Long Non-Coding RNA MEG3 Polymorphisms and Risk of Prostate Cancer in Chinese Han Population. *Urol J* (2020) 18(2):176–80. doi: 10.22037/uj.v16i7.5585
24. Ali MA, Shaker OG, Alazrak M, AbdelHafez MN, Khalefa AA, Hemeda NF, et al. Association Analyses of a Genetic Variant in Long Non-Coding RNA MEG3 With Breast Cancer Susceptibility and Serum MEG3 Expression Level in the Egyptian Population. *Cancer Biomark* (2020) 28(1):49–63. doi: 10.3233/CBM-191072
25. Mazraeh SA, Gharesouran J, Fard SG, Ketab FNG, Hosseinzadeh H, Moradi M, et al. Association Between WT1 and MEG3 Polymorphisms and Risk of Acute Myeloid Leukemia. *Meta Gene* (2020) 23:100636. doi: 10.1016/j.mgene.2019.100636
26. Wei ZM. Analysis of the Effect of Genetic Variants of MEG3-P53-MDM2-PTEN Molecular Pathway on Hepatocarcinogenesis. Master's degree dissertation. (2019). doi: 10.27690/d.cnki.ggdyk.2019.000060
27. Zhuo ZJ, Zhang R, Zhang J, Zhu J, Yang T, Zou Y, et al. Associations Between lncRNA MEG3 Polymorphisms and Neuroblastoma Risk in Chinese Children. *Aging (Albany NY)* (2018) 10(3):481–91. doi: 10.18632/aging.101406
28. Zhang Q, Ai L, Dai YG. Association Between Polymorphism of Long Non-Coding RNA Maternally Expressed Gene 3 and Risk of Gastric Cancer. *Chin J Bases Clinics Gen Surg* (2018) 25:1323–6. doi: 10.7507/1007-9424.201805074
29. Cao X, Zhuang S, Hu Y, Xi L, Deng L, Sheng H, et al. Associations Between Polymorphisms of Long Non-Coding RNA MEG3 and Risk of Colorectal Cancer in Chinese. *Oncotarget* (2016) 7(14):19054–9. doi: 10.18632/oncotarget.7764
30. Du P, Li G, Zhu J, Luo Y, Fu N, Li Y. Association of lncRNA PRNCR1 Polymorphisms With Cancer Susceptibility: A Meta-Analysis of the Current Literature. *J Genet* (2021) 100:19. doi: 10.1007/s12041-021-01269-3
31. Li HN, Deng N, Zhao X, Liu J, He T, Ding XW. Contributions of HOTAIR Polymorphisms to the Susceptibility of Cancer. *Int J Clin Oncol* (2021) 26(6):1022–38. doi: 10.1007/s10147-021-01884-1
32. Li N, Cui Z, Huang D, Gao M, Li S, Song M, et al. Association of LINC00673 Rs11655237 Polymorphism With Cancer Susceptibility: A Meta-Analysis Based on 23,478 Subjects. *Genomics* (2020) 112(6):4148–54. doi: 10.1016/j.ygeno.2020.07.015
33. Wu X, Li J, Ren Y, Zuo Z, Ni S, Cai J. MEG3 can Affect the Proliferation and Migration of Colorectal Cancer Cells Through Regulating miR-376/PRKD1 Axis. *Am J Transl Res* (2019) 11:5740–51.
34. Jiao J, Zhang S. Long Non−Coding RNA MEG−3 Suppresses Gastric Carcinoma Cell Growth, Invasion and Migration *via* EMT Regulation. *Mol Med Rep* (2019) 20(3):2685 –93. doi: 10.3892/mmr.2019.10515
35. Dong S, Ma M, Li M, Guo Y, Zuo X, Gu X, et al. LncRNA MEG3 Regulates Breast Cancer Proliferation and Apoptosis Through miR-141-3p/RBMS3 Axis. *Genomics* (2021) 113(4):1689–704. doi: 10.1016/j.ygeno.2021.04.015

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 December 2021
doi: 10.3389/fonc.2021.759586

Systematic Analysis of an Invasion-Related 3-Gene Signature and Its Validation as a Prognostic Model for Pancreatic Cancer

Dafeng Xu[1], *Yu Wang*[2], *Yuliang Zhang*[3], *Zhehao Liu*[1], *Yonghai Chen*[1] *and Jinfang Zheng*[1*]

[1] Department of Hepatobiliary and Pancreatic Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, China, [2] Geriatric Medicine Center, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, China, [3] Department of Otolaryngology Head and Neck Surgery, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, China

Background: Pancreatic adenocarcinoma (PAAD) is a malignant tumor of the digestive system that is associated with a poor prognosis in patients owing to its rapid progression and high invasiveness.

Methods: Ninety-seven invasive-related genes obtained from the CancerSEA database were clustered to obtain the molecular subtype of pancreatic cancer based on the RNA-sequencing (RNA-seq) data of The Cancer Genome Atlas (TCGA). The differentially expressed genes (DEGs) between subtypes were obtained using the limma package in R, and the multi-gene risk model based on DEGs was constructed by Lasso regression analysis. Independent datasets GSE57495 and GSE62452 were used to validate the prognostic value of the risk model. To further explore the expression of the hub genes, immunohistochemistry was performed on PAAD tissues obtained from a large cohort.

Results: The TCGA-PAAD samples were divided into two subtypes based on the expression of the invasion-related genes: C1 and C2. Most genes were overexpressed in the C1 subtype. The C1 subtype was mainly enriched in tumor-related signaling pathways, and the prognosis of patients with the C1 subtype was significantly worse than those with the C2 subtype. A 3-gene signature consisting of *LY6D*, *BCAT1*, and *ITGB6* based on 538 DEGs between both subtypes serves as a stable prognostic marker in patients with pancreatic cancer across multiple cohorts. *LY6D*, *BCAT1*, and *ITGB6* were over-expressed in 120 PAAD samples compared to normal samples.

Conclusions: The constructed 3-gene signature can be used as a molecular marker to assess the prognostic risk in patients with PAAD.

Keywords: Pancreatic adenocarcinoma (PAAD), invasive-related genes, LY6D, BCAT1, ITGB6, prognosis

OPEN ACCESS

Edited by:
Carlos A. Vaccaro, Italian Hospital of Buenos Aires, Argentina

Reviewed by:
Bangli Hu, Guangxi Medical University Cancer Hospital, China
Adam Lauko, Cleveland Clinic, United States

***Correspondence:**
Jinfang Zheng
zhengjf71@sina.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *16 August 2021*
Accepted: *24 November 2021*
Published: *15 December 2021*

Citation:
Xu D, Wang Y, Zhang Y, Liu Z, Chen Y and Zheng J (2021) Systematic Analysis of an Invasion-Related 3-Gene Signature and Its Validation as a Prognostic Model for Pancreatic Cancer. Front. Oncol. 11:759586. doi: 10.3389/fonc.2021.759586

Abbreviations: PAAD, Pancreatic adenocarcinoma; DEGs, Differentially expressed genes; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; *GSEA, Gene Set Enrichment Analysis;* FDR, False discovery rate; AIC, Akaike Information Criterion;KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; OS, Overall survival; CDF, Cumulative distribution function; ROC, Receiver operating characteristic.

BACKGROUND

Pancreatic adenocarcinoma (PAAD) is a malignant tumor of the digestive tract and is the fourth leading cause of cancer-related deaths worldwide (1). Since the early symptoms of PAAD are not obvious, its diagnosis is often difficult, and the primary tumor exhibits vascular invasion. Approximately 80-85% of patients with pancreatic cancer present with distant metastases at the time of diagnosis, thus making radical resection ineffective (2). Therefore, the search for more accurate and effective diagnostic and prognostic markers is of great significance for the stratification and individualization of patients with pancreatic cancer in the clinical setting.

At present, the prognostic prediction of patients with pancreatic cancer is mainly based on clinicopathologic features. However, the prognosis of patients with the same clinical stage and grade differs because of the high heterogeneity of pancreatic cancer. Moreover, the malignant progression of pancreatic cancer is accompanied by genetic changes. Therefore, the study of the molecular mechanisms underlying pancreatic cancer progression is key to prolonging the overall survival of patients with pancreatic cancer (3). However, the effect of a single gene in predicting the prognosis of a pancreatic tumor is often unsatisfactory and presents with some limitations; the combined detection of multiple genes is expected to facilitate the prognostic prediction of patients with pancreatic cancer. With the rapid development of bioinformatics and sequencing technology, an increasing number of studies have provided potential prognostic assessments for patients with pancreatic cancer. Li et al. (4) constructed a 9-gene signature using macrophage phenotypic switch-related genes in patients with pancreatic cancer. Wang et al. (5) constructed a 9-gene signature for predicting PAAD based on the expression of immune-related genes. However, most prognostic models include a large number of genes, which greatly increases the cost of medical treatment in clinical practice. Moreover, most studies are based on a comprehensive analysis of public databases and lack experimental data to verify and explore the role of the identified genes in the development of pancreatic cancer.

In this study, a molecular subtype of pancreatic cancer was constructed based on invasion-related genes using gene expression data from The Cancer Genome Atlas (TCGA), Gene Omnibus Expression (GEO), and other public databases. The relation between molecular subtypes, prognosis, and clinical features was further analyzed. A 3-gene prognostic model, composed of *LY6D*, *BCAT1*, and *ITGB6*, constructed with differentially expressed genes (DEGs) between the PAAD subtypes, could be used to evaluate the prognosis of patients with PAAD.

MATERIALS AND METHODS

Data Source and Preprocessing

RNA-sequencing (RNA-seq) data and clinical follow-up information data from TCGA-PAAD samples were downloaded from the TCGA database. The expression data and clinical information from the GSE57495, GSE62452 and GSE28735 datasets were downloaded from the GEO database. A total of 97 invasion-related genes were collected from the CancerSEA website (**Supplement Table 1**).

The RNA-seq data from the TCGA-PAAD dataset was processed through the following steps: 1) Samples with no clinical follow-up information were removed; 2) The ENSEMBL gene IDs were converted to the Gene Symbol format; 3) The median value was calculated with multiple Gene Symbol expressions.

The following steps were used to process the GEO dataset: 1) Samples without clinical follow-up information were removed; 2) The probe IDs were converted to the Gene Symbol format; 3) Probes that corresponded to multiple genes were removed. 4) When multiple probes correspond to one gene, take the average value as the gene expression.

After preprocessing, we enrolled 176 samples from TCGA-PAAD, 63 samples from GSE57495 data set, 66 samples from GSE62452 data set, and 42 samples from GSE28735 dataset. The clinical characteristics of the patient samples are listed in **Table 1**.

Consistency Clustering Algorithm and Gene Set Enrichment Analysis (GSEA)

The expression profiles of 97 invasion-related genes were extracted from the TCGA-PAAD dataset, and univariate Cox regression analysis was performed to select significant prognostic genes using coxph function in R ($p < 0.05$). Next, the genes with significant results from the univariate Cox analysis were clustered using ConsensusClusterPlus (V1.48.0; parameters: reps = 100, pitem = 0.8, pfeature = 1, and distance = "Canberra"). The Pam and Canberra distances were used as a clustering algorithm and distance measure, respectively.

The gene set c2.cp.kegg.v7.0.symbols.gmt was selected, and significantly enriched pathways between different molecular subtypes were analyzed by GSEA. PAAD samples were divided into either a C1 or C2 subtype based on gene expression data from the TCGA-PAAD dataset in the GSEA input file. The thresholds for pathway enrichment analysis were $p < 0.05$ and false discovery rate (FDR) < 0.25.

Identification of DEGs

DEGs between C1 and C2 subtypes were calculated using the limma package (6), and the filtering thresholds were FDR < 0.05 and | log 2 fold-change (FC) | > 1. The identified DEGs were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Gene Ontology (GO) enrichment analysis using the WebGestaltR (v0.4.2) package in R software.

Construction of a Risk Model Based on Invasion-Related Genes

Random Grouping of Training Set Samples

The 176 samples in the TCGA-PAAD dataset were divided into a training set and validation set. To avoid the effect of random assignment bias on the stability of subsequent modeling, 200

TABLE 1 | Clinical characteristics of patient samples.

Clinical Features	TCGA-PAAD	GSE57495	GSE62452	GSE28735
OS				
0	84	21	16	13
1	92	42	50	29
T Stage				
T1	7			
T2	24			
T3	140			
T4	3			
TX	2			
N Stage				
N0	49			
N1	122			
NX	5			
M Stage				
M0	79			
M1	4			
MX	93			
Stage				
I	21			
II	145			
III	3			
IV	4			
X	3			
Grade				
G1	30			
G2	94			
G3	48			
G4	2			
GX	2			
Gender				
Male	96			
Female	80			
Age				
≤65	93			
>65	83			
Alcohol				
YES	100			
NO	64			
Unknown	12			
Chemotherapy				
YES	116			
NO	60			
Radiation therapy				
YES	32			
NO	101			
Unknown	43			

samples were assigned to random groups. The samples were grouped according to a training set: validation set ratio of 3:2. After dividing the samples, there were 106 samples in the training set and 70 samples in the validation set.

Lasso Regression Analysis and Stepwise Regression Analysis of Training Set Data

Univariate Cox regression analysis was performed for each DEG (538 in total) using the coxph function in R to identify prognostic genes, and $p < 0.05$ was selected as the threshold for filtering. Lasso regression analysis was performed to further reduce the number of genes in the risk model using the glmnet package in R (7). In stepwise regression analysis, the selection of the model starts with the most complex model from which one variable is removed at a time to reduce the number of parameters according to the Akaike Information Criterion (AIC). The smaller the p-value of the regression model, the more superior the model. This indicates that the regression model fits the data well with fewer parameters. The prognostic model is made fit for clinical applications by performing stepwise regression to further reduce the number of genes.

The prognostic model was constructed based on the following equation:

$$\text{risk score} = \sum_{i=1}^{n} \beta i \times \exp(G_i)$$

where n refers to the number of genes identified for the multivariate Cox regression model; exp(Gi) is the expression value of gene i; and βi is the coefficient for gene i.

Immunohistochemistry

To verify the expression of the candidate three genes, tissue microarrays (TMA) comprised of 120 PAAD tissues and 30 normal samples were obtained from Shanghai Outdo Biotech Co., Ltd. (Shanghai, China). The clinicopathological details of 120 PAAD tissues were shown in **Table 2**. The studies were conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS), and the research protocols were approved by the

TABLE 2 | The clinicopathological details of 120 PAAD tissues.

Clinical Features	PAAD-IHC
T Stage	
T1	4
T2	30
T3	61
T4	1
TX	24
N Stage	
N0	54
N1	63
NX	3
M Stage	
M0	112
M1	8
MX	0
Stage	
I	21
II	90
III	1
IV	8
X	0
Grade	
G1	1
G2	76
G3	38
G4	0
GX	5
Gender	
Male	66
Female	54
Age	
≤65	77
>65	43

Ethics Committee of Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University.

The TMA slides were dried overnight at 37°C, dewaxed in xylene, and dehydrated in a gradient ethanol series. Antigens retrieval was performed by heating the tissue sections in a microwave oven inside a vessel filled with EDTA antigen retrieval buffer (pH 9.0). Subsequently, the tissue sections were immersed in 3% hydrogen peroxide for 25 min to block the activity of endogenous peroxides. Next, the TMA tissues were coated with 3% bovine serum albumin (BSA) and sealed at room temperature for 30 min to reduce non-specific staining. Then, the TMA slides were incubated with anti-LY6D (1: 200 dilution; Novus Biologicals, NBP1-84029), anti-BCAT1 (1:50 dilution; Abcam, ab197941), and anti-ITGB6 (1:10 dilution; Abcam, ab197672) overnight at 4°C.

The tissues were rinsed with 0.01 mol/L phosphate buffer saline (PBS; pH = 7.4) for 5 min each. The tissues were incubated at room temperature for 50 min with horseradish peroxidase (HRP)-labeled goat anti-rabbit secondary antibody (1:200 dilution, ServiceBio, GB23303). Then, the tissues were washed in PBS and stained with 3,3-diaminobenzidine (DAB). Finally, the TMA sections were counterstained with Mayer's hematoxylin, dehydrated, and fixed. To evaluate IHC staining, semi-quantitative scoring criteria were used.

The stained sections were scored by three pathologists who were blinded to the patients' clinical characteristics. The scoring system was based on the proportion of positively stained cells in all tissues and the staining intensity of these positively stained cells. The staining intensity was classified as follows: 0 (negative), 1 (weak), 2 (moderate), or 3 (strong). The staining ratio of positive cells was classified as follows: 0 (<5%), 1 (5%-25%), 2 (26%-50%), 3 (51%-75%), or 4 (> 75%). According to the staining intensity and the proportion of positively stained cells, the tissues were graded as follows: 0-1 grade, negative (-); > 1-4, weakly positive (+); > 4-8, moderately positive (++), and > 8- 12, strongly positive (+++).

RESULTS

Identification of Molecular Subtypes Based on Invasion-Related Genes

Thirty-five genes were found to be significantly associated with the prognosis of pancreatic cancer using univariate Cox analysis (**Supplement Table 2**). Consistent cluster analysis showed that the samples could be clustered together at k=2 (**Figures 1A, B**). The expression levels of the invasion-related genes were significantly different between the C1 and C2 subtypes, and most genes were overexpressed in the C1 subtype (**Figure 1C**). The relationship between the subtypes and prognosis was further analyzed, and results showed that there were significant differences in survival times between the C1 and C2 subtypes (**Figures 1D, E**, log-rank $p < 0.05$).

The results of the GSEA analysis showed the activation of more tumor-related pathways in the C1 subtype, such as pathways in cancer, notch signaling pathway, focal adhesion, extracellular matrix (ECM)-receptor interaction, and TGF-β signaling pathway (**Figure 1F**), suggesting that the C1 subtype is more closely related to cancer than the C2 subtype.

Analysis of DEGs Between Subtypes

According to the thresholds mentioned in the methods section, 538 DEGs were obtained, of which 531 genes were upregulated and 7 genes were downregulated (**Supplement Table 3**). The results demonstrated that the C1 subtype contains more upregulated genes than the C2 subtype. The volcano map of upregulated and downregulated DEGs between the two subtypes is shown in **Supplementary Figure 1A**. The expression patterns of the top 50 upregulated DEGs and all the downregulated DEGs were shown in a heatmap (**Supplementary Figure 1B**). The results of the GO enrichment analysis of DEGs showed that 548 Biological Process (BP) terms were significantly different between the two subtypes (FDR < 0.05). The first 15 BP terms were plotted (FDR < 0.05), as shown in **Supplementary Figure 1C**. The first 15 Cellular Component (CC) terms were plotted, as shown in **Supplementary Figure 1D**. Fifty-two Molecular Function (MF) terms were significantly different between the two subtypes (FDR < 0.05). The results of the first 15 MF terms are shown in **Supplementary Figure 1E**. The KEGG pathway analysis of DEGs showed 27 significantly enriched pathways (FDR < 0.05). Further visualization of the top 10 enriched pathways showed that genes were significantly enriched in tumor-related pathways such as the ECM-receptor interaction pathway, focal adhesion, and the PI3K-Akt signaling pathway (**Supplementary Figure 1F**).

Comparison of Immune Score Between Molecular Subtypes

To identify the relationship between molecular subtypes and immune scores in the TCGA-PADD dataset, the ESTIMATE package was used to evaluate the three immune scores: stromal, immune, and estimate scores. MCPcounter was used to evaluate 10 types of immune cells, and the single-sample GSEA (ssGSEA) method in the GSCA package was used to evaluate 28 types of immune cells (8). Meanwhile, the difference in immune scores between the two molecular subtypes was compared. The results showed that the immune scores of the C1 subtype were higher than those of the C2 subtype (**Figures 2A–C**). The heatmap of the immune scores of the two subtypes is shown in **Figure 2D**.

Risk Model of Pancreatic Cancer Based on Invasion-Related Genes

By performing univariate Cox analysis of the DEGs between the C1 and C2 subtypes, 18 prognostic genes were identified. Lasso regression analysis was performed to further reduce the number of prognostic genes. The locus of each independent variable is shown in **Supplementary Figure 2A**. As the value of lambda (λ) increased, the number of independent variables tending to zero also increased. A 10-fold cross-validation was performed to construct the model, and the confidence interval under each λ is shown in **Supplementary Figure 2B**. The model was found to be optimal when λ = 0.05667557, so a λ of 0.0567557 was chosen for further analysis of the prognostic genes. Six genes, namely *LY6D*, *DKK1*,

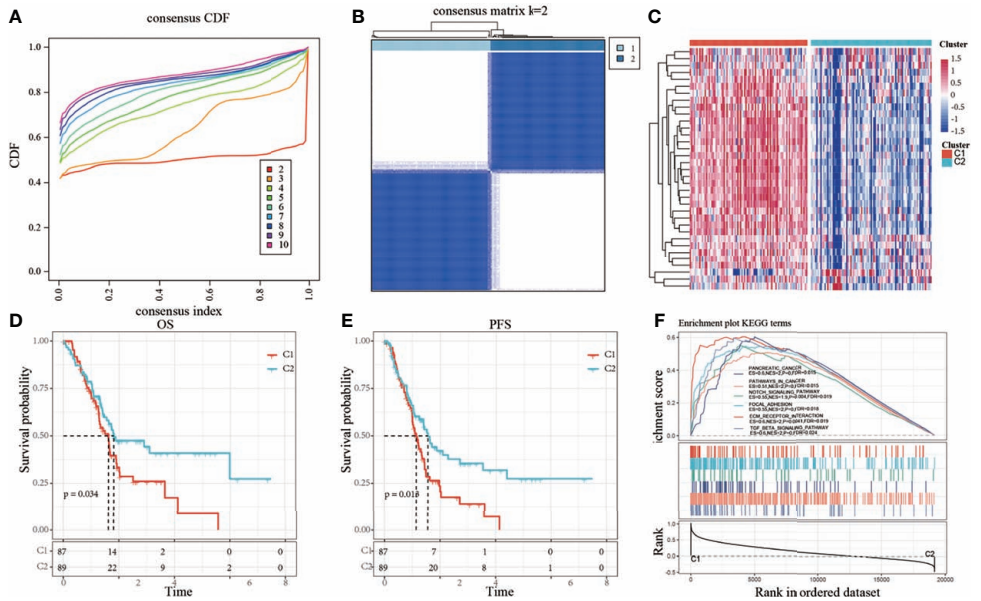


FIGURE 1 | **(A)** Cumulative distribution function (CDF); **(B)** Consistent clustering heatmap when k = 2; **(C)** Cluster heatmap of 35 prognostic genes; **(D)** Overall survival (OS) curve based on molecular subtypes in all samples of The Cancer Genome Atlas-pancreatic adenocarcinoma (TCGA-PAAD) dataset; **(E)** Progression-free survival (PFS) curve based on molecular subtypes in all TCGA-PAAD samples; **(F)** Involvement of tumor-related pathways between molecular subtypes of the TCGA dataset.

BICC1, *BCAT1*, *ITGB6*, and *PTGES* were identified as the hub genes when $\lambda = 0.0567557$. The number of model genes was further reduced by stepwise regression, and finally, three genes were obtained: *LY6D*, *BCAT1*, and *ITGB6*. The risk score based on the final 3-gene prognostic model was calculated as follows: Risk score = 0.1627483 * *LY6D* + 0.2210480 * *BCAT1* + 0.2005339 * *ITGB6*.

Risk scores of each sample were calculated based on the expression level of *LY6D*, *BCAT1*, and *ITGB6*, and a risk score distribution was plotted for each sample, as shown in **Figure 3A**. The results showed that a higher risk score was associated with worse outcomes, and high expression levels of *LY6D*, *BCAT1*, and *ITGB6* were associated with a higher risk score. The timeROC package was used to analyze the receiver operating characteristic (ROC) curve of risk score; the 1-, 2-, and 3-year predictive classification efficiencies were 0.76, 0.78, and 0.75, respectively, as shown in **Figure 3B**. The samples were divided into a high-risk group and a low-risk group based on the risk scores. Finally, 50 and 56 samples were placed into the high- and low-risk groups, respectively. The KM curve showed a significant difference in the expression of DEGs between the high- and low-risk groups ($p < 0.01$) (**Figure 3C**).

Verification of Robustness of the 3-Gene Prognostic Model Using Internal and External Datasets

Verification of the Robustness of the 3-Gene Prognostic Model Using Internal Datasets

To determine the robustness of the model, the risk score distribution of the TCGA validation set and all dataset samples was calculated using the same coefficients as those of the training set. The risk score distribution of the TCGA validation set suggested that samples with a high risk score are associated with a worse prognosis, as shown in **Figure 4A**. The 1-, 2-, and 3-year predictive classification efficiencies of the risk scores were 0.67, 0.76, and 0.87, respectively (**Figure 4B**). These results demonstrated that the prognosis of the high-risk group was significantly worse than that of the low-risk group (**Figure 4C**).

The risk score distribution trend of all TCGA datasets was consistent with those of the training set (**Figure 5A**). The predictive classification efficiencies of the 1-, 2-, and 3-year ROCs were 0.73, 0.77, and 0.81, respectively (**Figure 5B**). According to the above classification, 89 and 87 samples were categorized into the high- and low-risk groups, respectively, in all TGGA datasets. The prognosis of the high-risk group was significantly worse than that of the low-risk group (**Figure 5C**).

Validation of the Robustness of the 3-Gene Prognostic Model Using Three Independent Cohorts

The robustness of the model was further verified with three independent validation cohorts GSE57495, GSE62452 and GSE28735. The 1-, 3-, and 5-year ROCs in the GSE57495 dataset were 0.63, 0.74, and 0.78, respectively (**Figure 6A**). The 1-, 3-, and 5-year ROCs in the GSE62452 were 0.56, 0.71, and 0.84, respectively (**Figure 6C**). The 1-, 3-, and 5-year ROCs in the GSE28735 were 0.61, 0.72, and 0.68, respectively (**Figure 6E**). Therefore, the predictive performance of the model was stable in different cohorts. Finally, the samples with a risk score greater than zero after zscore method were classified into the high-risk group and those with a risk score less than 0 were classified into

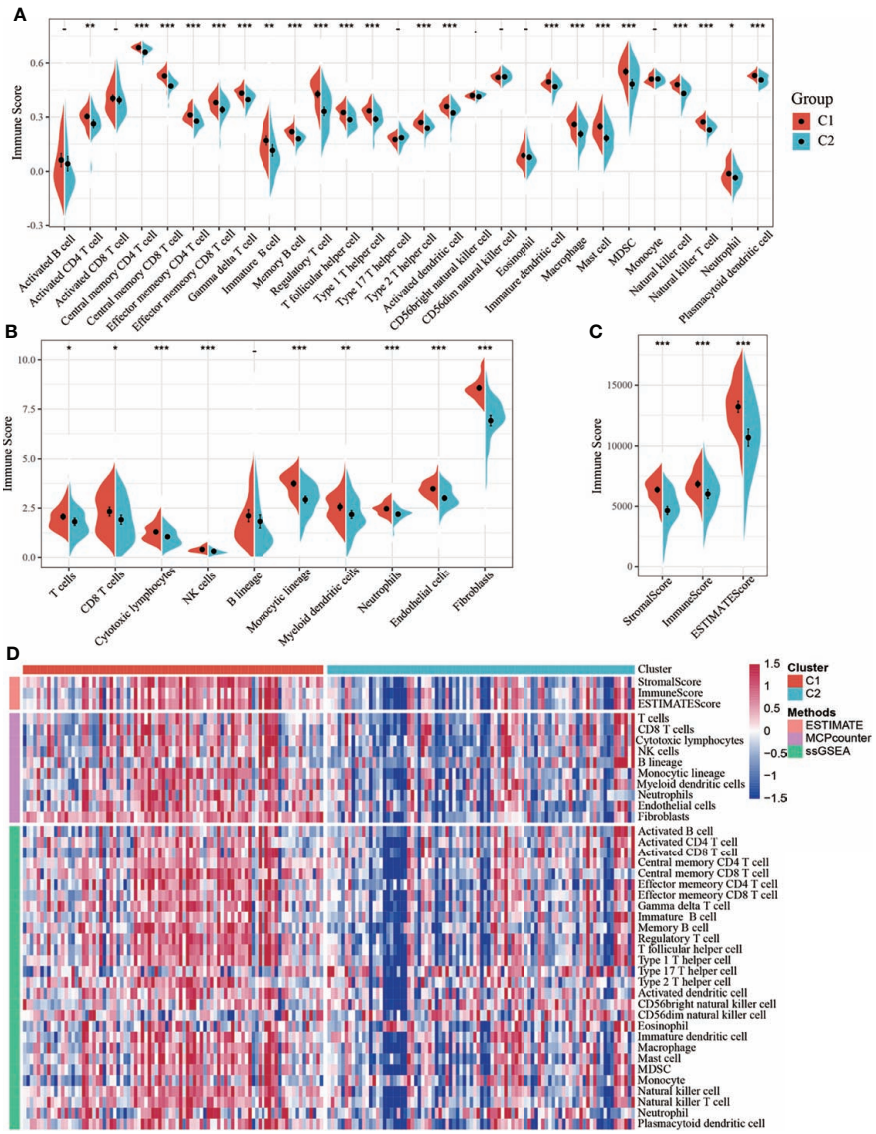


FIGURE 2 | **(A)** Comparison of single-sample gene set enrichment analysis (ssGSEA) immune scores between molecular subtypes in all samples of The Cancer Genome Atlas-pancreatic adenocarcinoma (TCGA-PAAD) dataset; **(B)** Comparison of MCPcounter immune scores between molecular subtypes of the TCGA dataset; **(C)** Comparison of estimated immune scores between molecular subtypes of the TCGA dataset; **(D)** Heat map comparing three software immune scores among molecular subtypes of the TCGA dataset. *P < 0.05; **P < 0.01; ***P < 0.001.

the low-risk group. In the GSE57495 cohort, 30 and 33 samples were categorized into the high- and low-risk groups, respectively, with significant prognostic differences between the two groups (**Figure 6B**). In the GSE62452 cohort, 33 samples each were categorized into the high and low-risk groups, respectively, with significant prognostic differences between the two groups (**Figure 6D**). In the GSE28735 cohort, 21 samples each were categorized into the high and low-risk groups, respectively, with significant prognostic differences between the two groups (**Figure 6F**).

Risk Model and Prognostic Analysis of Clinical Features

Further analysis of the relationship between the risk score and clinical features showed that the 3-gene prognostic model could significantly distinguish between age, sex, TNM stage, clinical stage, tumor grade, alcohol consumption, chemotherapy, and radiation therapy between the high- and low-risk groups (**Figures 7A–P**, $p < 0.05$). This suggests that the model also has good predictive power in distinguishing different clinical features.

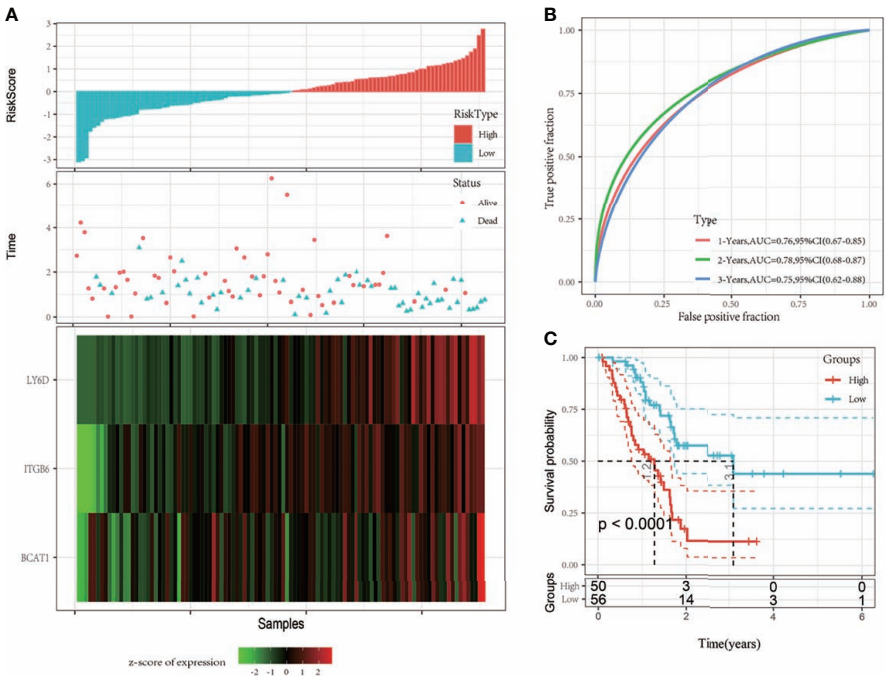


FIGURE 3 | **(A)** The risk score, survival time and state, and expression of the 3-gene signature were studied in The Cancer Genome Atlas (TCGA) training set. **(B)** Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the 3-gene signature; **(C)** The Kaplan-Meier (KM) survival curve distribution of the 3-gene signature in the training set.

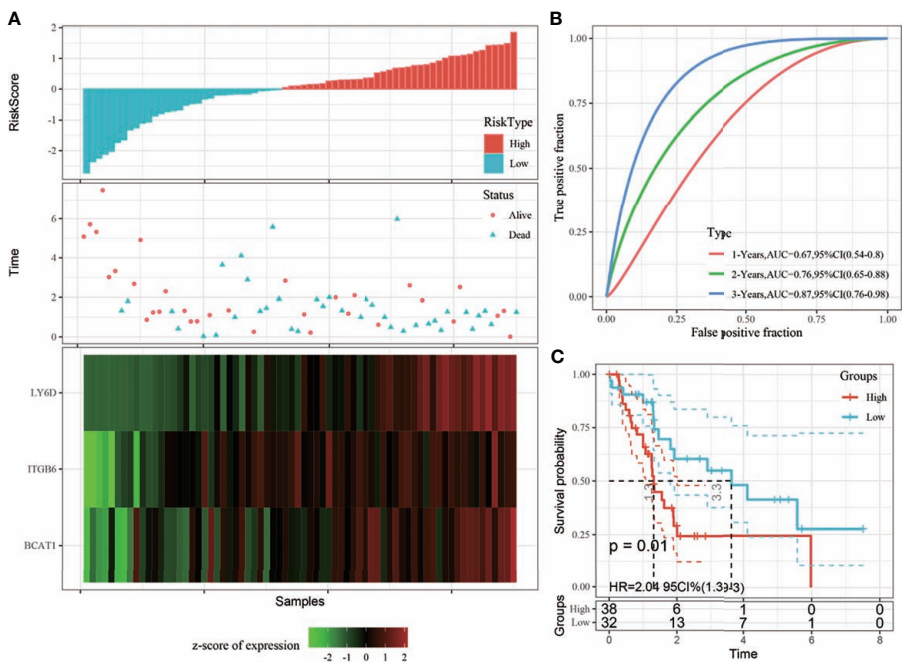


FIGURE 4 | **(A)** Risk score, survival time, survival status, and 3-gene signature expression in The Cancer Genome Atlas (TCGA) training set; **(B)** ROC curve and area under the curve (AUC) of the 3-gene signature; **(C)** Distribution of the Kaplan-Meier (KM) survival curve of the 3-gene signature in the TCGA validation set.

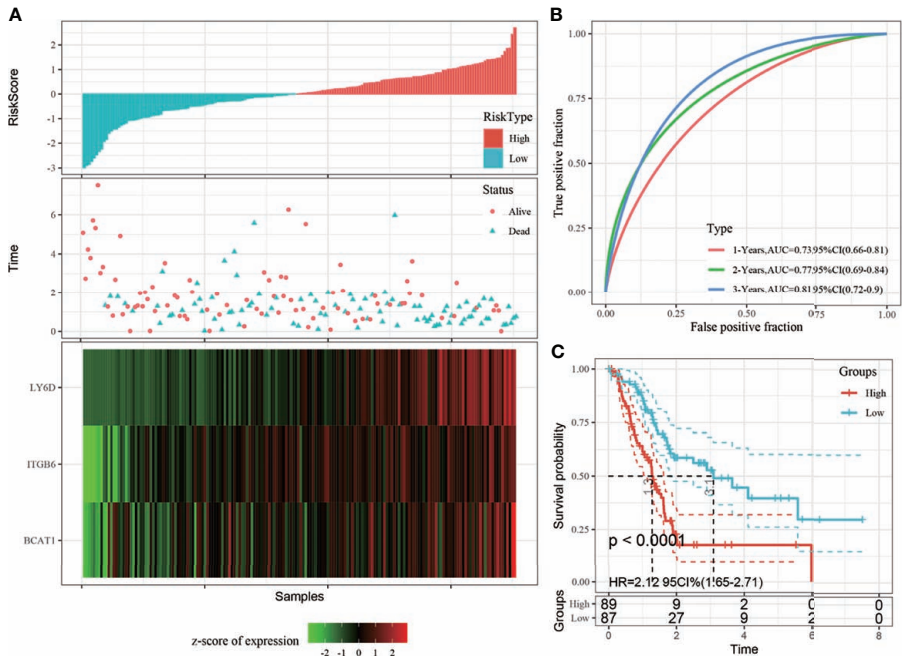


FIGURE 5 | **(A)** Risk score, survival time, and 3-gene signature expression in all The Cancer Genome Atlas (TCGA) datasets; **(B)** Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the 3-gene signature; **(C)** Distribution of the Kaplan-Meier (KM) survival curve of 3-gene signature in all TCGA datasets.

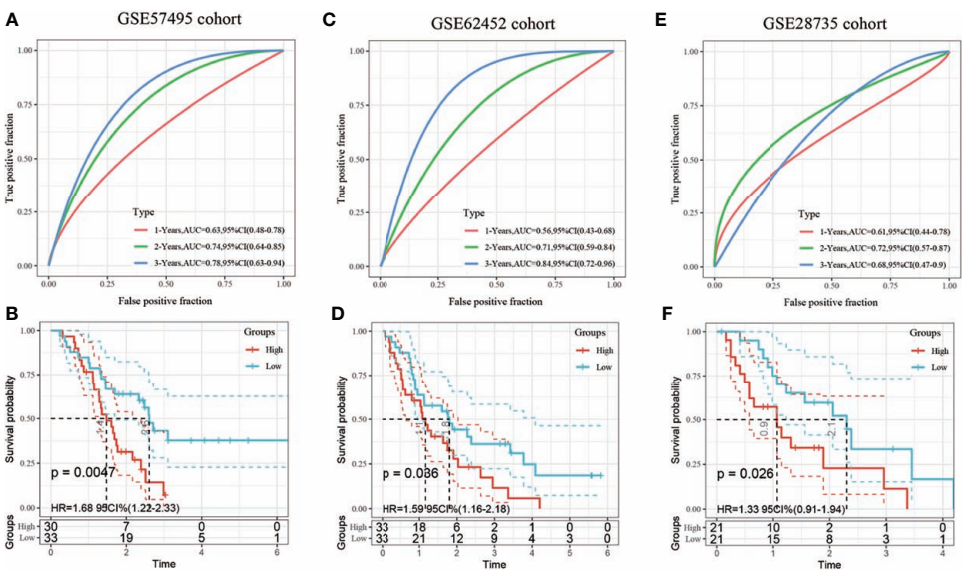


FIGURE 6 | **(A)** Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the 3-gene signature; **(B)** Distribution of the Kaplan-Meier (KM) survival curve of the 3-gene signature in the GSE57495 dataset; **(C)** Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the 3-gene signature; **(D)** Distribution of the Kaplan-Meier (KM) survival curve of the 3-gene signature in the GSE62452 independent validation set; **(E)** Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the 3-gene signature; **(F)** Distribution of the Kaplan-Meier (KM) survival curve of the 3-gene signature in the GSE28735 independent validation set.

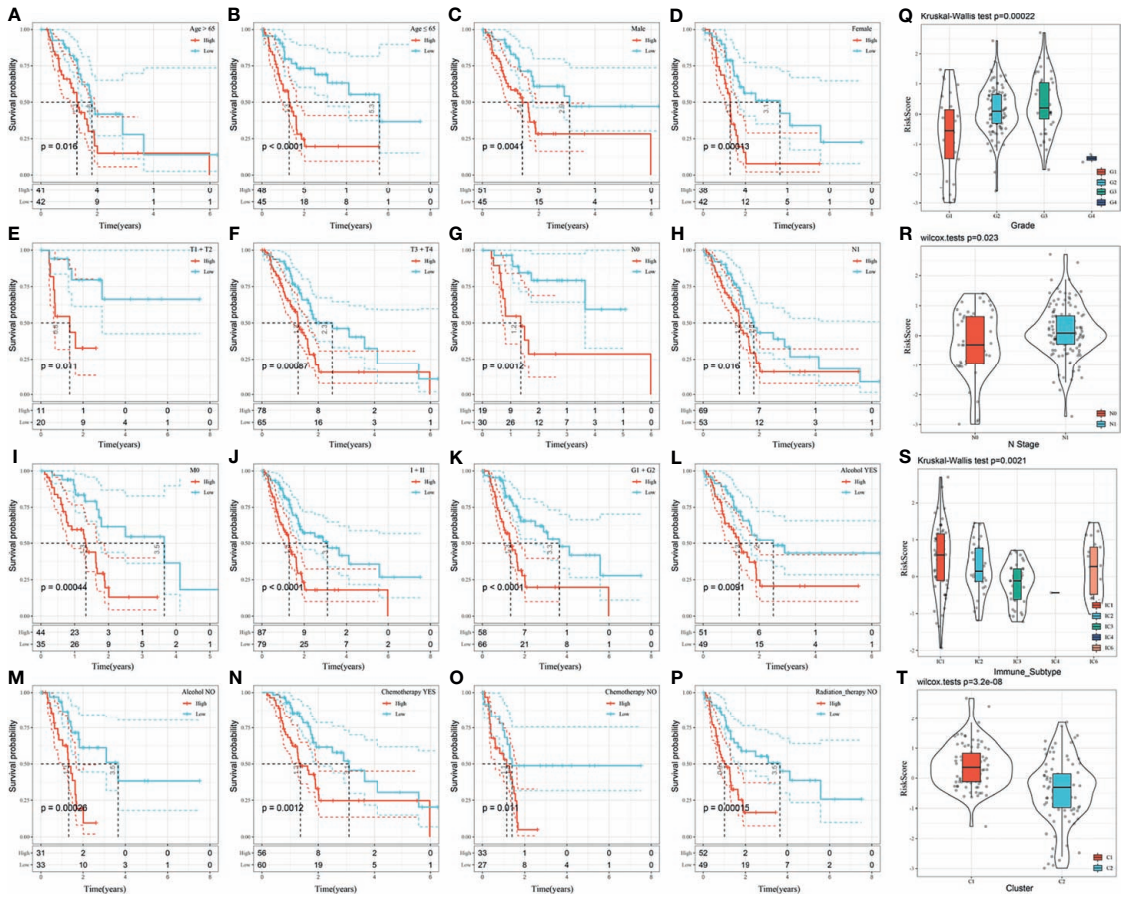


FIGURE 7 | **(A–P)** Performance of the risk model in distinguishing different clinical characteristics of patients; **(Q)** Comparison of the risk score between the samples grouped according to the tumor grade; **(R)** Comparison of the risk score comparison between the samples grouped according to the N-Stage; **(S)** Comparison of the risk score in existing immune molecular subtypes between grouped samples; **(T)** Comparison of the risk score between samples of the molecular subtypes identified in this study.

The distribution of the risk score among the clinical features of the two groups was further compared. The results demonstrated that the risk score is significantly different between the N-stage and tumor grade ($p < 0.05$). The higher the tumor grade, the higher the risk score (**Figure 7Q**). The risk score of N1 was significantly higher than that of N0 (**Figure 7R**). The risk score of the C1 subtype with a poor prognosis was significantly higher than that of the C2 subtype with a good prognosis (**Figure 7T**). Moreover, the risk score was significantly different among existing immune molecular subtypes (**Figure 7S**).

Construction of the Nomogram

In the TCGA-PAAD dataset, the univariate Cox regression analysis showed a significant correlation between the risk type and survival, while the multivariate Cox regression analysis showed a significant correlation between the risk score (Hazard ratio [HR] = 1.94, 95% confidence interval [CI] = 1.25–3.01, and $p = 0.003$) and survival. These results demonstrate the good predictive performance of the identified 3-gene prognostic model in clinical applications. Furthermore, the N stage (HR = 4.17, 95% 1.16–14.93, and $p = 0.028$) and grade (HR = 3.06, 95% CI = 1.3–7.21, and $p = 0.011$) were identified as independent prognostic risk factors for patients with pancreatic cancer. Chemotherapy (HR = 0.13, 5% CI = 0.05–0.36, and $p < 0.001$) was identified as an independent prognostic protective factor (**Figures 8A, B**).

The nomogram, which displays the results of the risk model directly and effectively, can be conveniently applied to the prediction of an outcome. The nomogram uses the length of the line to indicate the degree of influence that different variables have on the result and the influence of different values of variables on the result. According to the results of the univariate and multivariate analyses, the nomogram was constructed with the following clinical features: N stage, tumor grade, chemotherapy, and risk score (**Figure 8C**). The results showed that the risk score has the greatest effect on survival prediction, indicating that the risk model based on the 3-gene signature can accurately predict the prognosis of patients with pancreatic cancer. A calibration diagram was used to visualize the nomogram. The results showed that the nomogram

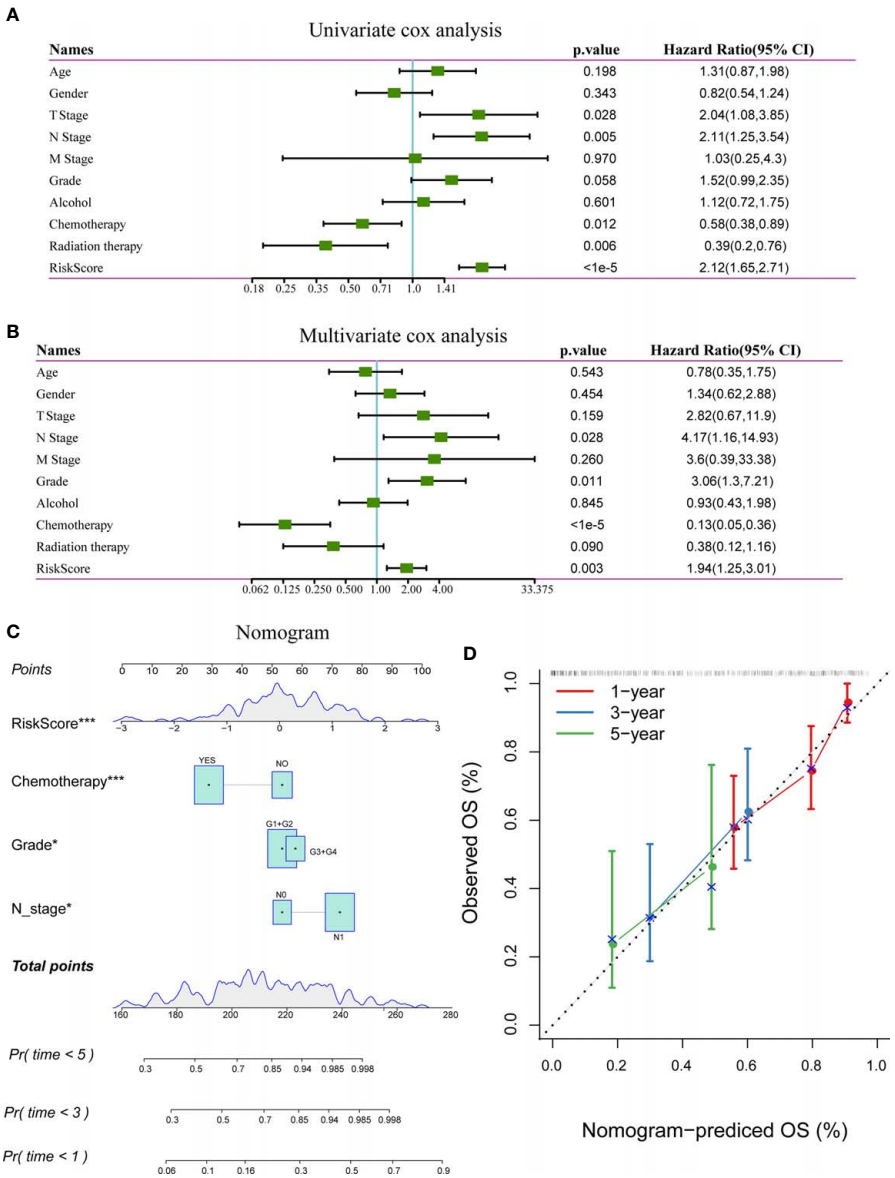


FIGURE 8 | **(A)** Results of univariate analysis of clinical characteristics and risk scores; **(B)** Results of multivariate analysis clinical characteristics and risk scores; **(C)** Nomogram based on clinical characteristics and risk scores; **(D)** Nomogram for predicting survival rate of patients with pancreatic cancer along with correction factors. *P < 0.05; ***P < 0.001.

performed well in determining the prognostic risk of patients with pancreatic cancer (**Figure 8D**).

Comparison of Risk Model With Other Models

Four prognostic risk models, including 15-gene signature (Chen) (9), 7-gene signature (Cheng) (10), and 6-gene signature (Stratford) (11) models, were compared with the identified 3-gene prognostic model. To facilitate comparison among the models, the risk score of each TCGA-PAAD sample was calculated using the same method, and the risk score was zscored according to the corresponding gene in all three models. Genes with a risk score greater than zero were categorized into a high-risk group and those with a risk score less than zero were categorized into a low-risk group. The prognosis difference between the two groups was further analyzed. There were significant differences in outcomes between the high-risk and low-risk groups in all three risk models (**Figures 9B, D, F**, log-rank $p < 0.05$), the area under the curve (AUC)s at 1-, 2-, and 3-year of Cheng and Stratford models were lower than that of our model (**Figures 9C, E**). Although our 1-year AUC is smaller than the Chen model (0.73 vs 0.74), the AUC at 2 and 3 years is larger

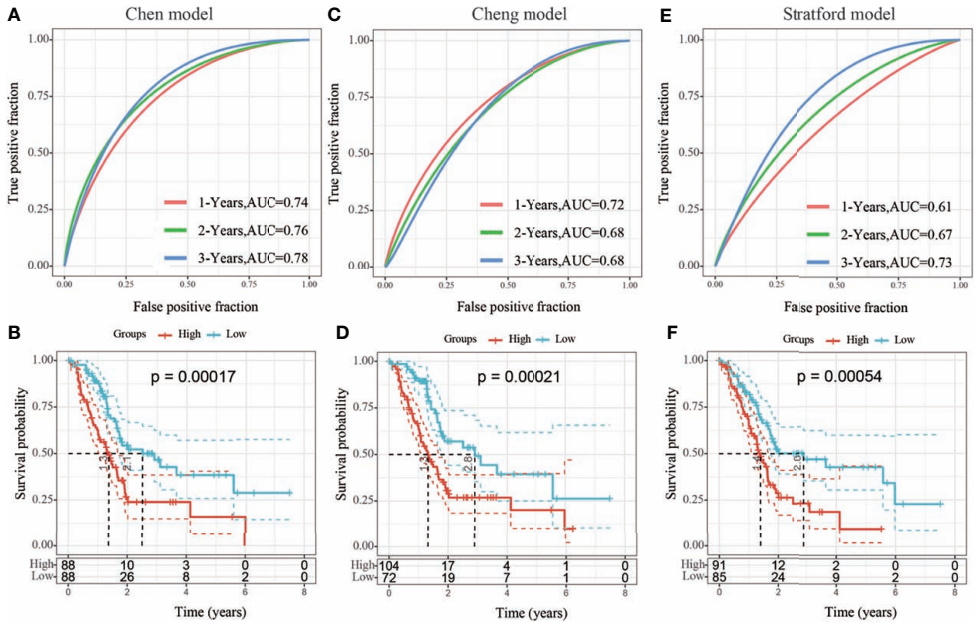


FIGURE 9 | **(A, B)** Receiver operating characteristic (ROC) curve of the 15-gene signature (Chen) risk model and KM curve of High/Low-risk samples; **(C, D)** ROC of the 7-gene signature (Cheng) risk model and the Kaplan-Meier (KM) curve of samples from high- and low-risk groups; **(E, F)** ROC of 6-gene signature (Stratford) risk model and the KM curve of samples from high- and low-risk groups.

than his (0.77 vs 0.76, 0.81 vs 0.78, respectively) (**Figure 9A**) (**Supplement Table 4**). Therefore, the 3-gene signature identified in this study represents a more reasonable and efficient model to determine the prognostic risk of patients with pancreatic cancer with the use of fewer genes.

Expression of LY6D, BCAT1, and ITGB6 in Pancreatic Cancer

The differences in the expression of the *LY6D*, *BCAT1*, and *ITGB6* genes in PAAD and adjacent tissues were investigated. The expressions of *LY6D*, *BCAT1*, and *ITGB6* in 120 cases of pancreatic cancer and 30 cases of para-carcinoma were detected by immunohistochemistry. The results showed that *BCAT1*, *LY6D*, and *ITGB6* were significantly overexpressed in cancer tissues (**Figures 10A–C**). Many cases in the TMA cohort were not effectively followed up. Therefore, to compensate for this limitation, the Kaplan-Meier plotter database was used to obtain 177 samples with overall survival data and 69 cases with recurrence-free survival data. The results showed that patients with high expression of *LY6D*, *BCAT1*, and *ITGB6* genes have a significantly worse prognosis than those with a low expression both in terms of overall survival and recurrence-free survival (**Figures 10D–I**). Our immunohistochemical results demonstrated that LY6D, BCAT1, and ITGB6 proteins were all overexpressed in PAAD samples compared to normal samples. Therefore, it can be speculated that these genes act as oncogenes in pancreatic cancer, and the upregulation of these genes is associated with a significantly worse prognosis in patients with pancreatic cancer.

Flow Chart of Research Methodology

A flowchart has been drawn to allow readers to better understand the research process of this study (**Figure 11**).

DISCUSSION

Pancreatic cancer is a highly aggressive malignancy that is associated with a high mortality rate and poor prognosis. The 5-year survival rate for patients with pancreatic cancer is less than 10% (12). In 2021, 60,430 new pancreatic cancer cases and 48,220 pancreatic cancer-related deaths are expected in the United States (12). By 2030, pancreatic cancer is estimated to be the second most common cause of cancer-related deaths in the United States (13). The malignant progression of pancreatic cancer is often accompanied by changes in the expression of multiple genes, and the abnormal expression of specific genes may affect the prognosis of patients with pancreatic cancer. These genes may also serve as effective targets for personalized cancer therapy (14, 15). In recent years, with the rapid development of sequencing technology, high-throughput genomics has allowed for the exploration of key genes involved in cancer tumorigenesis and development. Moreover, high-throughput genomics allows for further analysis of the mechanisms related to tumorigenesis and development.

In this study, 176 TCGA-PAAD samples were genotyped based on 97 invasion-related genes, and two subtypes (C1 and C2) were obtained. The C1 subtype with a poor prognosis was more associated with the involvement of tumor-related pathways

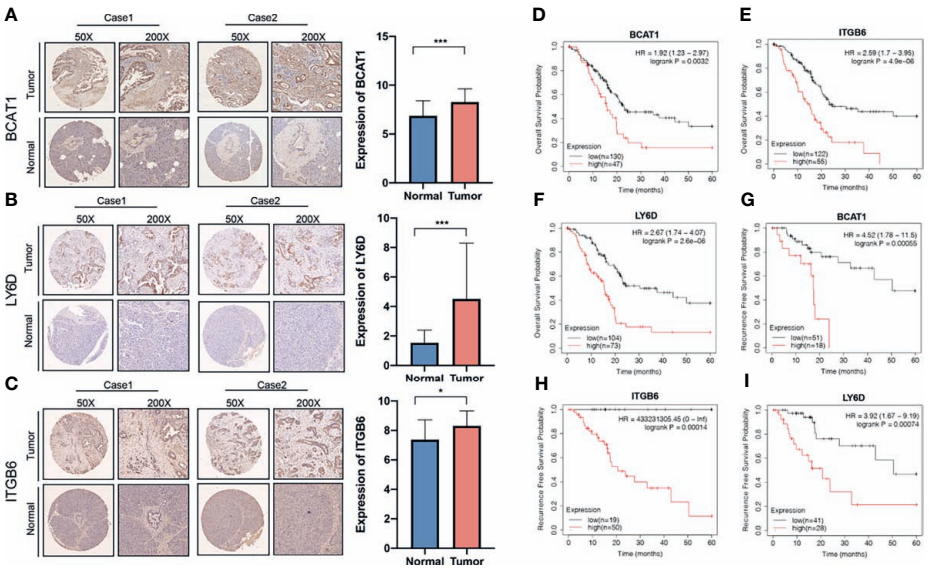


FIGURE 10 | Association of the expression of invasion-related genes with prognosis of pancreatic cancer. Expression of **(A)** *BCAT1*, **(B)** *LY6D*, and **(C)** *ITGB6* genes in pancreatic cancer and normal tissues. The relationship between the expressions of **(D)** *BCAT1*, **(E)** *LY6D*, and **(F)** *ITGB6* genes with overall survival. The relationship between the expressions of **(G)** *BCAT1*, **(H)** *LY6D*, and **(I)** *ITGB6* genes with recurrence-free survival. ***$p < 0.001$, *$p < 0.05$.

such as the Notch signaling pathway and ECM-receptor interaction. The Notch signaling pathway plays an important role in the maintenance of pancreatic tumor phenotypes (16), and the downregulation of the Notch receptor is associated with decreased proliferation, increased apoptosis, anchor-dependent growth, and decreased invasiveness of pancreatic cancer cells (17). However, matrix proteins derived from tumor cells may promote the development and metastasis of ductal adenocarcinoma of the pancreas (18). Five hundred and thirty-eight DEGs between the C1 and C2 subtypes were identified using the limma package, of which 531 genes were upregulated, and 7 genes were downregulated. We constructed a 3-gene signature using the *LY6D*, *BCAT1*, and *ITGB6* genes out of the 538 identified DEGs.

Lymphocyte 6 (Ly6) complex is a group of alloantigens, and LY6D is an important member of the Ly6 family. LY6D plays an important role in the maintenance of phenotypic and transcriptome heterogeneity of progenitor cells and the proliferation and differentiation of lymphocyte B during the early stages of lymphogenesis (19, 20). LY6D also plays an important role in cancer; it serves as a prognostic marker for advanced prostate cancer (21) and stage I non-small cell lung carcinoma (NSCLC) (22), drug resistance-associated marker for laryngeal squamous cell carcinoma (23), long-range metastasis marker for patients with ESR1-positive breast cancer (24), and a marker of urothelial and squamous cell differentiation (25). Apart from its involvement in cell adhesion, LY6D also regulates important interactions between endothelial cells and head and neck squamous cell carcinoma cells (26). In addition to glucose and fatty acid metabolism, amino acid metabolism plays an important role in tumor metabolic reprogramming.

The study has shown that the metabolism of Branched-chain amino acids (BCAA) is potentially linked with development of pancreatic ductal adenocarcinoma (27), and BCAT1, an enzyme involved in the degradation of branched-chain amino acids, is responsible for initiating the catabolism of such amino acids (28).

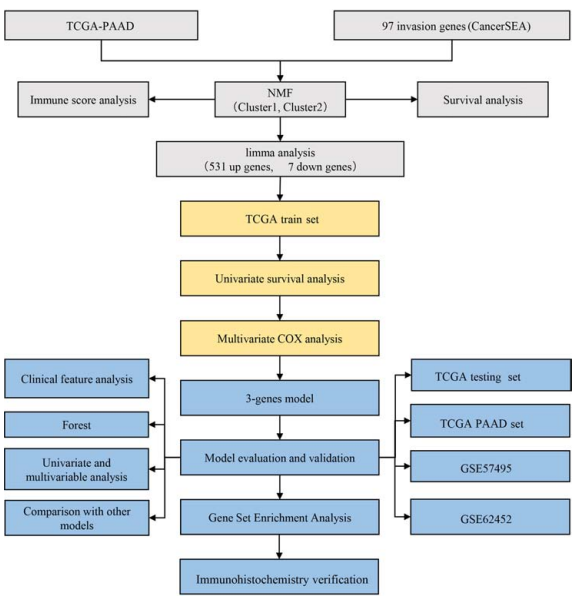


FIGURE 11 | Flow chart of research methodology.

It has been reported that pancreatic ductal adenocarcinoma cells reprogram fibroblasts to upregulate the expression of BCAT1, to meet the cancer cells' demand for branched-chain α-ketoacid (BCKAs) under BCAA deprivation (29).

The expression of BCAT1 is also upregulated in hepatocellular carcinoma (HCC) (30), breast cancer (31), and NSCLC (32), and indicates a poor prognosis. In HCC, BCAT1 plays a pathogenic role by promoting cell proliferation and chemoresistance (33). BCAT1 regulates mTOR-mediated autophagy *via* branched-chain amino acid metabolism, thus reducing the sensitivity of cancer cells to cisplatin (34).

As a member of the integrin β (ITGB) superfamily, the overexpression of ITGB6 is associated with the upregulation of the Notch signaling pathway in pancreatic cancer and is associated with immunosuppression in pancreatic cancer (35). Nine genetic markers, including ITGB6, can be used to predict the overall survival of patients with pancreatic cancer (36). ITGB6, which is highly expressed in colorectal cancer, is associated with a poor prognosis (37). ITGB6 can also be used as a tumor-specific surface antigen (TSA) to identify cell surface targets of CAR-T cell therapy and antibody-drug conjugates in breast cancer (38). Studies have shown that ITGB6 was a liver-metastasis-related gene for PAAD patients (39) and the overexpression of ITGB6 was significantly associated with advanced AJCC stage and histologic grade, and worse prognosis in pancreatic cancer (40). Our immunohistochemical results showed that LY6D, BCAT1, and ITGB6 were all overexpressed in pancreatic cancer, which was consistent with the previous results.

Although there are many multi-gene prognostic models for PAAD, there is no model based on invasion-related gene signature to predict the prognosis of pancreatic cancer. Invasion genes play an important role in metastasis as well as the development of cancer. Moreover, some prognostic signatures contain multiple genes (15-gene signature, 7-gene signature, and 6-gene signature), indicating that it is necessary to assess the expression profile of more genes in a patient-specific manner, which adds extra cost to medical care. Our 3-gene prognostic model has a higher ROC than the above models in terms of prediction of 1-, 2-, and 3-year survival rates of patients with pancreatic cancer, while having fewer genes. Therefore, our model has certain advantages in PAAD.

However, our model also presents certain limitations. First, information in the TCGA database is primarily limited to Caucasian and African populations; therefore, and data from the Asian population are missing from this study. Additionally, our study was a retrospective study of patients with pancreatic cancer, and prospective studies should be conducted to validate the prognostic characteristic and confirm the stable performance of the 3-gene prognostic model. Finally, the molecular mechanisms by which LY6D, BCAT1, and ITGB6 drive the malignant progression of pancreatic cancer require further verification.

CONCLUSIONS

In this study, we divided the TCGA-PAAD samples into two subtypes based on the differential expression of the invasion-related genes and constructed a prognostic molecular signature consisting of three genes, including *LY6D*, *BCAT1*, and *ITGB6*, based on the DEGs between the two subtypes. The *LY6D*, *BCAT1*, and *ITGB6* genes were upregulated in pancreatic cancer samples. The 3-gene prognostic model also exhibited a good AUC in both the training and validation sets. Therefore, this 3-gene prognostic model, based on the expression of three invasion-related genes, may be used to assess the prognosis of patients with pancreatic cancer. This will help in the stratification of patients for personalized cancer therapy.

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.

AUTHOR CONTRIBUTIONS

DX and YW designed the study, performed data analysis, and wrote the manuscript. YZ, ZL, and YC performed data collection. JZ supervised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Natural Science Foundation of Hainan Province (820MS130); Construction Project of Hainan Provincial Clinical Medical Research Center for malignant tumors of Digestive tract system.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.759586/full#supplementary-material

Supplementary Figure 1 | The volcano map of upregulated and downregulated DEGs between the two subtypes.

Supplementary Figure 2 | (A) For each independent variable, the horizontal axis represents the log value of the independent variable lambda, and the vertical axis represents the coefficient of the independent variable. **(B)** Confidence interval under each lambda.

REFERENCES

1. Zhou S, Ma X, Wang ZJ, Zhang WY, Jiang H, Li SD, et al. Research on the Establishment of a TPM3 Monoclonal Stable Transfected PANC-1 Cell Line and the Experiment of the EMT Occurrence in Human Pancreatic Cancer. *Onco Targets Ther* (2019) 12:5577–87. doi: 10.2147/OTT.S212689
2. Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic Cancer. *Lancet* (2020) 395(10242):2008–20. doi: 10.1016/S0140-6736(20)30974-0

3. Sinha V, Shinde S, Saxena S, Thakur S, Walia T, Dixit V, et al. A Comprehensive Review of Diagnostic and Therapeutic Strategies for the Management of Pancreatic Cancer. *Crit Rev Oncog* (2020) 25(4):381–404. doi: 10.1615/CritRevOncog.2020035971
4. Li MX, Wang HY, Yuan CH, Ma ZL, Jiang B, Li L, et al. Establishment of a Macrophage Phenotypic Switch Related Prognostic Signature in Patients With Pancreatic Cancer. *Front Oncol* (2021) 11:619517. doi: 10.3389/fonc.2021.619517
5. Wang W, Yan L, Guan X, Dong B, Zhao M, Wu J, et al. Identification of an Immune-Related Signature for Predicting Prognosis in Patients With Pancreatic Ductal Adenocarcinoma. *Front Oncol* (2020) 10:618215. doi: 10.3389/fonc.2020.618215
6. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
7. Kukreja SL, Löfberg J, Brenner MJ. A Least Absolute Shrinkage and Selection Operator (LASSO) for Nonlinear System Identification. *IFAC Proc volumes* (2006) 39:814–9. doi: 10.3182/20060329-3-AU-2901.00128
8. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
9. Chen DT, Davis-Yadley AH, Huang PY, Husain K, Centeno BA, Permuth-Wey J. Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma. *PloS One* (2015) 10(8):e0133562. doi: 10.1371/journal.pone.0133562
10. Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, et al. Identification of Candidate Diagnostic and Prognostic Biomarkers for Pancreatic Carcinoma. *EBioMedicine* (2019) 40:382–93. doi: 10.1016/j.ebiom.2019.01.003
11. Stratford JK, Bentrem DJ, Anderson JM, Fan C, Volmar KA, Marron JS, et al. A Six-Gene Signature Predicts Survival of Patients With Localized Pancreatic Ductal Adenocarcinoma. *PloS Med* (2010) 7(7):e1000307. doi: 10.1371/journal.pmed.1000307
12. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
13. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res* (2014) 74(11):2913–21. doi: 10.1158/0008-5472.CAN-14-0155
14. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer Genome Landscapes. *Science* (2013) 339(6127):1546–58. doi: 10.1126/science.1235122
15. Lucito R, Suresh S, Walter K, Pandey A, Lakshmi B, Krasnitz A, et al. Copy-Number Variants in Patients With a Strong Family History of Pancreatic Cancer. *Cancer Biol Ther* (2007) 6(10):1592–9. doi: 10.4161/cbt.6.10.4725
16. Vo K, Amarasinghe B, Washington K, Gonzalez A, Berlin J, Dang TP. Targeting Notch Pathway Enhances Rapamycin Antitumor Activity in Pancreas Cancers Through PTEN Phosphorylation. *Mol Cancer* (2011) 10:138. doi: 10.1186/1476-4598-10-138
17. Wang Z, Zhang Y, Banerjee S, Li Y, Sarkar FH. Notch-1 Down-Regulation by Curcumin is Associated With the Inhibition of Cell Growth and the Induction of Apoptosis in Pancreatic Cancer Cells. *Cancer* (2006) 106(11):2503–13. doi: 10.1002/cncr.21904
18. Tian C, Ohlund D, Rickelt S, Lidstrom T, Huang Y, Hao L, et al. Cancer Cell-Derived Matrisome Proteins Promote Metastasis in Pancreatic Ductal Adenocarcinoma. *Cancer Res* (2020) 80(7):1461–74. doi: 10.1158/0008-5472.CAN-19-2578
19. Mansson R, Zandi S, Welinder E, Tsapogas P, Sakaguchi N, Bryder D, et al. Single-Cell Analysis of the Common Lymphoid Progenitor Compartment Reveals Functional and Molecular Heterogeneity. *Blood* (2010) 115(13):2601–9. doi: 10.1182/blood-2009-08-236398
20. Borghesi L, Hsu LY, Miller JP, Anderson M, Herzenberg L, Herzenberg L, et al. B Lineage-Specific Regulation of V(D)J Recombinase Activity Is Established in Common Lymphoid Progenitors. *J Exp Med* (2004) 199 (4):491–502. doi: 10.1084/jem.20031800
21. Barros-Silva JD, Linn DE, Steiner I, Guo G, Ali A, Pakula H, et al. Single-Cell Analysis Identifies LY6D as a Marker Linking Castration-Resistant Prostate Luminal Cells to Prostate Progenitors and Cancer. *Cell Rep* (2018) 25 (12):3504–3518 e6. doi: 10.1016/j.celrep.2018.11.069
22. Lu Y, Lemon W, Liu PY, Yi Y, Morrison C, Yang P, et al. A Gene Expression Signature Predicts Survival of Patients With Stage I non-Small Cell Lung Cancer. *PloS Med* (2006) 3(12):e467. doi: 10.1371/journal.pmed.0030467
23. Wang J, Fan J, Gao W, Wu Y, Zhao Q, Chen B, et al. LY6D as a Chemoresistance Marker Gene and Therapeutic Target for Laryngeal Squamous Cell Carcinoma. *Stem Cells Dev* (2020) 29(12):774–85. doi: 10.1089/scd.2019.0210
24. Mayama A, Takagi K, Suzuki H, Sato A, Onodera Y, Miki Y, et al. OLFM4, LY6D and S100A7 as Potent Markers for Distant Metastasis in Estrogen Receptor-Positive Breast Carcinoma. *Cancer Sci* (2018) 109(10):3350–9. doi: 10.1111/cas.13770
25. Andersson N, Ohlsson J, Wahlin S, Nodin B, Boman K, Lundgren S, et al. Lymphocyte Antigen 6 Superfamily Member D is a Marker of Urothelial and Squamous Differentiation: Implications for Risk Stratification of Bladder Cancer. *biomark Res* (2020) 8:51. doi: 10.1186/s40364-020-00232-1
26. Eshel R, Zanin A, Kapon D, Sagi-Assif O, Brakenhoff R, van Dongen G, et al. Human Ly-6 Antigen E48 (Ly-6D) Regulates Important Interaction Parameters Between Endothelial Cells and Head-and-Neck Squamous Carcinoma Cells. *Int J Cancer* (2002) 98(6):803–10. doi: 10.1002/ijc.10301
27. Li JT, Yin M, Wang D, Wang J, Lei MZ, Zhang Y, et al. BCAT2-Mediated BCAA Catabolism is Critical for Development. Of Pancreatic Ductal Adenocarcinoma. *Nat Cell Biol* (2020) 22(2):167–74. doi: 10.1038/s41556-019-0455-6
28. Papathanassiu AE, Ko JH, Imprialou M, Bagnati M, Srivastava PK, Vu HA, et al. BCAT1 Controls Metabolic Reprogramming in Activated Human Macrophages and is Associated With Inflammatory Diseases. *Nat Commun* (2017) 8:16040. doi: 10.1038/ncomms16040
29. Zhu Z, Achreja A, Meurs N, Animasahun O, Owen S, Mittal A, et al. Tumour-Reprogrammed Stromal BCAT1 Fuels Branched-Chain. Ketoacid Dependency in Stromal-Rich PDAC Tumours. *Nat Metab* (2020) 2(8):775–92. doi: 10.1038/s42255-020-0226-5
30. Ji D, Jiang C, Zhang L, Liang N, Jiang T, Yang B, et al. LncRNA CRNDE Promotes Hepatocellular Carcinoma Cell Proliferation, Invasion, and Migration Through Regulating miR-203/ BCAT1 Axis. *J Cell Physiol* (2019) 234(5):6548–60. doi: 10.1002/jcp.27396
31. Zhang L, Han J. Branched-Chain Amino Acid Transaminase 1 (BCAT1) Promotes the Growth of Breast Cancer Cells Through Improving mTOR-Mediated Mitochondrial Biogenesis and Function. *Biochem Biophys Res Commun* (2017) 486(2):224–31. doi: 10.1016/j.bbrc.2017.02.101
32. Lin X, Tan S, Fu L, Dong Q. BCAT1 Overexpression Promotes Proliferation, Invasion, and Wnt Signaling in Non-Small Cell Lung Cancers. *Onco Targets Ther* (2020) 13:3583–94. doi: 10.2147/OTT.S237306
33. Zheng YH, Hu WJ, Chen BC, Grahn TH, Zhao YR, Bao HL, et al. BCAT1, a Key Prognostic Predictor of Hepatocellular Carcinoma, Promotes Cell Proliferation and Induces Chemoresistance to Cisplatin. *Liver Int* (2016) 36 (12):1836–47. doi: 10.1111/liv.13178
34. Luo L, Sun W, Zhu W, Li S, Zhang W, Xu X, et al. BCAT1 Decreases the Sensitivity of Cancer Cells to Cisplatin by Regulating mTOR-Mediated Autophagy via Branched-Chain Amino Acid Metabolism. *Cell Death Dis* (2021) 12(2):169. doi: 10.1038/s41419-021-03456-7
35. Zhuang H, Zhou Z, Ma Z, Li Z, Liu C, Huang S, et al. Characterization of the Prognostic and Oncologic Values of ITGB Superfamily Members in Pancreatic Cancer. *J Cell Mol Med* (2020) 24(22):13481–93. doi: 10.1111/jcmm.15990
36. Wu M, Li X, Zhang T, Liu Z, Zhao Y. Identification of a Nine-Gene Signature and Establishment of a Prognostic Nomogram Predicting Overall Survival of Pancreatic Cancer. *Front Oncol* (2019) 9:996. doi: 10.3389/fonc.2019.00996
37. Bengs S, Becker E, Busenhart P, Spalinger MR, Raselli T, Kasper S, et al. Beta6 -Integrin Serves as a Novel Serum Tumor Marker for Colorectal Carcinoma. *Int J Cancer* (2019) 145(3):678–85. doi: 10.1002/ijc.32137
38. Schettini F, Barbao P, Braso-Maristany F, Galvan P, Martinez D, Pare L, et al. Identification of Cell Surface Targets for CAR-T Cell Therapies and Antibody-Drug Conjugates in Breast Cancer. *ESMO Open* (2021) 6(3):100102. doi: 10.1016/j.esmoop.2021.100102
39. Dong Y, Tian J, Yan B, Lv K, Li J, Fu D. Liver-Metastasis-Related Genes are Potential Biomarkers for Predicting the Clinical Outcomes of Patients With Pancreatic Adenocarcinoma. *Pathol Oncol Res* (2021) 27:1609822. doi: 10.3389/pore.2021.1609822

40. Lenggenhager D, Bengs S, Fritsch R, Hussung S, Busenhart P, Endhardt K, et al. Beta6-Integrin Serves as a Potential Serum Marker for Diagnosis and Prognosis of Pancreatic Adenocarcinoma. *Clin Transl Gastroenterol* (2021) 12 (8):e00395. doi: 10.1111/jcmm.15990

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/fonc.2021.809035

Orphan Medicinal Products for the Treatment of Pancreatic Cancer: Lessons Learned From Two Decades of Orphan Designation

Jorn Mulder[1], Tobias van Rossum[1], Segundo Mariz[2], Armando Magrelli[3], Anthonius de Boer[1,4], Anna M. G. Pasmooij[1] and Violeta Stoyanova-Beninska[1]*

[1] Medicines Evaluation Board, Utrecht, Netherlands, [2] European Medicines Agency, Amsterdam, Netherlands, [3] National Center for Drug Research and Evaluation, Istituto Superiore di Sanità, Rome, Italy, [4] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute of Pharmaceutical Sciences, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Carlos A. Vaccaro, Italian Hospital of Buenos Aires, Argentina

Reviewed by:
Sabina Antonela Antoniu, Grigore T. Popa University of Medicine and Pharmacy, Romania
Maria Stefanova Kamusheva, Medical University of Sofia, Bulgaria

***Correspondence:**
Jorn Mulder
j.mulder@cbg-meb.nl

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *04 November 2021*
Accepted: *02 December 2021*
Published: *20 December 2021*

Citation:
Mulder J, van Rossum T, Mariz S, Magrelli A, de Boer A, Pasmooij AMG and Stoyanova-Beninska V (2021) Orphan Medicinal Products for the Treatment of Pancreatic Cancer: Lessons Learned From Two Decades of Orphan Designation. Front. Oncol. 11:809035. doi: 10.3389/fonc.2021.809035

Pancreatic cancer has a dismal prognosis and only a few treatment options are available. In the European Union, pancreatic cancer classifies as a rare disease, allowing drug developers to apply for orphan medicinal product (OMP) designation. The aim of this study was to provide more detail on OMPs for pancreatic cancer. All applications for OMP designation submitted to the EMA between 2000 and 2019 were identified. For each medicinal product that received an OMP designation, the mode of drug action, use of protocol assistance, and current life cycle status was determined. Fifty-two medicinal products received an OMP designation. At the time of submission, eighteen OMPs were at the non-clinical and 34 OMPs were at the clinical stage of development. At least fourteen kinds of mode of action were explored in the condition. For eighteen out of 52 OMPs protocol assistance was sought. At the time of data analysis, one OMP received marketing authorisation and 24 OMPs were ongoing in development. Many medicinal products for pancreatic cancer received an OMP designation and the majority of these products was already in the clinical stage of development. Nonetheless, the success rate of OMPs for pancreatic cancer that reach the market is low, and increasing this rate is something to aspire. Fortunately, development is still ongoing for a part of the OMPs, and a few developers are planning to submit a marketing authorisation application in the near future. This however does not guarantee success, as pancreatic cancer remains a difficult disease to treat. Developers are advised to make optimal use of incentives such as protocol assistance, establishing (early) dialogue between regulators and drug developers and to agree on important topics such as clinical trial design.

Keywords: orphan designation, rare disease, pancreatic cancer, european medicines agency, drug development, committee for orphan medicinal products

INTRODUCTION

Pancreatic cancer has a poor prognosis and is currently the seventh leading cause of cancer-related deaths worldwide (1). The most common type of pancreatic cancer is pancreatic ductal adenocarcinoma (2), and many patients are diagnosed when the cancer is already in the advanced stage of the disease (3). A reason for late diagnosis is that patients often do not experience any symptoms in the earlier stages of the disease (4, 5).

A few treatment options exist for patients with pancreatic cancer. Curative treatment is only optional in those that have a resectable tumour at the time of diagnosis; the minority of patients. Palliative treatment can be considered for patients with advanced or metastatic disease. Dependent on the performance status (PS) of the patient, FOLFIRINOX (PS 0 or 1), albumin-bound paclitaxel in combination with gemcitabine (PS 0 or 1) or gemcitabine monotherapy (PS 2 and/or bilirubin higher than 1.5 x upper limit normal) can be considered as a first-line treatment option, according to clinical practice guidelines (6). The only recommended second-line treatment option is liposomal irinotecan in combination with 5-fluoruracil (7). The median overall survival for first-line therapy varies between 6 and 11 months, dependent on the therapy that is administered (8). Despite available therapies, overall survival is generally poor, as reflected by the median OS being less than 1 year in patients with advanced pancreatic cancer. Hence, there is a clear unmet medical need.

According to the European Union Orphan Regulation, pancreatic cancer is classified as a rare disease (9), allowing drug developers to submit an application for orphan medicinal product (OMP) designation to the European Medicines Agency (EMA). Drug developers can submit an application for OMP designation if their product meets a couple of criteria. These criteria concern the seriousness of the disease, the prevalence of the disease, and the existence of a satisfactory method of diagnosis, prevention or treatment of the condition. Once an application is submitted to the EMA, the Committee for Orphan Medicinal Products (COMP) – one of the committees of the EMA – will examine the application. The final COMP opinion on OMP designation will be send to the European Commission (EC), and the EC decides whether the OMP designation will be granted (10). A range of incentives is offered by the EC through the Orphan Regulation. These incentives include protocol assistance (PA), fee reductions for regulatory procedures and market exclusivity (11). Protocol assistance is a kind of scientific advice specifically for OMPs (12). The aim of the Orphan Regulation is to stimulate research and development of medicinal products for rare diseases and ensure that effective medicinal products are authorised for diseases with a high unmet medical need.

To date, the COMP has approximately 20 years of experience with applications for OMP designation for pancreatic cancer. Through the years, many applications have been submitted to the EMA, and we are of the opinion that this orphan condition deserves further attention. The aim of this study was to provide a detailed overview on OMPs for pancreatic cancer, which can be of value for various stakeholders, including regulators and drug developers. Of special interest were the use of PA incentive and the current life cycle status.

METHODS

Data Sources

Internal and publicly available documents from the EMA were used in this study. Internal data was derived from EMA/COMP summary reports on applications for OMP designation, PA letters, and annual reports on designated OMPs. Publicly available data was retrieved from public summaries of positive opinion for orphan designation and European public assessment reports (EPARs); both available at www.ema.europa.eu.

Data Collection

All applications for OMP designation for medicinal products for the treatment of pancreatic cancer submitted to the COMP between 17 April 2000 and 31 December 2019 were included in this study.

From the summary reports the following information was obtained: date of submission, final COMP opinion, MoA, and stage of development at time of submission. In addition to the summary reports, information on MoAs was also obtained from public summaries. If the MoA was not clearly described in the summary report and/or public summary, literature describing the MoA was sought *via* PubMed.

PA letters were used to determine how many developers made use of this incentive and if advice on clinical development was sought.

From the annual reports the (development) status and the planned submission date was subtracted.

EPARs provided insight in the number of marketing authorisation applications (MAAs) submitted to the EMA. The time from OMP designation to Committee for Medicinal Products for Human Use (CHMP) opinion or withdrawal was determined by calculating the days between the date of the OMP designation and the date of final CHMP opinion or withdrawal of the MAA. Public summaries enabled the identification of OMPs that were withdrawn from the Community Register of designated Orphan Medicinal Products (access date: 12 March 2021).

Statistics

Descriptive statistics were used.

RESULTS

Applications for OMP Designation

Between 2000 and 2019, a total of 80 applications for OMP designation for pancreatic cancer were evaluated by the COMP. Of the 80 applications, 52 received a positive opinion on OMP designation, two received a negative opinion on OMP designation and 26 were withdrawn by the applicant prior to final COMP opinion. Seven applications were resubmitted to the agency after

the first application was withdrawn; six applications were resubmitted once and one application was resubmitted twice. Of these, six were granted positive opinion on OMP designation; these positive opinions were already included in the total number of positive opinions mentioned above. The other application resulted in a second withdrawal and eventually a negative opinion; this negative opinion was already included in the total number of negative opinions mentioned above. All medicinal products that received positive opinion by the COMP were granted OMP designation by the EC (**Supplementary Table 1**).

Simplified Mode of Action

Table 1 shows the simplified MoAs of the OMPs for pancreatic cancer. The OMPs either 'stimulate an immune response'; 'block signalling pathway(s)'; 'inhibit DNA synthesis'; 'infiltrate tumour cells and replicate therein'; 'improve the effectiveness of existing medicinal products'; 'induce DNA lesions'; 'counter migration of tumour cells'; 'induce cell cycle arrest'; 'deplete hyaluronan in tumour stroma'; 'deplete an essential amino acid required for cell growth'; 'deliver radiation specifically to tumour cells'; 'collapse mitochondrial metabolism'; 'trigger apoptosis'; or 'induces oxidative stress'. The remaining OMPs had multiple MoAs. Additional information on the MoA can be found in **Supplementary Table 1**.

Stage of Drug Development at Time of Orphan Designation

To determine which data was considered sufficient to grant OMP designation, the stage of development was identified for the 52 OMPs. At the time of submission, 18 medicinal products were at the non-clinical and 34 medicinal products were at the clinical stage of development. For the medicinal products in the non-clinical stage of development, one was investigated in an *in vitro* study and 17 were investigated in one or more *in vivo* ± *in vitro* studies (**Figure 1A**). For the medicinal products in the clinical stage of development, phase I, II and III clinical trials were ongoing/completed for 7, 25 and 2 medicinal products, respectively (**Figure 1B**).

TABLE 1 | Mode of drug action of OMPs for the treatment of pancreatic cancer.

Mode of drug action (simplified)	Number of OMPs
Stimulates an immune response	12
Blocks signalling pathway(s)	8
Inhibits DNA synthesis	5
Infects tumor cells and replicates therein	5
Improves the effectiveness of existing medicinal products	4
Multiple mechanisms	4
Induces DNA lesions	3
Counters migration of tumor cells	2
Induces cell cycle arrest	2
Delivers radiation specifically to tumour cells	2
Depletes hyaluronan in tumour stroma	1
Depletes an essential amino acid required for cell growth	1
Collapses mitochondrial metabolism	1
Triggers apoptosis	1
Induces oxidative stress	1

Use of Incentives

For 18 OMPs PA on the development of the product was sought. In total, PA was requested 23 times, including two follow-up advices and three additional advices for products for which PA was already requested previously. Nineteen of the PA requests contained questions concerning the clinical development of the OMP. Of these, twelve contained questions concerning a planned phase III trial. For four OMPs a question on a conditional marketing authorisation was included in the PA. For six OMPs a question on significant benefit was included in the PA.

Current Status of the Orphan Medicinal Products

At the time of analysis, 36 medicinal products still had an OMP designation and 16 medicinal products were withdrawn from the EC Community Register. Of the medicinal products that still had an OMP designation, 1 was authorised in the EU for the treatment of pancreatic cancer, namely Onyvide (**Figure 2**). For two OMPs, Masiviera and Orathecin, a MAA was submitted to the EMA, but these applications did not result in a marketing authorisation. For Onyvide, Masiviera and Orathecin, the time from OMP designation to final CHMP opinion or withdrawal of the MAA was 1687, 1669, and 955 days, respectively. The development status was determined for the remaining 33 OMPs. Development was ongoing for 24 OMPs, stopped for 2 OMPs and could not be determined for 7 OMPs. Development was stopped due to financial or strategic reasons. Development status was undetermined due to the absence of an annual report, while still being included in the community register.

Ongoing OMPS and Planned Submissions

A planned submission date for a MAA was included in the latest annual report for 14 out of 24 OMPs that were ongoing in development. Of the fourteen annual reports that included a planned submission date, six developers planned to submit a MAA before 2021 and eight developers planned to submit a MAA in 2021 or thereafter (**Figure 3**). The remaining sponsors did not specify a planned submission date.

DISCUSSION

To date, the COMP has two decades of experience with OMPs for pancreatic cancer, which prompted our interest in these products and their life cycle status. Through the years, a total of 52 medical products for pancreatic cancer were granted OMP designation. The major findings regarding these OMPs will be discussed in detail below.

Many of the medicinal products (65%) were already in the clinical stage of development when the developers applied for an OMP designation. This finding is however not solely confined to OMPs for the treatment of pancreatic cancer. Pauwels and colleagues revealed that the majority of anti-cancer medicinal products were in the clinical stage of development at the time of submission for OMP designation (13). Additionally, Mariz and

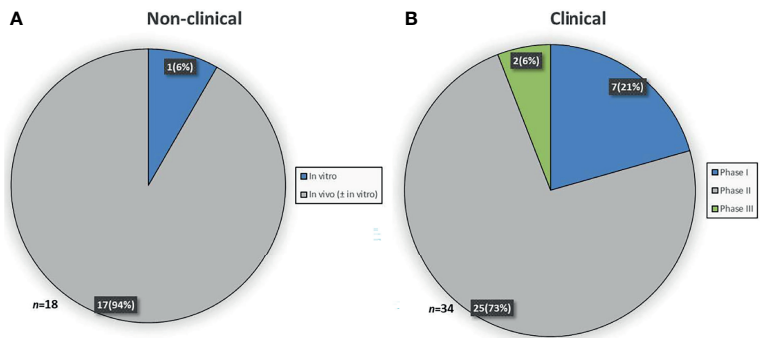


FIGURE 1 | Stage of development at time of designation. **(A)** Study(ies) conducted in the non-clinical stage of development. **(B)** Latest study ongoing or completed in the clinical stage of development.

colleagues showed that 68% of the applications for OMP designation were supported by preliminary clinical data (14). It may appear promising that many of the OMPs are already in the clinical stage of development, but it should be noted that the later stages of clinical development are often the most challenging. Hence, success cannot be guaranteed, in spite of encouraging non-clinical and preliminary clinical data. This is particularly the case for pancreatic cancer, as it is a notoriously difficult disease to treat with a high failure rate in drug development (15).

Our results show that OMPs for pancreatic cancer had distinct MoAs. The most commonly investigated OMPs included those that stimulate an immune response, block signalling pathways, infect tumour cells and replicate therein, and inhibit DNA synthesis. These OMPs can be classified as immunotherapy, targeted therapy, oncolytic virus therapy and chemotherapy, respectively. Chemotherapy continues to play an important role in the treatment of pancreatic cancer. However, other types of therapy have unfortunately not yet demonstrated definitive efficacy in pancreatic cancer, which concerns both OMPs as well as medicinal products without an OMP designation. Targeted therapy could be considered an exception, as a phase III clinical trial showed a statistically significant improvement in overall survival for erlotinib plus gemcitabine compared to gemcitabine monotherapy (16). However, the clinical relevance of this outcome is questioned, as the gain in median overall survival is approximately 2 weeks (6). There are several reasons why pancreatic cancer is such a difficult disease to treat. For instance, it is reported that a considerable part of the tumour mass is made up of a highly fibrotic stroma and this is associated with poor survival outcome (17). Furthermore, within the stroma, macrophages and inflammatory cells construct an immunosuppressive microenvironment, preventing an anti-tumour immune response (18, 19). Developing effective medicinal products remains challenging, despite the attempts to overcome these hurdles, as also seen by the MoAs of the OMPs included in this study. Therefore, a better understanding of the disease remains of importance.

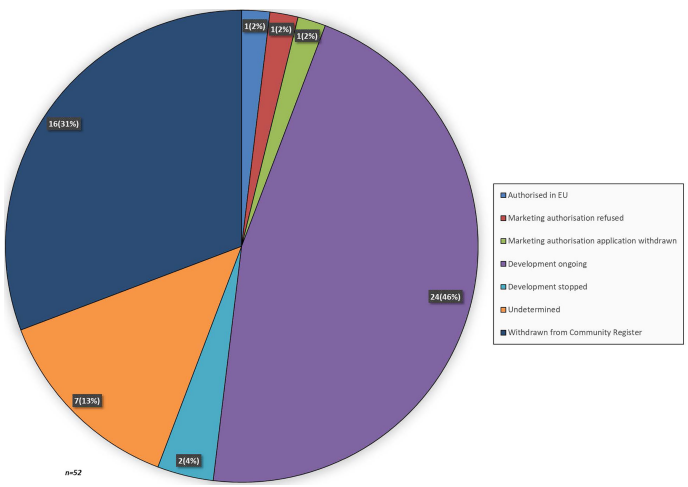


FIGURE 2 | Lifecycle status of medicinal products that received an OMP designation for pancreatic cancer. When a recent annual report was absent the development lifecycle status was labelled as undetermined.

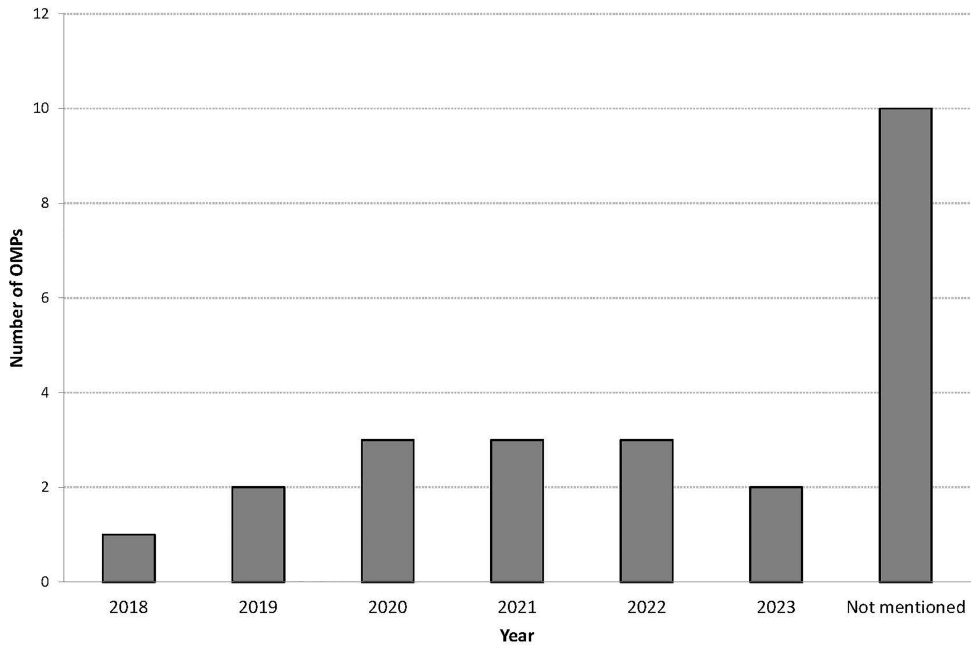


FIGURE 3 | Planned submission date for an application for MA for OMPs ongoing in development.

To stimulate the development of medicinal products for rare diseases, incentives have been implemented in the EU Orphan Drug legislation (20). We found that PA, one of these incentives, was sought only for the minority of OMPs (35%). Moreover, almost all of the PAs requests included questions on the clinical development, including questions on the design of phase III clinical trials. Hence, it appears that developers are more likely to seek PA when their product is transitioning to the late stage of clinical development. This is not surprising, as agreement between regulators and developers on the design of phase III trials – the confirmatory trial – is of importance when considering potential future MAAs. There might be several reasons why not all of the developers have requested PA, including no advancement in development, financial limitations, or lack of efficacy in previously ongoing clinical trials. Besides, developers might not be aware of the benefit of PA and hence do not make use of this incentive. An analysis performed by Hofer and colleagues showed that compliance with PA was associated with a higher probability for MA. They advised that drug developers should make use of the incentive, as the development plan could be discussed and amended. This may prevent major outstanding issues during the evaluation of a MAA (21). Therefore, it remains important that developers continue to seek PA, considering the benefit of this incentive.

Even though the majority of medicinal products was already in the clinical stage of development when the developers applied for OMP designation, to date, only one OMP for pancreatic cancer received MA, namely irinotecan hydrochloride trihydrate (22). Irrespective of orphan condition, the success rate of medical products that reach the market as OMPs is estimated to be 8% (23), which is four times higher than our finding. These data highlight that – despite the efforts of developers – not many OMPs eventually will reach the market, especially not those for pancreatic cancer. Nonetheless, the lower success rate is of course related to the difficulties in treating the condition. This is further highlighted by the fact that the CHMP was of the opinion that the benefit-risk balance was not considered positive for two out of three OMP for pancreatic cancer considered for MA, namely rubitecan and masitinib (24, 25). This resulted in a withdrawal of MA application and a refusal on MA, respectively.

A positive finding in our results is that development is still ongoing for almost half of the OMPs (46%), and a couple of developers are even planning to submit an application for MA in the near future. Of all these, a few developers indicated their plans to submit a MAA in previous years, but this did not happen so far. The reasons for this might be delayed of failed development. For the remaining OMPs it could not be determined whether development is still ongoing, as the annual reports were absent or OMPs were withdrawn from the Community Register. It remains difficult to speculate on the reasons behind this, but plausible reasons could be either failure in development or financial considerations. At least for those products that have received an OMP designation a while ago.

This study has a few limitations, one of which is the lack of correction for time. For example, some medicinal products have received an OMP designation recently, while others have received OMP designation years ago. Therefore, products that have recently been granted OMP designation might still face potential developmental challenges in the future. Another limitation is the incompleteness of our overview on status of drug development,

which is due to the lack of (recent) annual reports for a part of the OMPs. Determining whether the OMP is still in the drug pipeline of the developer would provide a more definitive answer on the life cycle status than is currently provided in our study.

CONCLUSION

The success rate of medical products for pancreatic cancer that reach the market as OMPs is lower than for OMPs in general and increasing this success rate is something to aspire. Despite that pancreatic cancer is such a difficult disease to treat, a substantial number of applications has been submitted to the EMA for this condition, which indicates interest among drug developers. Development is still ongoing for a part of the OMPs, and for a few of these OMPs a submission for MAA in planned in the near future. It should be reminded that an OMP designation is supported by promising non-clinical and/or preliminary clinical data, but efficacy and safety still needs to be determined and the late stages of development are often the most challenging. Therefore an OMP designation is not a guarantee for a successful MA. In this respect, developers are advised to make optimal use of incentives inherent with an OMP designation, such as PA, establishing (early) dialogue between regulators and drug developers to agree on important topics such as clinical trial design. In addition, developers are strongly encouraged to provide yearly updates on advancements in development. Close monitoring of the drug development through the annual reports and transparency regarding the reason(s) for stopping development are crucial for saving human and financial resources and redirecting efforts in promising concepts.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions, the data that support the findings of this study are only available from the EMA through a confidentiality agreement which the authors signed. Requests to access these datasets should be directed to science@cbg-meb.nl.

AUTHOR CONTRIBUTIONS

JM, AP, and VS-B designed the study. JM and TV performed the research. JM, AP, and VS-B analysed the data. JM, TV, SM, AM, AD, AP, and VS-B wrote the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Kristina Larsson for the insightful discussion on potential research questions for this study and providing relevant EMA documents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.809035/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Rishi A, Goggins M, Wood LD, Hruban RH. Pathological and Molecular Evaluation of Pancreatic Neoplasms. *Semin Oncol* (2015) 42:28–39. doi: 10.1053/j.seminoncol.2014.12.004
3. Freelove R, Walling AD. Pancreatic Cancer: Diagnosis and Management. *Am Fam Physician* (2006) 73:485–92.
4. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic Cancer. *Lancet* (2016) 388:73–85. doi: 10.1016/S0140-6736(16)00141-0
5. Lancet Oncology. Pancreatic Cancer: A Neglected Killer? *Lancet Oncol* (2010) 11:1107. doi: 10.1016/S1470-2045(10)70281-3
6. Ducreux M, Cuhna AS, Caramella C, Hollebecque A, Burtin P, Goéré D, et al. Cancer of the Pancreas: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2015) 26:v56–68. doi: 10.1093/annonc/mdv295
7. *Eupdate – Cancer of the Pancreas Treatment Recommendations*. Available at: https://www.esmo.org/guidelines/gastrointestinal-cancers/pancreatic-cancer/eupdate-cancer-of-the-pancreas-treatment-recommendations2.
8. Neoptolemos JP, Kleeff J, Michl P, Costello E, Greenhalf W, Palmer DH. Therapeutic Developments in Pancreatic Cancer: Current and Future Perspectives. *Nat Rev Gastroenterol Hepatol* (2018) 15:333–48. doi: 10.1038/s41575-018-0005-x
9. *Community Register of Orphan Medicinal Products*. Available at: https://ec.europa.eu/health/documents/community-register/html/reg_od_act.htm?sort=n.
10. European Medicines Agency. *Orphan Designation: Overview*. Available at: https://www.ema.europa.eu/en/human-regulatory/overview/orphan-designation-overview.
11. Fregonese L, Greene L, Hofer M, Magrelli A, Naumann-Winter F, Larsson K, et al. Demonstrating Significant Benefit of Orphan Medicines: Analysis of 15 Years of Experience in Europe. *Drug Discovery Today* (2018) 23:90–100. doi: 10.1016/j.drudis.2017.09.010
12. European Medicines Agency. *Orphan Incentives*. Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/orphan-designation/orphan-incentives.
13. Pauwels K, Huys I, Casteels M, Larsson K, Voltz C, Penttila K, et al. Are Products With an Orphan Designation for Oncology Indications Different From Products for Other Rare Indications? A Retrospective Analysis of European Orphan Designations Granted Between 2002-2012. *Orphanet J Rare Dis* (2017) 12:36. doi: 10.1186/s13023-017-0578-4
14. Mariz S, Reese JH, Westermark K, Greene L, Goto T, Hoshino T, et al. Worldwide Collaboration for Orphan Drug Designation. *Nat Rev Drug Discov* (2016) 15:440–1. doi: 10.1038/nrd.2016.80
15. Thota R, Maitra A, Berlin JD. Preclinical Rationale for the Phase III Trials in Metastatic Pancreatic Cancer: Is Wishful Thinking Clouding Successful Drug Development for Pancreatic Cancer? *Pancreas* (2017) 46:143–50. doi: 10.1097/MPA.0000000000000753
16. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, et al. Erlotinib Plus Gemcitabine Compared With Gemcitabine Alone in Patients With Advanced Pancreatic Cancer: A Phase III Trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* (2007) 25:1960–6. doi: 10.1200/JCO.2006.07.9525
17. Kokkinos J, Mistica RMC, Sharbeen G, Boyer C, Gonzales-Aloy E, Goldstein D, et al. Targeting the Undruggable in Pancreatic Cancer Using Nano-Based

Gene Silencing Drugs. *Biomaterials* (2020) 240:119742. doi: 10.1016/j.biomaterials.2019.119742
18. Gore J, Korc M. Pancreatic Cancer Stroma: Friend or Foe? *Cancer Cell* (2014) 25:711–2. doi: 10.1016/j.ccr.2014.05.026
19. Oberstein PE, Olive KP. Pancreatic Cancer: Why Is it So Hard to Treat? *Therap Adv Gastroenterol* (2013) 6:321–37. doi: 10.1177/1756283X13478680
20. Westermark K, Holm BB, Söderholm M, Llinares-Garcia J, Rivière F, Aarum S, et al. European Regulation on Orphan Medicinal Products: 10 Years of Experience and Future Perspectives. *Nat Rev Drug Discov* (2011) 10:341–9. doi: 10.1038/nrd3445
21. Hofer MP, Hedman H, Mavris M, Koenig F, Vetter T, Posch M, et al. Marketing Authorisation of Orphan Medicines in Europe From 2000 to 2013. *Drug Discov Today* (2018) 23:424–33. doi: 10.1016/j.drudis.2017.10.012
22. European Medicines Agency. *Assessment Report Onivyde*. European Medicines Agency (2016). Available at: https://www.ema.europa.eu/en/documents/assessment-report/onivyde-epar-public-assessment-report_en.pdf.
23. European Medicines Agency. *Rare Diseases, Orphan Medicines*. Available at: https://www.ema.europa.eu/en/documents/other/rare-diseases-orphan-medicines-getting-facts-straight_en.pdf.
24. European Medicines Agency. *Withdrawal Assessment Report Orathecin*. European Medicines Agency (2007). Available at: https://www.ema.europa.eu/en/documents/withdrawal-report/withdrawal-assessment-report-orathecin_en.pdf.
25. European Medicines Agency. *Assessment Report Masiviera*. European Medicines Agency (2014). Available at: https://www.ema.europa.eu/en/documents/assessment-report/masiviera-epar-public-assessment-report_en.pdf.

Author Disclaimer: The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of, or reflecting the position of the EMA or one of its committees, working parties, or any of the national agencies.

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fonc.2022.835581

Case Report Series: Aggressive HR Deficient Colorectal Cancers Related to BRCA1 Pathogenic Germline Variants

*Maria Valeria Freire[1], Marie Martin[1], Romain Thissen[1], Cédric Van Marcke[2,3], Karin Segers[1], Edith Sépulchre[1], Natacha Leroi[1], Céline Lété[1], Corinne Fasquelle[1], Jean Radermacher[4], Yeter Gokburun[5], Joelle Collignon[6], Anne Sacré[7], Claire Josse[6], Leonor Palmeira[1] and Vincent Bours[1]**

[1] Department of Human Genetics, GIGA Research Center – University of Liège and Centre Hospitalier Universitaire (CHU) Liège, Liège, Belgium, [2] Institute for Experimental and Clinical Research (Institut de Recherche Expérimentale et Clinique (IREC), Pôle Molecular Imaging, Radiotherapy and Oncology (MIRO)), Université Catholique de Louvain (UCLouvain), Brussels, Belgium, [3] Department of Medical Oncology, Institut Roi Albert II, Cliniques Universitaires Saint-Luc, Brussels, Belgium, [4] Department of Pathology, Institut de Pathologie et de Génétique, Charleroi, Belgium, [5] Department of Gastroenterology, Centre Hospitalier Régional Sambre et Meuse, Namur, Belgium, [6] Department of Medical Oncology, GIGA Research Center – University of Liège and Centre Hospitalier Universitaire (CHU) Liège, Liège, Belgium, [7] Onco-Hematology Department, Centre Hospitalier Régional (CHR) Verviers, Verviers, Belgium

Objective: The link between *BRCA1* and homologous recombination deficiency (HRD) in cancer has gained importance with the emergence of new targeted cancer treatments, while the available data on the role of the gene in colorectal cancer (CRC) remain contradictory. The aim of this case series was to elucidate the role of known pathogenic *BRCA1* variants in the development of early-onset CRC.

Design: Patients were evaluated using targeted next generation sequencing, exome sequencing and chromosomal microarray analysis of the paired germline and tumor samples. These results were used to calculate the HRD score and the frequency of mutational signatures in the tumors.

Results: Three patients with metastatic CRC were heterozygous for a previously known *BRCA1* nonsense variant. All tumors showed remarkably high HRD scores, and the HRD-related signature 3 had the second highest contribution to the somatic pattern of variant accumulation in the samples (23% in 1 and 2, and 13% in sample 3).

Conclusions: A *BRCA1* germline pathogenic variant can be involved in CRC development through HRD. Thus, *BRCA1* testing should be considered in young patients with a personal history of microsatellite stable CRC as this could further allow a personalized treatment approach.

Keywords: colorectal (colon) cancer, *BRCA1*, homologous recombination deficiency (HRD), exome sequencing (ES), case report

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Moom Roosan, Chapman University, United States
Natalija Dedić Plavetić University Hospital Centre Zagreb, Croatia

***Correspondence:**
Vincent Bours
vbours@uliege.be

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *14 December 2021*
Accepted: *01 February 2022*
Published: *24 February 2022*

Citation:
Freire MV, Martin M, Thissen R, Van Marcke C, Segers K, Sépulchre E, Leroi N, Lété C, Fasquelle C, Radermacher J, Gokburun Y, Collignon J, Sacré A, Josse C, Palmeira L and Bours V (2022) Case Report Series: Aggressive HR Deficient Colorectal Cancers Related to BRCA1 Pathogenic Germline Variants. Front. Oncol. 12:835581. doi: 10.3389/fonc.2022.835581

INTRODUCTION

BRCA1 is a tumor suppressor gene encoding a large protein that coordinates several cellular pathways including DNA repair, transcriptional regulation, cell-cycle control, centrosome duplication, and apoptosis (1). Pathogenic germline variants in *BRCA1* gene have been associated with familial risk of breast and ovarian cancers (OMIM: 604370) (2, 3). As early as in 1994, it was observed that women with a history of breast, endometrial, or ovarian cancer presented a statistically significant although small risk for subsequent colorectal cancer (CRC), suggesting the existence of common etiologic factors for the development of these tumors (4).

Data concerning young patients with *BRCA1* variants that develop CRC have been scarce. Germline pathogenic variants in *BRCA1* gene have not been causally linked to an increased risk of familial colorectal cancer, but the reports on the subject are contradictory (5–9). Indeed, patients carrying a germline *BRCA1* variant can develop a sporadic tumor, independently of *BRCA1* loss of function, highlighting the need to demonstrate the causal role of the variant in the cancer development (10).

The aim of this case series was to gain insight into the role of known pathogenic *BRCA1* variants in the development of early-onset CRC.

CASE DESCRIPTION

Three patients were diagnosed in 2020 and 2021 with aggressive early-onset CRC. The demographic, familial, clinical, histopathological, and molecular characteristics, as well as the treatment regimens of these patients are presented in **Table 1**.

The first case was referred to oncogenetic consultation due to the young age of presentation of an aggressive disease without evidence of Lynch syndrome (no mismatch repair deficiency or microsatellite instability) and history of a *BRCA1* pathogenic

TABLE 1 | Patient characteristics.

Parameter	Case 1	Case 2	Case 3
Age (year)	31	56	35
Sex	Female	Female	Male
Medical history	None	Breast cancer at 36 y/o, contralateral breast cancer at 41 y/o	Ulcerative colitis
Family history	Maternal side: aunt breast cancer, grandmother CRC, great-grandmother uterine cancer	Paternal side: aunt CRC, grandmother ovary cancer, grandmother's sister breast cancer	Maternal side: five aunts breast cancer, grandmother ovary cancer. Paternal side: grandmother CRC
CRC localization and type	Right colon moderately differentiated adenocarcinoma	Well to moderately differentiated rectum adenocarcinoma	Mucinous appendix adenocarcinoma
TNM tumor staging	pT4aN2aM1a	cT3N1M1b	pT4bN0M0 at diagnosis, peritoneal relapse at month 5
IHC and molecular tests on the tumor			
MSI-H	No	No	No
MLH1, MSH2, MSH6 and PMS2 protein expression	Normal	Normal	Normal
Identified variants			
Somatic pathogenic variants (heterozygous)	*KRAS* c.35G>A (p. Gly12Asp) *TP53* c.524G>A (p.Arg175His)	*KRAS* c.35G>A (p. Gly12Asp)	–
Germline pathogenic variants (heterozygous)	*BRCA1* c.1016dup (p.Val340Glyfs*6)	*BRCA1* c.3756_3759del (p.Ser1253Argfs*10)	*BRCA1* c.3841C>T (p.Gln1281*)
Somatic CMA array results	Partial gains and losses on Chr 1-3, 5- 9, 12, 13, 15-20 and X	Entire and partial gains and losses on Chr 1, 7, 8, 12, 13 and 18-20	Normal
HRD evaluation			
HRD score	59	61.15	66
Proportions of mutational signatures with a proposed etiology			
SBS1	24%	28%	14%
SBS3	23%	22%	13%
SBS5	0%	0%	20%
Treatment			
Surgical	Right colectomy with lymph node dissection, ileocolonic anastomosis and metastasectomy of liver segments	Anterior rectum resection and hepatic surgery	Ileocolectomy with a lymph node dissection firstly and a posterior debulking surgery with IPCH after discovery of a peritoneal carcinomatosis
Chemotherapy	Pseudo-adjuvant chemotherapy with capecitabine-oxaliplatin followed by 7 cycles of chemotherapy with FOLFOX-bevacizumab	6 cycles of FOLFOXIRI	Adjuvant chemotherapy with capecitabine-oxaliplatin regimen (Xelox)

CMA, chromosomal microarray analysis; FOLFOX, folinic acid; fluorouracil and oxaliplatin; FOLFOXIRI, fluorouracil; folinic acid; oxaliplatin; and irinotecan; HRD, homologous recombination deficiency; IHC, Immunohistochemistry; IPCH, Intraperitoneal chemohyperthermia; MMR, mismatch repair; MSI, microsatellite instability; SBS, Single Base Substitution; TNM, TNM Classification of Malignant Tumors; y/o, years old. Reference transcripts: BRCA1 NM_007294.3; KRAS NM_004985.5; TP53 NM_000546.6.

variant in the family. Given the age at the diagnosis of CRC, genes associated with familial polyposis (*NTHL1, RNF43, SMAD4, BMPR1A*), CRC (*POLE, POLD1*) and Li-Fraumeni syndrome (*TP53*) were analyzed. However, the patient only carried the heterozygous *BRCA1* pathogenic variant NM_007294.3(BRCA1_v001):c.1016dup (p.Val340Glyfs*6) identified in her maternal aunt.

To further evaluate the disease, targeted next-generation sequencing (NGS) and a high-resolution (180K) chromosomal microarray analysis (CMA) were performed on the DNA extracted from the tumor (estimated proportion of tumor cells in the sample - 50%). After sequencing, the familial pathogenic variant *BRCA1* c.1016dup was identified at an allele frequency (AF) of 70%, suggesting a loss of heterozygosity at the *BRCA1* locus. Further analysis revealed a somatic variant of *TP53* NM_000546.6(TP53):c.524G>A (p.Arg175His) at an AF of 40%. The CMA showed multiple rearrangements indicating genomic instability (chromosomal partial gains and losses on chromosomes 1-3, 5- 9, 12, 13, 15-20 and X).

The personal and family history of cancer in case 2 already led in 2011 to the identification of the pathogenic *BRCA1* germline variant NM_007294.3(BRCA1_v001):c.3756_3759del (p.Ser1253Argfs*10). Taking this information into account, a CMA and NGS of the tumor DNA (estimated tumor infiltration – 30%) were performed, identifying the known germline *BRCA1* variant with an AF of 35% and an additional NM_004985.5 (KRAS_v001):c.35G>A variant with an AF of 23%. The CMA results were monosomies 18 and 19, trisomies 1q, 7, 8, 12, 13 and 20, partial chromosomal losses in the 1p region and partial chromosomal gains in the 1p region.

In case 3, CRC was diagnosed from a surgical specimen obtained after an appendectomy with the subsequent identification of a tumor-like lesion with low-grade dysplasia at the base of the cecum. Considering that the patient's mother carried a *BRCA1* germline variant, the patient DNA was tested, confirming the presence of the heterozygous BRCA1 pathogenic variant NM_007294.3(BRCA1_v001):c.3841C>T (p.Gln1281*). Subsequently, *BRCA1* sequencing and CMA array on tumor DNA (sample estimated tumor infiltration – 20%) showed the *BRCA1* c.3841C>T family variant with an AF of 43%, while the CMA was normal.

The three variants are predicted to cause truncation of the translation in exon 10 (out of a total of 23) which will result in a severely shortened or absent protein due to nonsense-mediated decay of the mRNA. *BRCA1* protein truncations downstream of this position have been described as pathogenic (11, 12). *BRCA1* c.1016dup and *BRCA1* c.3841C>T variants were absent in 251174 control chromosomes in gnomAD, whereas *BRCA1* c.3756_3759del was present at an AF of 1.267e-05. *BRCA1* c.1016dupA has been reported in the literature as a founder variant in Norway and Canada (13, 14) and also in multiple individuals affected with hereditary breast and ovarian cancer syndrome in other populations (15–18). Case 2 four-nucleotide deletion was widely reported in the literature in Polish and French-Canadian gynecological cancer patients (19, 20). The *BRCA1* variant present in case 3 has been reported as a France, Belgium, and Holland founder variant (21). ClinVar submitters including an expert panel (ENIGMA) cite the three variants as pathogenic. These data indicate that the three variants are highly likely to be associated with high breast and ovarian cancer risk.

Homologous recombination deficiency (HRD) evaluation can be performed using HRD score, an aggregate score of loss of heterozygosity (LOH), telomeric-allelic imbalance (TAI) and large-scale state transitions (LST). To confirm the HRD score in the CRC samples we used an alternative method of HRD detection by investigating single base substitution (SBS) signatures.

To assess homologous recombination deficiency (HRD) in CRC samples, a paired germline and tumoral DNA exome sequencing using Twist Comprehensive Exome Panel and Twist Human RefSeq Panel (according to the manufacturer's instructions) from all three patients was performed. We used Sequenza (22) to detect and quantify copy number variation and estimate tumor cellularity and ploidy. These results were used as an input to calculate the HRD score with a threshold of positivity ≥33 (23). Mutational signatures in the samples were analyzed Using MutationalPatterns R package (24) and COSMIC v2 signatures (25), taking only the somatic variants into account.

Through Sequenza, the estimated tumor cellularity was of 95% in the case 3 sample, while this value was lower for cases 1 and 2 – 22% and 27%, respectively. All three samples showed remarkably high HRD scores (59, 61.15 and 66, respectively), while no somatic copy number alteration was identified in *PALB2, BRCA1* and *BRCA2*.

The three most frequent SBS signatures with a proposed etiology in the samples were SBS1, 3 and 5 (see **Figure 1**). Signature 3 was the second most frequent signature with a contribution to 23% of the somatic pattern of variant accumulation in samples 1 and 2, and 13% in sample 3. While signature 1 and 5 reflect clock-like accumulation of somatic variants, signature 3 has been directly related to HRD (25).

DISCUSSION

The existing data linking germline pathogenic variants in the *BRCA1* gene to an increased risk of CRC are scarce. Two large studies reported that *BRCA1* variants conferred approximatively a fivefold increased risk for CRC, especially in young patients from high-risk families (6, 26). Out of three recent meta-analyses, one of them found an increased risk of colorectal cancer associated with *BRCA1* variants (odds ratio = 1.49, 95% CI = 1.19 to 1.85, $P < 0.001$) (8), while the other two did not identify any increase in CRC risk among patients carrying a *BRCA1* variant (7, 9). A study evaluating a cohort of *BRCA1* or *BRCA2* pathogenic variant carriers mostly of Ashkenazy ancestry concluded that they may be prone to developing anal carcinoma and left-sided mucinous histology CRC (27). One single publication reported a young male patient with a *BRCA1* germinal variant who presented with rectal adenocarcinoma

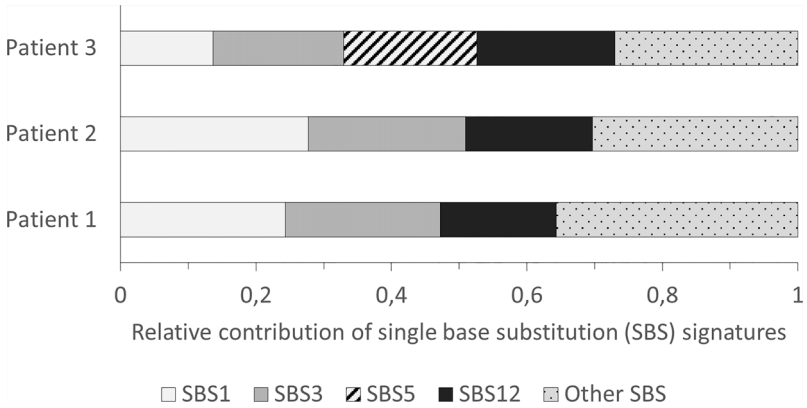


FIGURE 1 | Single base substitution (SBS) signatures identified in the patients with a relative contribution ≥ 15% of the global pattern.

and showed an excellent response to oxaliplatin-containing neoadjuvant therapy (28). These data thus remain contradictory and do not allow to recommend to screen for CRC in BRCA1 variants heterozygotes, or to consider *BRCA1* pathogenic variants as a factor predisposing to familial CRC.

Given the frequency of CRC and of *BRCA1* variant heterozygotes in European populations (29), co-occurrence may be incidental rather than indicative of a causal relationship, as suggested previously (30). However, a few lines of evidence indicate that co-occurrence might be relevant.

Recently, a large report investigated the frequencies of various cancers, including CRCs, in 6902 men with BRCA variants (31). The probability for developing a CRC was, according to this report, two times lower in men with *BRCA2* variants than in *BRCA1* variant heterozygotes. As it seems unlikely that *BRCA2* variants had a protective role against CRCs, these data could indicate a slightly but significantly increased risk of these cancers in men with *BRCA1* variants.

In our samples, we did not evaluate BRCA1 protein expression. Although we describe patients with aggressive metastatic cancer, the presence of low levels of BRCA1 protein had a worse prognosis even in early-stage CRC (32).

In our study, we not only confirmed that the *BRCA1* germline variants were still present in the tumor (with evidence of positive selection in case 1), but we also demonstrate scars of HRD in the three tumors. Indeed, the presence of germline variants in HRD-associated genes alone is not sufficient to predict clinically relevant HRD. We highlighted the presence of specific mutational signatures (COSMIC signature 3) (33) and genomic instability characteristics (LOH, TAI and LST) (34–36), reflecting significant HRD, comparable with that observed in ovarian cancers with a *BRCA1* or *BRCA2* pathogenic variants. Interestingly, the initial somatic NGS analysis of cases 2 and 3 was not conclusive, possibly because of low tumor infiltration, but it could also be indicative of an epigenetic event leading to loss of *BRCA1* function and demonstrates the role of HRD testing even in cases where the mechanism driving HRD is not fully elucidated. Taken together, these observations indicate that germline *BRCA1* variants may, in a small proportion of variant carriers, play a driver role in CRC development or progression and that these patients might thus benefit from a treatment with poly (ADP-ribose) polymerase-inhibitors (PARPi). Indeed, clinical trials clearly demonstrated the efficacy of platinum-based chemotherapy and PARPi to treat BRCA mutated and/or HRD positive cancers inside the spectrum of BRCA-related cancers (37). Further evidence demonstrating that some CRC could be linked to BRCA deficiencies could open new perspectives for treatment with PARPi of these rare aggressive tumors.

The small number of patients and the bias in recruitment are the main limitations of our study, precluding to justify any specific surveillance or screening program in the absence of a personal or family history.

In conclusion, our data indicate that a *BRCA1* germline pathogenic variant can be involved in CRC development through HRD. Thus, *BRCA1* testing should be considered in young patients with a personal history of microsatellite stable CRC. This could further allow a personalized treatment approach with a PARPi.

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 "Comité d'Ethique Hospitalo-facultaire Universitaire de Liège" (CHU/University of Liège) approved the study. 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

Conceptualization: VB. Data curation: MF, MM, and VB. Formal analysis: MF. Funding acquisition: VB. Investigation: MF, MM, RT, CM, KS, ES, NL, CL, and CF. Methodology: MF, MM, RT, CM, CJ, and LP. Project administration: VB. Resources: VB. Supervision: CJ, LP, and VB. Validation: JR, YG, JC, and AS. Writing-original draft: MF and VB. Writing-review and editing: MF, MM, RT, CM, KS, ES, NL, CL, JR, YG, JC, AS, CJ, LP, VB, and CF. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a Télévie fellowship (MV.F., grant N° 7451419F), a grant from the CHU Liége (N° 981481200) and the WALGEMED grant (Région Wallonne, grant N° 1710180).

ACKNOWLEDGMENTS

The CHU of Liége is member of the GENTURIS (GENetic TUmor RISk) European Reference Network).

REFERENCES

1. Wu W, Koike A, Takeshita T, Ohta T. The Ubiquitin E3 Ligase Activity of BRCA1 and Its Biological Functions. *Cell Div* (2008) 3:1. doi: 10.1186/1747-1028-3-1
2. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene Brca1. *Science* (1994) 266(5182):66–71. doi: 10.1126/science.7545954
3. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, et al. Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. *J Natl Cancer Inst* (2013) 105(11):812–22. doi: 10.1093/jnci/djt095
4. Schoen RE, Weissfeld JL, Kuller LH. Are Women With Breast, Endometrial, or Ovarian Cancer at Increased Risk for Colorectal Cancer? *Am J Gastroenterol* (1994) 89(6):835–42.
5. Mersch J, Jackson M, Park M, Nebgen D, Peterson SK, Singletary C, et al. Cancers Associated With BRCA1 and BRCA2 Mutations Other Than Breast and Ovarian. *Cancer* (2015) 121(2):269–75. doi: 10.1002/cncr.29041
6. Phelan CM, Iqbal J, Lynch HT, Lubinski J, Gronwald J, Moller P, et al. Incidence of Colorectal Cancer in BRCA1 and BRCA2 Mutation Carriers: Results From a Follow-Up Study. *Br J Cancer* (2014) 110(2):530–4. doi: 10.1038/bjc.2013.741
7. Cullinane CM, Creavin B, O'Connell EP, Kelly L, O'Sullivan MJ, Corrigan MA, et al. Risk of Colorectal Cancer Associated With BRCA1 and/or BRCA2 Mutation Carriers: Systematic Review and Meta-Analysis. *BJS (British J Surgery)* (2020) 107(8):951–9. doi: 10.1002/bjs.11603
8. Oh M, McBride A, Yun S, Bhattacharjee S, Slack M, Martin JR, et al. BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-Analysis. *J Natl Cancer Inst* (2018) 110(11):1178–89. doi: 10.1093/jnci/djy148
9. Lee Y-C, Lee Y-L, Li C-Y. BRCA Genes and Related Cancers: A Meta-Analysis From Epidemiological Cohort Studies. *Medicina* (2021) 57(9):905. doi: 10.3390/medicina57090905
10. Curtit E, Benhamo V, Gruel N, Popova T, Manie E, Cottu P, et al. First Description of a Sporadic Breast Cancer in a Woman With BRCA1 Germline Mutation. *Oncotarget* (2015) 6(34):35616–24. doi: 10.18632/oncotarget.5348
11. Song H, Cicek MS, Dicks E, Harrington P, Ramus SJ, Cunningham JM, et al. The Contribution of Deleterious Germline Mutations in BRCA1, BRCA2 and the Mismatch Repair Genes to Ovarian Cancer in the Population. *Hum Mol Genet* (2014) 23(17):4703–9. doi: 10.1093/hmg/ddu172
12. de Juan Jiménez I, García Casado Z, Palanca Suela S, Esteban Cardeñosa E, López Guerrero JA, Segura Huerta Á, et al. Novel and Recurrent BRCA1/BRCA2 Mutations in Early Onset and Familial Breast and Ovarian Cancer Detected in the Program of Genetic Counseling in Cancer of Valencian Community (Eastern Spain). Relationship of Family Phenotypes With Mutation Prevalence. *Familial Cancer* (2013) 12(4):767–77. doi: 10.1007/s10689-013-9622-2
13. Simard J, Tonin P, Durocher F, Morgan K, Rommens J, Gingras S, et al. Common Origins of BRCA1 Mutations in Canadian Breast and Ovarian Cancer Families. *Nat Genet* (1994) 8(4):392–8. doi: 10.1038/ng1294-392
14. Dørum A, Heimdal K, Hovig E, Inganäs M, Møller P. Penetrances of BRCA1 1675dela and 1135insa With Respect to Breast Cancer and Ovarian Cancer. *Am J Hum Genet* (1999) 65(3):671–9. doi: 10.1086/302530
15. Foretova L, Machackova E, Navratilova M, Pavlu H, Hruba M, Lukesova M, et al. BRCA1 and BRCA2 Mutations in Women With Familial or Early-Onset Breast/Ovarian Cancer in the Czech Republic. *Hum Mutat* (2004) 23(4):397–8. doi: 10.1002/humu.9226
16. Machackova E, Foretova L, Lukesova M, Vasickova P, Navratilova M, Coene I, et al. Spectrum and Characterisation of BRCA1 and BRCA2 Deleterious Mutations in High-Risk Czech Patients With Breast and/or Ovarian Cancer. *BMC Cancer* (2008) 8:140. doi: 10.1186/1471-2407-8-140
17. Laraqui A, Uhrhammer N, Lahlou-Amine I, EL Rhaffouli H, El Baghdadi J, Dehayni M, et al. Mutation Screening of the BRCA1 Gene in Early Onset and Familial Breast/Ovarian Cancer in Moroccan Population. *Int J Med Sci* (2012) 10(1):60–7. doi: 10.7150/ijms.5014
18. Janavičius R. Founder BRCA1/2 Mutations in the Europe: Implications for Hereditary Breast-Ovarian Cancer Prevention and Control. *EPMA J* (2010) 1(3):397–412. doi: 10.1007/s13167-010-0037-y
19. Behl S, Hamel N, de Ladurantaye M, Lepage S, Lapointe R, Mes-Masson A-M, et al. Founder BRCA1/BRCA2/PALB2 Pathogenic Variants in French-Canadian Breast Cancer Cases and Controls. *Sci Rep* (2020) 10:6491. doi: 10.1038/s41598-020-63100-w
20. Łukomska A, Menkiszak J, Gronwald J, Tomiczek-Szwiec J, Szwiec M, Jasiówka M, et al. Recurrent Mutations in BRCA1, BRCA2, RAD51C, PALB2 and CHEK2 in Polish Patients With Ovarian Cancer. *Cancers (Basel)* (2021) 13(4):849. doi: 10.3390/cancers13040849
21. Caputo S, Benboudjema L, Sinilnikova O, Rouleau E, Béroud C, Lidereau R. Description and Analysis of Genetic Variants in French Hereditary Breast and Ovarian Cancer Families Recorded in the UMD-BRCA1/BRCA2 Databases. *Nucleic Acids Res* (2012) 40(Database issue):D992–1002. doi: 10.1093/nar/gkr1160
22. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: Allele-Specific Copy Number and Mutation Profiles From Tumor Sequencing Data. *Ann Oncol* (2015) 26(1):64–70. doi: 10.1093/annonc/mdu479
23. Coleman RL, Fleming GF, Brady MF, Swisher EM, Steffensen KD, Friedlander M, et al. Veliparib With First-Line Chemotherapy and as Maintenance Therapy in Ovarian Cancer. *New Engl J Med* (2019) 381:2403–15. doi: 10.1056/NEJMoa1909707
24. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: Comprehensive Genome-Wide Analysis of Mutational Processes. *Genome Med* (2018) 10(1):33. doi: 10.1186/s13073-018-0539-0
25. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of Mutational Processes in Human Cancer. *Nature* (2013) 500(7463):415–21. doi: 10.1038/nature12477
26. Sopik V, Phelan C, Cybulski C, Narod SA. BRCA1 and BRCA2 Mutations and the Risk for Colorectal Cancer. *Clin Genet* (2015) 87(5):411–8. doi: 10.1111/cge.12497
27. Grinshpun A, Halpern N, Granit RZ, Hubert A, Hamburger T, Laitman Y, et al. Phenotypic Characteristics of Colorectal Cancer in BRCA1/2 Mutation Carriers. *Eur J Hum Genet* (2018) 26(3):382–6. doi: 10.1038/s41431-017-0067-1
28. Soyano AE, Baldeo C, Kasi PM. BRCA Mutation and Its Association With Colorectal Cancer. *Clin Colorectal Cancer* (2018) 17(4):e647–50. doi: 10.1016/j.clcc.2018.06.006
29. Maxwell KN, Domchek SM, Nathanson KL, Robson ME. Population Frequency of Germline BRCA1/2 Mutations. *JCO* (2016) 34(34):4183–5. doi: 10.1200/JCO.2016.67.0554

30. Evans DG, Clancy T, Hill J, Tischkowitz M. Is There Really an Increased Risk of Early Colorectal Cancer in Women With BRCA1 Pathogenic Mutations? *Clin Genet* (2016) 89(3):399–9. doi: 10.1111/cge.12687
31. Silvestri V, Leslie G, Barnes DRCIMBA Group, , Agnarsson BA, Aittomäki K, et al. Characterization of the Cancer Spectrum in Men With Germline BRCA1 and BRCA2 Pathogenic Variants: Results From the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *JAMA Oncol* (2020) 6 (8):1218–30. doi: 10.1001/jamaoncol.2020.2134
32. Du C, Peng Y, He Y, Chen G, Chen H. Low Levels of BRCA1 Protein Expression Predict a Worse Prognosis in Stage I–II Colon Cancer. *Int J Biol Markers* (2021) 36(1):47–53. doi: 10.1177/1724600820986572
33. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The Repertoire of Mutational Signatures in Human Cancer. *Nature* (2020) 578(7793):94–101. doi: 10.1038/s41586-020-1943-3
34. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of Genomic Loss of Heterozygosity Predict Homologous Recombination Repair Defects in Epithelial Ovarian Cancer. *Br J Cancer* (2012) 107(10):1776–82. doi: 10.1038/bjc.2012.451
35. Birkbak NJ, Wang ZC, Kim J-Y, Eklund AC, Li Q, Tian R, et al. Telomeric Allelic Imbalance Indicates Defective DNA Repair and Sensitivity to DNA Damaging Agents. *Cancer Discov* (2012) 2(4):366–75. doi: 10.1158/2159-8290.CD-11-0206
36. Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-Like Breast Carcinomas With BRCA1/2 Inactivation. *Cancer Res* (2012) 72 (21):5454–62. doi: 10.1158/0008-5472.CAN-12-1470
37. Rose M, Burgess JT, O'Byrne K, Richard DJ, Bolderson E. PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance. *Front Cell Dev Biol* (2020) 8:879. doi: 10.3389/fcell.2020.564601

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 23 March 2022
doi: 10.3389/fonc.2022.855463

Proof-of-Concept Pilot Study on Comprehensive Spatiotemporal Intra-Patient Heterogeneity for Colorectal Cancer With Liver Metastasis

*Ioannis D. Kyrochristos[1,2], Georgios K. Glantzounis[3], Anna Goussia[4], Alexia Eliades[5], Achilleas Achilleos[5], Kyriakos Tsangaras[5], Irene Hadjidemetriou[5], Marilena Elpidorou[5], Marios Ioannides[5], George Koumbaris[5], Michail Mitsis[3,6], Philippos C. Patsalis[5,7] and Dimitrios Roukos[1,3,8]**

[1] Centre for Biosystems and Genome Network Medicine, Ioannina University, Ioannina, Greece, [2] Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University Munich, Munich, Germany, [3] Department of Surgery, Ioannina University Hospital, Ioannina, Greece, [4] Department of Pathology, Ioannina University Hospital, Ioannina, Greece, [5] NIPD Genetics, Nicosia, Cyprus, [6] Cancer Biobank Centre, Ioannina University, Ioannina, Greece, [7] Department of Human Genetics, School of Medicine, University of Nicosia Medical School, Nicosia, Cyprus, [8] Department of Systems Biology, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Introduction: The mechanisms underlying high drug resistance and relapse rates after multi-modal treatment in patients with colorectal cancer (CRC) and liver metastasis (LM) remain poorly understood.

Objective: We evaluate the potential translational implications of intra-patient heterogeneity (IPH) comprising primary and matched metastatic intratumor heterogeneity (ITH) coupled with circulating tumor DNA (ctDNA) variability.

Methods: A total of 122 multi-regional tumor and perioperative liquid biopsies from 18 patients were analyzed *via* targeted next-generation sequencing (NGS).

Results: The proportion of patients with ITH were 53% and 56% in primary CRC and LM respectively, while 35% of patients harbored *de novo* mutations in LM indicating spatiotemporal tumor evolution and the necessity of multiregional analysis. Among the 56% of patients with alterations in liquid biopsies, *de novo* mutations in cfDNA were identified in 25% of patients, which were undetectable in both CRC and LM. All 17 patients with driver alterations harbored mutations targetable by molecularly targeted drugs, either approved or currently under evaluation.

Conclusion: Our proof-of-concept prospective study provides initial evidence on potential clinical superiority of IPH and warrants the conduction of precision oncology trials to evaluate the clinical utility of I PH-driven matched therapy.

Keywords: actionable mutations, circulating variability, comprehensive intra-patient heterogeneity, intratumor heterogeneity, next-generation sequencing, precision cancer medicine

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
David Stephen Guttery, University of Leicester, United Kingdom
Beifang Niu, Computer Network Information Center (CAS), China
Chen Wang, Mayo Clinic, United States

****Correspondence:***
Dimitrios Roukos
droukos@uoi.gr

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *15 January 2022*
Accepted: *21 February 2022*
Published: *23 March 2022*

Citation:
Kyrochristos ID, Glantzounis GK, Goussia A, Eliades A, Achilleos A, Tsangaras K, Hadjidemetriou I, Elpidorou M, Ioannides M, Koumbaris G, Mitsis M, Patsalis PC and Roukos D (2022) Proof-of-Concept Pilot Study on Comprehensive Spatiotemporal Intra-Patient Heterogeneity for Colorectal Cancer With Liver Metastasis. Front. Oncol. 12:855463. doi: 10.3389/fonc.2022.855463

INTRODUCTION

Dynamic evolution of genomic clones underlying cancer cell subpopulations and intratumor heterogeneity (ITH), as well as metastasis originating from tumor cells shed in the circulation and therapeutic resistance, represent the major causes of relapse and cancer-related death (1, 2). The capacity of next generation sequencing (NGS) studies to identify multi-regional ITH and serial circulating cell-free DNA (cfDNA) or circulating tumor DNA (ctDNA) mutations responsible for intrinsic and acquired drug resistance has transformed cancer biology and translational research (3, 4). We have developed and proposed a spatiotemporal concept of comprehensive intra-patient heterogeneity (IPH) with potential translation into Precision Oncology (5). In this pilot study we evaluate the translational efficacy of our IPH-based protocol to characterize and compare, for the first time, the ITH of primary colorectal cancer (CRC) and matched liver metastases (LM), in conjunction with the ctDNA mutational landscape in the perioperative setting. This holistic approach enables the detection of dynamic evolution of cancer genomes in time and space enabling the identification and potential targeting of all actionable mutations at different time points over the disease course.

Despite the widespread establishment of primary prevention, CRC remains the second leading cause of cancer-related death in industrialized western countries (6). In more than 50% of patients with CRC, the cancer metastasizes to the liver over the disease course, with half of the metastases being synchronous and half metachronous (7). To this day, liver resection remains the cornerstone of potentially curative treatment. However, only 15-20% of these patients are candidates for surgery aiming to complete tumor resection at diagnosis (8). Overall, treatment of colon cancer with resectable LM consists of surgery, upfront or after neoadjuvant chemotherapy, and adjuvant chemotherapy, while for rectal cancer, surgery after neoadjuvant chemoradiotherapy is the standard of care (9). A new addition to the guidelines is the option for neoadjuvant immunotherapy with nivolumab/ipilimumab or pembrolizumab in patients with high microsatellite instability (MSI), albeit based on limited data (9). For resectable metachronous metastases, treatment consists of resection and adjuvant chemotherapy with or without neoadjuvant chemotherapy (9). Notably, no molecularly targeted agent has been approved for use in resectable disease. Nevertheless, recurrence- free survival for patients with resectable CRC-LM at 5 years remains only 30%, even after multimodal treatment (8).

Over the past decade, an explosion in genome sequencing studies has provided accumulating data suggesting that a shift from single tumor biopsy to multi-regional tumor and liquid biopsy analysis could enable accurate genetic diagnosis to improve therapeutic decisions towards Precision Oncology (10, 11). Established dynamic evolution of cancer genomes in time and space before and after treatment is reflected in cancer phenotypes, subclonal ITH and circulating plasma mutational variability (3, 12, 13). Indeed, this rapid progress is being translated into multiple underway clinical trials testing the efficacy of intratumor heterogeneity analysis and serial liquid biopsies to guide more effective individualized treatment (14). Considering the discovery of thousands of drug targets *via* NGS and genome editing technologies (15, 16), IPH-matched therapy could maximize clinical benefit (5, 17). Based on intra-lesion and serial ctDNA variability, comprehensive patient-specific tumor and ctDNA analysis could empower the optimization of decision-making on the selection of targeted drugs (13).

To assess ITH and serial ctDNA mutational heterogeneity in the perioperative setting, we designed a prospective protocol encompassing multiple intra-lesional and matched plasma samples for each individual patient. We enrolled patients with CRC and LM who underwent resection of the primary and metastatic tumors with curative intent, after neo-adjuvant treatment. The comparisons between primary, metastatic and plasma mutational variability can dissect the dynamic evolution of genomic clones in time and space, orchestrating individual cancer phenotypes and drug resistance. Therefore, the concept of IPH proposed in our pilot study could enable the shift from single tumor to multiple tumor and liquid biopsy sampling, potentially improving diagnostic guidelines, and providing translational implications for personalized novel drug combinations.

MATERIALS AND METHODS

A total of 28 patients diagnosed with metastatic colorectal adenocarcinoma were treated in our surgical department in the University Hospital of Ioannina between January 2017 and December 2019 and were enrolled in this study. All patients signed a consent form for analysis of biomaterials provided by our institution's ethics committee. After initial histopathological quality control for adequacy of both primary and metastatic tumor tissue 10 patients were excluded due to inadequate tissue availability from the primary, metastatic or both sites. Thus, 18 patients were included in our final analysis, 10 women and 8 men, with an average age of 63.8 years (range 41 to 84 years). For anonymization purposes, a unique code was assigned to each patient (AA to AR). In our cohort, 13 patients had synchronous and the remaining 5 metachronous liver metastases (LM). All patients were previously subjected to neo-adjuvant chemotherapy, according to the recommendations of our institution's Multidisciplinary Tumor Board. For patients who were subjected to primary tumor resection before the initiation of the study, multi-regional primary tumor (PT) samples were retrospectively collected in our Department of Pathology. In total, 94 FFPE samples were collected from both PTs and LMs. Eighty-six out of 94 FFPE samples passed quality control (QC) requirements (DNA yield, DNA quality) to be further subjected to downstream analysis. For 16 out of 18 patients, plasma was collected at multiple time points during treatment, before and after surgery, to assess the molecular dynamics of the disease using liquid biopsies. A total of 38 plasma samples were collected out of which 36 samples passed QC requirements. Overall, out of 132 samples in total, data were collected and analyzed for 122 samples (92.4% QC success). A flowchart summarizing our study population and sample analysis is delineated in **Figure 1**.

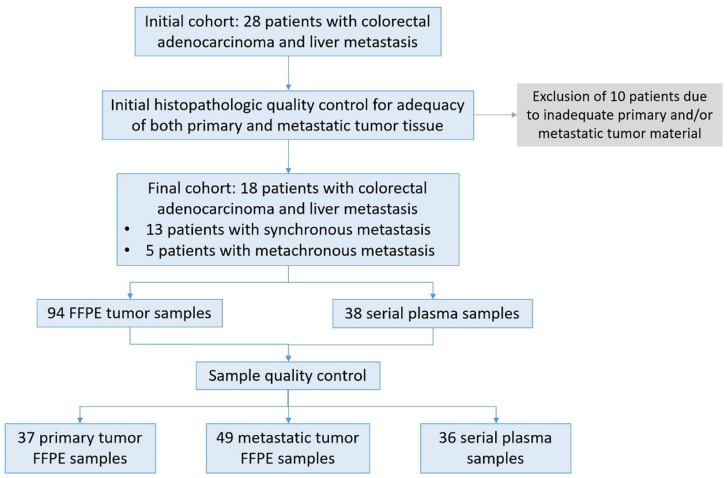


FIGURE 1 | Description of patient samples. Twenty-eight patients diagnosed with colorectal cancer and liver metastasis were enrolled in the study. After initial histopathologic quality control for adequacy of tumor tissue 18 patients were subjected to further analysis. For 16 patients both FFPE tissue as well as blood samples were collected at different time points from diagnosis, before and after surgery and during treatment and monitoring. For two patients FFPE samples were available from different sites of the primary and metastatic site. A total of 122 samples passed QC requirements (92.4%) for NGS analysis.

All samples were subjected to targeted next- generation sequencing (tNGS) using a custom 77-cancer gene panel (**Table 1**).

DNA Preparation

DNA was extracted from FFPE tissue sections using the GeneRead DNA FFPE Kit (Qiagen) following the manufacturer's instructions. DNA was quantified with a fluorometric based assay for FFPE tissue-derived DNA (Qubit flex fluorometer, Qubit dsDNA high sensitivity assay, Thermo Scientific). ctDNA was extracted from 4ml plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) following the manufacturer's instructions.

Library Preparation, Enrichment, and Sequencing

DNA libraries from FFPE and plasma samples were prepared using established protocols. For FFPE samples a commercially available kit was used for library preparation (Integrated DNA Technologies) according to manufacturer's instructions. Briefly, 60-250ng of DNA were subjected to enzymatic fragmentation, at 32oC for 7 minutes followed by adaptor ligation at 20oC for 20 minutes and clean up using magnetic beads. Next, samples were further subjected to indexing PCR and final beads-based clean up. DNA libraries from plasma were prepared using NEB reagents for dA-tailing, adaptor ligation and indexing PCR (New England Biolabs). Briefly, all the amount of extracted cfDNA (typically around 20-30 ng per 4ml plasma) is subjected to dA-tailing followed by adaptor ligation using a pool of unique adaptors for 15 minutes at 20oC. Following magnetic beads-based clean up, the samples were further subjected to indexing PCR and a final clean up step. Evaluation of library samples was performed using the 4150 Agilent Tapestation system (D1000 ScreenTape, Agilent). DNA enrichment for the genomic regions of interest was carried out using an in solution-hybridization based method using TACS (TArget Capture Sequences) specifically designed to capture selected loci in the genes of interest. TACS were then immobilized on streptavidin-coated magnetic beads for subsequent hybridization with the DNA libraries (18). A custom NIPD Genetics tumor profile gene assay was used for the identification of single nucleotide variants (SNVs), small insertions and deletions (indels), copy number alterations

TABLE 1 | Custom 77-gene panel for targeted next-generation sequencing analysis.

Single Nucleotide Variants (SNVs) / Insertions andDeletions (Indels) (67 genes)	Copy-Number Alterations (17 genes)	Translocations (10 genes)
AKT1, ALK, APC, AR, ARAF, ATM, ATRX, BARD1, BRAF, BRCA1,BRCA2, BRIP1, CDH1, CDKN2A, CHEK2, CIC, CTNNB1,DDR2, DICER1, EGFR, ERBB2, ERBB3, ERBB4, ESR1, FBXW7, FOXA1, FOXL2, FUBP1, GATA3,GNA11, GNAQ, GNAS, H3F3A, IDH1, IDH2, JAK2, KEAP1, KIT, KRAS, MAP2K1, MAP3K1, MET, MLH1, MRE11A, MSH2, MSH6, MTOR, NBN, NF1, NRAS, NTRK1, PALB2, PIK3CA, PIK3CB, PMS2, POLE, PTEN, RAD51C, RAD51D, RAF1, RET, RUNX1, SMAD4, SPOP, STK11, TERT, TP53	AR, CDKN2A, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, KIT, KRAS, MET, MYC, MYCN, PIK3CA, PTEN, RB1, TP53	ALK, BRAF, FGFR3, NF1, NTRK1, NTRK2, NTRK3, RET, ROS1, TMPRSS2

**Includes MSI assessment.*

(CNAs) and rearrangements and microsatellite instability (MSI) detection (**Table 1**). Eluted samples were amplified using outer-bound adaptor primers. Enriched DNA libraries were then normalized and subjected to sequencing on an Illumina sequencing platform.

Bioinformatics Analysis

Tissue Biopsy Analysis Pipeline

Sequencing data were de-multiplexed with bcl2fastq (v.2.16.0) and paired-end DNA sequencing reads were processed to remove adapter sequences and poor-quality reads. The remaining sequences were aligned to the human reference genome build (hg19) using the Burrows-Wheeler alignment algorithm (19). Duplicate read entries were removed (20) and aligned reads files were converted to a binary (BAM) format. For FFPE samples, the average unique read depth was approximately 950x. Variant calling was performed using a versatile somatic variant caller (21). Annotation of variants is performed using VEP (22). Variants with total base read depth less than 20X or variant count less than 4 are removed. All detected variants were filtered, annotated and classified based on well known, publicly available, disease databases [COSMIC(v92) and ClinVar(20201020)]. Benign or likely benign variants were filtered out. Only variants with strong or potential clinical significance according to AMP/ASCO/CAP guidelines (TierI/TierII) were reported for each tested sample. Gene- level Copy Number Variants (CNVs) were detected using an in-house bioinformatics pipeline that implements a circular binary segmentation method (23). Translocation calling is performed by utilizing discordant pair and split- read alignments following local assembly, realignment and an in-house filtering pipeline to refine the set of candidate events (24–27).

cfDNA Analysis Pipeline

Sequencing data were de-multiplexed with bcl2fastq (v.2.16.0) and paired-end DNA sequencing reads were processed to remove adapter sequences and poor-quality reads. The remaining sequences were aligned to the human reference genome build (hg19) using the Burrows-Wheeler alignment algorithm. Duplicate reads were identified, grouped by their families and processed to produce consensus reads per family (fgbio). The average unique read depth of plasma samples was approximately 4000X. Allelic count information for all targeted loci was used to calculate the variant allele frequency for each substitution and short insertion/deletion. A statistical error-correction model (at a base-pair resolution) was subsequently applied to refine the set of positive variant calls. The threshold for variant calling was set to 0.1%VAF. Variant annotation and classification was performed as described in the tissue biopsy analysis pipeline. Translocation calling was performed by utilizing discordant pair and split-read alignments following local assembly, realignment and a filtering pipeline to refine the set of candidate events. Transformed read depth information on pre-defined genomic windows spanning the regions of interest were normalized utilizing a, multistep statistical method (applying a within- and between-samples normalization approach). Normalized read depth data were processed to detect copy number changes. Our Bioinformatics analysis pipeline is delineated in **Figure 2**.

RESULTS

Based on our analysis, mutations were detected in 17/18 (94.4%) of patients in our cohort, and the number of variants per patient ranged from 0 to 18. Overall, mutations in 28 genes were identified, adding up to an average 4.7 variants per patient. In agreement with published literature, the most frequently mutated genes were APC and TP53, followed by PIK3CA and KRAS, while 3 or fewer variants were identified for all other genes. Importantly, the majority of identified variants (19/28, 69%) were rare and detected in only one patient. Moreover,

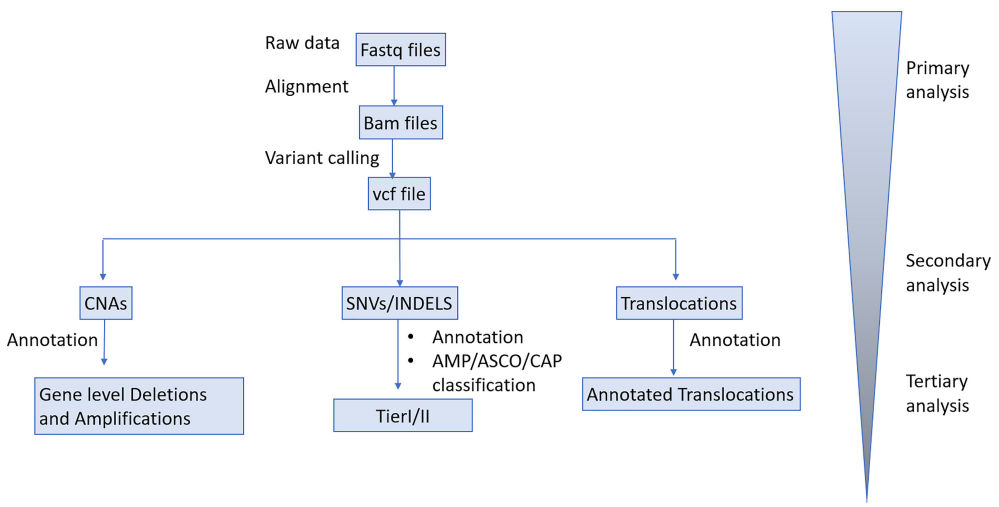


FIGURE 2 | Bioinformatics analysis pipeline.

a significant proportion of variants was identified in potentially clinically relevant genes not routinely tested in day-to-day practice. The majority (60%) of alterations regarded missense mutations, while other alteration types included frameshift (14.1%) and nonsense (11.8%) mutations, amplifications (8.2%) and others. **Figure 3** summarizes our findings on the frequency of mutated genes, mutation types and patient variants.

Intra-Tumor Heterogeneity

Initially, we aimed to detect and characterize the extent of ITH within matched primary and metastatic lesions. Concerning primary tumor analysis, we analyzed 2-4 multiregional PT samples from 15 patients, while regarding liver metastases, we analyzed 2-5 multiregional LM samples in each of 16 patients. For one patient (AG), we analyzed two distinct liver metastases, one synchronous and one metachronous. Due to quality control exclusion, three patients had single PT and/or LM samples. There was no PT sample available for one patient. In our cohort, increased number of multiregional samples did not correlate with the frequency of observed ITH as compared to dual samples in primary or metastatic tumors, although our relatively small cohort sample warrants further large-scale investigation. Quite notably, 53% (8/15) of patients with multi-regional primary biopsies featured ITH between spatially distinct geographical regions of the PT at variable degrees, with homogeneity of observed variants between regions ranging from 83% to under 15%. Similarly, variable multi-regional ITH of the LMs was observed in 56% (9/16) of patients between different regions of the same metastasis. These results indicate that analysis of multiple spatially distinct samples is meaningful in more than half of lmCRC patients, with potentially significant clinical implications as discussed below.

Circulating Tumor DNA Variability

Based on the highly compelling potential for a non-invasive blood-based patient monitoring strategy, we additionally analyzed 36 serially collected plasma samples before and after therapy from 16 of 18 patients utilizing NIPD Genetics custom gene assay, to explore the temporal dynamics of ctDNA variability. Plasma ctDNA analysis identified mutations in 60% (9/15) of patients with both pre- and post-operative plasma samples. Tumor mutations were detected in the ctDNA analysis in 8 and 2 of 16 patients pre- and post- operatively respectively. There was a clear tendency for pre-operative tumor mutations in plasma to be undetectable in ctDNA after complete tumor resection (R0 surgery) with curative intention, indicating the well-established association of plasma mutation detection and tumor burden. Persistence of ctDNA mutations post-operatively, as for example in patient AA for whom a staged hepatectomy was planned but not performed, and thus was not subjected to R0 resection, correlated with early relapse (<1 year) and adverse oncological outcome.

Comprehensive Intra-Patient Heterogeneity

The first most ambitious aim of our project was to dissect the comprehensive spatiotemporal genetic IPH for each individual patient, comprising the ITH of primary and metastatic lesions, as well as temporal ctDNA mutational heterogeneity. Most importantly, genetic differences between matched PT and LMs, meaning the presence of mutations in either the primary or metastatic tissue but not in both, were detected in 53% (9/17) of cases with both PT and matched metastatic tumor (MT) samples, with 35% (6/17) of patients harboring *de novo* mutations present only in LMs, indicating dynamic clonal evolution (**Figure 4**). More specifically, 5 of these patients

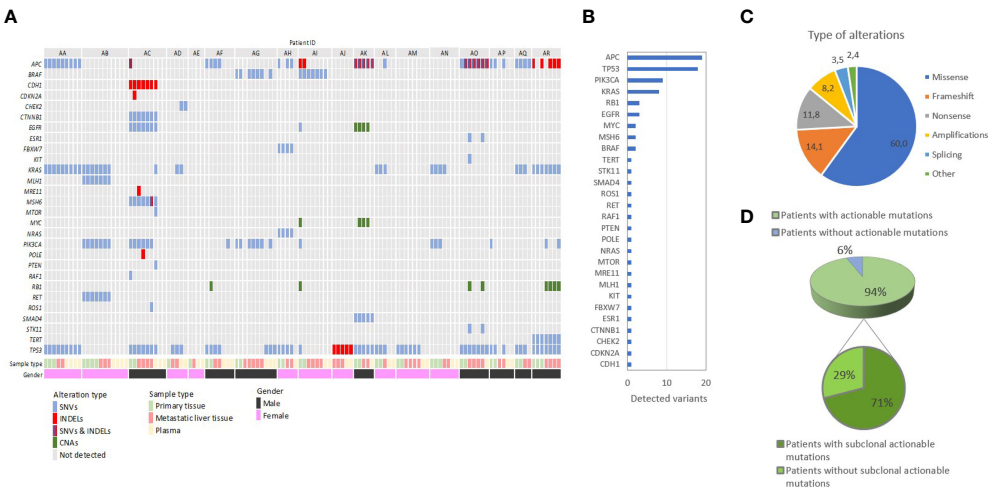


FIGURE 3 | Summary of the distribution of genetic alterations among all tissue and liquid biopsy samples from our cohort of 18 patients with metastatic colorectal adenocarcinoma. **(A)** Oncoplot shows the patients in a horizontal orientation and the gene and corresponding driver mutations in the vertical orientation. **(B)** Frequency of mutations observed per gene: The frequency of mutations is estimated based on the cumulative list of variants per patient. The number of variants identified in all patients is shown. **(C)** Distribution of different types of genetic alterations identified in the study. **(D)** Overall detection of potentially actionable mutations and their clonal status in our study: 71% of all patients with detectable actionable mutations harbored subclonal putative oncotargets, defined as being present in some but not all tumor samples.

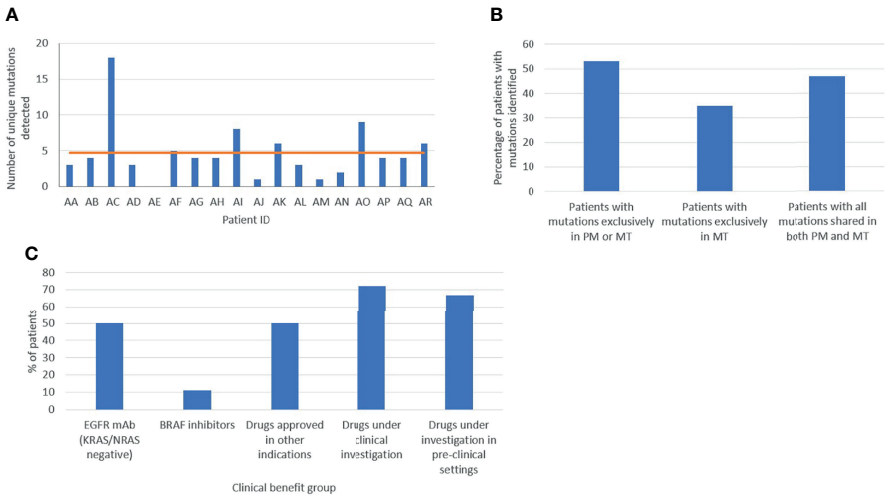


FIGURE 4 | Assessment of tumor heterogeneity **(A)** Total number of variants per patient in the cohort. **(B)** Molecular heterogeneity between primary (PM) and metastatic (MT) lesions. **(C)** Clinical actionability of genetic findings. Based on the variants identified, patients were allocated in different clinical benefit groups: EGFR monoclonal antibody (mAb), BRAF inhibitor, drugs approved in other indications, and drugs under clinical and pre-clinical investigation.

harbored several mutated genes only in the metastatic samples, including genes such as KRAS, TP53, APC, PIK3CA, RB1. This finding supports the need for multi-regional NGS analysis of metastases to potentially improve therapeutic decisions. Overall, 41% of variants observed in solid tumor samples were ubiquitously shared by all regions between primary and liver lesions. These results suggest that multiple biopsies of both primary and matched metastatic lesions are required to delineate cancer genetic diversity with potentially crucial implications for the clinic. Quite impressively, ctDNA analysis effectively dissected the complete variability of PTs and matched MTs in 25% (4/16) of cases, identifying all intra-tumorally identified mutations in these patients, suggesting the potential use of liquid biopsy for delineating ITH. Lastly, in 4 of 16 patients (patients AA, AD, AG, AJ), ctDNA analysis uncovered mutations, which were not detected in the primary or metastatic tumors and could represent the result of dynamic subclonal evolution, which needs further investigation. Notably, putatively actionable mutations identified exclusively in liquid biopsies, including KRAS, CHEK2, PIK3CA, TP53 and others, could provide important translational therapeutic implications. All of the above support our hypothesis that a holistic approach to the oncological patient requires rigorous combinatorial analysis of spatiotemporally collected primary tumoral, metastatic and plasma samples to improve the accuracy in characterizing the complete tumor genetic diversity.

Intra-Patient Mutational Heterogeneity-Driven Targeted Therapy

The second major aim of our study was to identify potential oncotargets and characterize their clonality within PT and matched LMs, in order to clarify whether a single biopsy could effectively guide therapeutic decision-making. In our cohort, all patients presented one or multiple therapeutic opportunities specifically for targeted treatment. Of the 18 patients, 9 (50%) were wild-type for KRAS/NRAS and could derive potential benefit from EGFR inhibitors, such as cetuximab and panitumumab. Two patients had BRAF mutations, presenting an opportunity for treatment with BRAF-inhibitors, already approved for use in non- resectable CRC (9). Moreover, 50% (9/18) of participants had variant targets of drugs already approved for other cancer types suggesting a potential benefit through drug repurposing (9). Additionally, 72% (13/18) and 67% (12/18) of patients could be matched to targeted drugs under clinical or pre-clinical evaluation respectively, in CRC or other cancer types (**Figure 4**). It is worth noting that, out of 19 detected rare variants, 10 were potentially actionable. Collectively, 94% (17/18) of participants in our cohort featured targets of not-yet-approved agents in the clinical or pre-clinical stages of development (**Figure 4**). Selected examples of known druggable targets are shown in 3D protein structure representation in **Supplementary Figure 1**.

A notable example is patient AC, a patient with established Lynch syndrome according to the Amsterdam criteria, as well as microsatellite instability previously identified using a panel of five markers (βατ-25, BAT-26, D5S346, D17S250 & D2S123) through PCR-based testing and confirmed by our assay. Our analysis also confirmed a mutated MSH6 gene in all intratumor samples as well as its presence in plasma. In this patient, more than 10 putatively actionable variants were identified, with the majority being subclonal, suggesting the potential for extensive drug combinations, including immunotherapy, guided by intratumor and circulating DNA heterogeneity.

A crucial parameter hindering the potential therapeutic utilization of genetic findings is the putative spatial and temporal subclonality of actionable mutations, which, if

validated, could not be addressed *via* a single tumor biopsy. Indeed, our study identified an average of 3.2 potentially actionable variants per patient, suggesting the capacity for combinatorial targeted therapy, following large-scale validation. However, based on comparative primary and metastatic tumor analysis, 71% (12/17) of patients with detectable mutations harbored subclonal putative oncotargets not identified in all tumor samples, which could potentially be masked by single-biopsy analysis (**Figure 3**) (28, 29). Analysis of ctDNA uncovered potentially druggable variants in all nine patients with detectable plasma mutations, indicating the potential role of ctDNA in guiding therapeutic decisions following validation. Quite notably, 33% and 22% of participants harbored putatively actionable variants only in LM and plasma samples respectively, which would not have been identified by PT analysis alone, although our study is limited in discovering plasma-exclusive variants, as described below. Caution should be taken when interpreting low frequency variants identified exclusively in plasma in genes previously shown to contribute to clonal hematopoiesis of intermediate potential (CHIP). In the absence of lymphocyte sample for testing we could not exclude the possibility of low frequency CHIP-derived mutations in genes such as KRAS, TP53 and PIK3CA. Overall, almost 65% of all oncotargets detected in our cohort featured diverse degrees of intratumor and circulating variability among samples from the same patient. Therefore, our study strongly supports the hypothesis that therapeutic decision-making should not be limited to single PT samples and warrants further evaluation of spatiotemporally collected samples from matched PTs, MTs and plasma in large-scale studies to establish the clinical utility of comprehensive IPH. A summary of our findings on genetic spatiotemporal heterogeneity in tumor and liquid samples is delineated in **Figure 3**. Moreover, an overview, as well as definitions, on our most crucial findings can be found in **Box 1**.

DISCUSSION

Our proof-of-concept pilot prospective study on comprehensive intra-patient genetic heterogeneity in patients with resected CRC and LM provides new diagnostic, predictive and therapeutic implications. In the present cohort, 53% and 56% of patients harbored ITH of the primary and metastatic tumors respectively, while 53% had genetic differences between matched primary and metastatic samples. Combining both primary and metastatic ITH with matched plasma cfDNA, 25% of patients had aberrations in plasma but not in tumor specimens which highlights the potential benefit of ctDNA analysis in capturing dynamic tumor heterogeneity which might be missed from analysis of only a few FFPE sections that offer solely a snapshot of the tumor's molecular profile. Almost all patients had cancer targets for approved drugs and/or agents under investigation in ongoing clinical trials (14) or pre-clinical studies. Our data are consistent with multiple genomic and transcriptomic studies on spatial and temporal dynamic evolution of cancer genomes (29, 30), underlying intratumor subclonal cell populations.

Multiple genomic studies have identified ITH as an integral part of cancer evolution in solid tumors (28, 31). This multi-regional variability has been strongly correlated with intrinsic and acquired drug resistance and relapse (29, 32, 33). Indeed, consistent with our work, several studies have uncovered extensive ITH of primary CRC as a prognostic factor for metastasis, as well as a predictor of drug resistance (34, 35). With regards to ITH of the liver metastasis, our study is among the very few published reports with a strict protocol, enabling the exact identification of metastatic genetic ITH in 56% of patients. Most available studies have evaluated the variability between matched primary and metastatic lesions through either single or multi-regional samples. Based on a single biopsy, data remain controversial regarding the degree of heterogeneity between

BOX 1 | Summary of the most crucial findings in our prospective cohort.

ITH of the primary tumor	The proportion of patients harbouring variable levels of ITH of the primary	8/15 patients with multi-regional samples from the primary tumor, 53%
ITH of the liver metastasis	The proportion of patients harbouring variable levels of ITH of the liver metastatic lesion	9/16 patients with multi-regional samples from liver metastases, 56%
Genetic heterogeneity between primary and matched metastatic	The proportion of patients harbouring mutations detected either in the primary or the matched metastatic tumor but not in both	9/17 patients with matched primary and metastatic tumor samples, 53%
De novo mutations in liver metastases	The proportion of patients harboring de novo mutations in liver metastases not found in the primary tumor, including potentially actionable variants indicating dynamic clonal evolution	6/17 patients with matched primary and metastatic tumor samples, 35%
Detection of cfDNA mutations	Proportion of patients in which tumor mutations were detected in cfDNA pre- operatively Proportion of patients in which tumor mutations were detected in cfDNA post- operatively	8/16 patients, 50% 2/16 patients, 12.5%
Potentially actionable mutations	9/18 (50%) and 17/18 (94%) patients could benefit from repurposing of already approved drugs and agents under clinical or pre-clinical development respectively Overall, an average of 3.2 actionable mutations per patient were identified.	12/17 (71%) patients with detectable mutations harbored potentially actionable alterations not ubiquitously shared by all tumor samples, indicating the need for spatiotemporal sampling to increase therapeutic accuracy

primary and matched metastatic lesions (36, 37). Similarly, intra-lesion sampling of both primary CRC and LM has provided contradictory findings. For instance, Siraj et al. demonstrated high genomic concordance between primary CRC and matched LM *via* whole-exome sequencing of a total of 191 samples (38). By contrast, Hu and colleagues, in a whole-exome sequencing cohort consisting of 118 biopsies from 23 patients, uncovered extensive inter- and intra-lesion heterogeneity (39), in accordance with our prospective cohort. Whether a shift from the current standard of single-tumor biopsies to multi-sampling from matched primary and metastatic lesions can enable more accurate diagnosis and decision-making on novel combinations of molecularly targeted drugs remains at the present unclear. By identifying putatively targetable mutations in nearly all patients, of which approximately 65% featured variable intratumor and ctDNA heterogeneity, our pilot study strongly supports the prospective evaluation of this concept within Precision Oncology trials. Indeed, the proportion of patients with clinically actionable mutations in recent large-scale consortia based on single- biopsy NGS, such as The Cancer Genome Atlas and the Pan-Cancer Analysis of Whole Genomes, ranged between 57% and over 75% (15, 40), highlighting that multi-sampling could substantially increase the discovery rates of cancer targets.

Targeted NGS, whole-exome sequencing and whole-genome sequencing of cfDNA or ctDNA, as well as analysis of circulating tumor cells, are receiving tremendous attention towards implementation into the clinical setting for early diagnosis, individualized prediction of drug response, patient monitoring to readily detect relapse and drug development (30, 41). Indeed, beyond large innovative projects (42), several comprehensive pan-cancer multi-gene panels, including for CRC, have already been developed and approved by federal regulatory institutions as companion diagnostics for tumor profiling within modern guidelines, highlighting the introduction of Precision Oncology into clinical practice (30). On this basis, the promising findings provided *via* our custom 77-gene panel could be incorporated into the clinical setting, following large- scale validation. Additionally, completed early-phase clinical trials, such as the TARGET (43) and the I-PREDICT (44) studies have demonstrated potential clinical benefit from liquid biopsy-guided drug target detection and molecularly matched therapy.

Based on our recent published work on the comprehensive model of IPH (5), in the present study, we have explored the potential translational implications of IPH, comprised by the ITH of primary and metastatic lesions in combination with plasma DNA mutational landscapes. This integrated framework has highlighted the necessity for multiple sample analysis from both tumor and ctDNA in the perioperative setting. This approach could potentially transform decision-making on neoadjuvant or adjuvant treatment following complete tumor resection to overcome the unmet clinical challenge of substantial therapeutic resistance and relapse rates among patients with resectable colorectal cancer with liver metastasis. In fact, 35% of patients in our cohort harbored *de novo* mutations in the metastases, potentially unraveling the capacity to improve therapeutic decisions. Moreover, *de novo* mutations were quite impressively identified in the ctDNA from 25% of patients but not in the multi-regional analysis of primary and metastatic tumors. However, this finding requires further evaluation *via* DNA analysis of white blood cells to differentiate cancer from germline mutations or potential clonal hematopoiesis (45). Additionally, on the basis of WGS on over 2,500 cancer samples, the PCAWG initiative identified actionable mutations in 60% of tumors (31), while in our integrated analysis of both multi-regional tumor and serial liquid biopsies, all patients with detectable mutations harbored actionable events, supporting the translational framework of detailed tumor and plasma analysis. Nevertheless, as demonstrated by our results, isolated analysis of primary CRC, metastatic tissue or ctDNA is unable to uncover the complete mutational landscape for each individual patient, highlighting the translational importance or our work.

Despite intriguing findings, our study is presented with several limitations. First, the number of enrolled patients and total samples analyzed is relatively small to extract definitive conclusions. Second, our plasma analysis is lacking in serial postoperative samples over the course of disease to evaluate the potential for early relapse detection and potential therapeutic targeting. And third, our analysis focuses on the detection of actionable mutations and matched targeted drugs, while not exploring putative immunotherapeutic implications. It should however be noted that our analysis was a pilot study, which was designed to explore the feasibility and potential clinical applications of the IPH concept, therefore, encouraging further extensive work and large-scale prospective studies and precision clinical trials.

Currently, ongoing projects and underway clinical trials are evaluating the clinical utility of cancer type- specific ITH and plasma DNA mutational heterogeneity, as well as tumor-infiltrating immune and stromal cells of the tumor microenvironment, to establish Precision Oncology and Immunology in the clinical setting (10, 14, 46, 47). Expanding our holistic IPH approach to include intratumor interactions between cancer and environmental immune cells, pioneering single-cell genome sequencing, editing and machine learning technologies, as well as liquid biopsies of peripheral blood for cfDNA, circulating tumor and immune cell analysis raise novel expectations to realize patient-specific optimal precision immuno-oncological treatment (17, 30, 48–51).

CONCLUSIONS

Our IPH-based concept with a pilot-level tNGS analysis of 122 multi-regional tumor and liquid biopsies with a custom 77-gene assay has provided initial evidence on the necessity of multiple spatiotemporal sampling in patients with resectable CRC with LM. This region-to-region NGS analysis of primary and matched metastatic lesions, as well as perioperative plasma samples, has enabled the dissection of primary and metastatic ITH and ctDNA variability. Indeed, there was substantial heterogeneity between primary, metastatic and circulating mutational landscapes, unraveling the dynamic evolution of cancer genomes, underlying tumor progression, metastasis, drug resistance and

relapse. Our data uncover the significance of ITH and cfDNA as predictors of response to molecularly targeted drugs. Moreover, the identification of clinically actionable mutations by ctDNA analysis highlights the importance of evaluating tumor dynamics and heterogeneity using liquid biopsy to guide therapy selection and better stratification of patients for clinical trial enrollment and treatment-response evaluation. The holistic approach in our work detected in average 3.2 actionable events per patient, warranting the conduction of precision clinical trials to assess the clinical utility of novel targeted drug combinations in the adjuvant or neo-adjuvant setting. In summary, comprehensive IPH-based translational research and clinical trials are expected to transform current treatment guidelines towards the implementation of Cancer Precision Medicine in the clinical setting, to overcome the unmet challenges of high drug resistance and relapse rates.

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

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

IK: conceptualization, formal analysis, investigation, methodology, project administration, validation, writing - original draft, and writing - review and editing. GG: methodology and resources. AG: data curation, investigation, and resources. AE: data curation, formal analysis, investigation, visualization, and writing- original draft. AA: software and validation. KT: funding acquisition and methodology. IH: investigation, methodology, and visualization. ME: data curation, methodology, and resources. MI: funding acquisition, resources, and software. GK: validation and visualization. MM: project administration, resources, and supervision. PP: funding acquisition, methodology, resources, software, and validation. DR: conceptualization, methodology, project administration, supervision, validation, writing- original draft, and writing - review and editing. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.855463/full#supplementary-material

REFERENCES

1. Klein CA. Selection and Adaptation During Metastatic Cancer Progression. *Nature* (2013) 501:365–72. doi: 10.1038/nature12628
2. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of Genomic Clones in Breast Cancer Patient Xenografts at Single-Cell Resolution. *Nature* (2015) 518:422–6. doi: 10.1038/nature13952
3. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-Invasive Analysis of Acquired Resistance to Cancer Therapy by Sequencing of Plasma DNA. *Nature* (2013) 497:108–12. doi: 10.1038/nature12065
4. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic Architecture and Evolution of Clear Cell Renal Cell Carcinomas Defined by Multiregion Sequencing. *Nat Genet* (2014) 46:225–33. doi: 10.1038/ng.2891
5. Roukos DH. Spatiotemporal Diversification of Intrapatient Genomic Clones and Early Drug Development Concepts Realize the Roadmap of Precision Cancer Medicine. *Drug Discovery Today* (2017) 22:1148–64. doi: 10.1016/j.drudis.2017.03.014
6. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal Cancer Statistic. *CA Cancer J Clin* (2020) 70:145–64. doi: 10.3322/caac.21601
7. Zarour LR, Anand S, Billingsley KG, Bisson WH, Cercek A, Clarke MF, et al. Colorectal Cancer Liver Metastasis: Evolving Paradigms and Future Directions. *Cell Mol Gastroenterol Hepatol* (2017) 3:163–73. doi: 10.1016/j.jcmgh.2017.01.006
8. Nordlinger B, Sorbye H, Glimelius B, Poston GJ, Schlag PM, Rougier P, et al. Perioperative FOLFOX4 Chemotherapy and Surgery Versus Surgery Alone for Resectable Liver Metastases From Colorectal Cancer (EORTC 40983): Long-Term Results of a Randomised, Controlled, Phase 3 Trial. *Lancet Oncol* (2013) 14:1208–15. doi: 10.1016/S1470-2045(13)70447-9
9. National Comprehensive Cancer Network, Inc. (2021). Available at: www.nccn.org.
10. Dugger SA, Platt A, Goldstein DB. Drug Development in the Era of Precision Medicine. *Nat Rev Drug Discov* (2018) 17:183–96. doi: 10.1038/nrd.2017.226
11. Chen HZ, Bonneville R, Roychowdhury S. Implementing Precision Cancer Medicine in the Genomic Era. *Semin Cancer Biol* (2019) 55:16–27. doi: 10.1016/j.semcancer.2018.05.009
12. Costanzo M, Kuzmin E, Van Leeuwen J, Mair B, Moffat J, Boone C, et al. Global Genetic Networks and the Genotype-To-Phenotype Relationship. *Cell* (2019) 177:85–100. doi: 10.1016/j.cell.2019.01.033
13. Chakravarty D, Solit DB. Clinical Cancer Genomic Profiling. *Nat Rev Genet* (2021) 22:483–501. doi: 10.1038/s41576-021-00338-8
14. *ClinicalTrials.Gov* (2021). Available at: https://clinicaltrials.gov/.
15. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* (2018) 173:371–385 e318. doi: 10.1016/j.cell.2018.02.060
16. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-Cancer Analysis of Whole Genomes. *Nature* (2020) 578:82–93. doi: 10.1038/s41586-020-1969-6
17. Kyrochristos ID, Roukos DH. Comprehensive Intra-Individual Genomic and Transcriptional Heterogeneity: Evidence-Based Colorectal Cancer Precision Medicine. *Cancer Treat Rev* (2019) 80:101894. doi: 10.1016/j.ctrv.2019.101894
18. Koumbaris G, Kypri E, Tsangaras K, Achilleos A, Mina P, Neofytou M, et al. Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal Sex. *Clin Chem* (2016) 62:848–85. doi: 10.1373/clinchem.2015.252502
19. Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA -MEM*. arXiv: Genomics (2013).
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352

21. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, Mcewen R, et al. VarDict: A Novel and Versatile Variant Caller for Next-Generation Sequencing in Cancer Research. *Nucleic Acids Res* (2016) 44:e108. doi: 10.1093/nar/gkw227
22. Mclaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* (2016) 17:122. doi: 10.1186/s13059-016-0974-4
23. Seshan VE, Adam O. *DNAcopy: DNA Copy Number Data Analysis. R Package Version 1.66.0.* (2021).
24. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: A Probabilistic Framework for Structural Variant Discovery. *Genome Biol* (2014) 15:R84. doi: 10.1186/gb-2014-15-6-r84
25. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: Rapid Detection of Structural Variants and Indels for Germline and Cancer Sequencing Applications. *Bioinformatics* (2016) 32:1220–2. doi: 10.1093/bioinformatics/btv710
26. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: Sensitive and Specific Genomic Rearrangement Detection Using Positional De Bruijn Graph Assembly. *Genome Res* (2017) 27:2050–60. doi: 10.1101/gr.222109.117
27. Gawroński AR, Lin YY, Mcconeghy B, Lebihan S, Asghari H, Koçkan C, et al. Structural Variation and Fusion Detection Using Targeted Sequencing Data from Circulating Cell Free DNA. *Nucleic Acids Res* (2019) 47:e38.
28. Iacobuzio-Donahue CA, Litchfield K, Swanton C. Intratumor Heterogeneity Reflects Clinical Disease Course. *Nat Cancer* (2020) 1:3–6. doi: 10.1038/s43018-019-0002-1
29. Black JRM, Mcgranahan N. Genetic and Non-Genetic Clonal Diversity in Cancer Evolution. *Nat Rev Cancer* (2021) 21:379–92. doi: 10.1038/s41568-021-00336-2
30. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid Biopsy Enters the Clinic - Implementation Issues and Future Challenges. *Nat Rev Clin Oncol* (2021) 18:297–312. doi: 10.1038/s41571-020-00457-x
31. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Characterizing Genetic Intra-Tumor Heterogeneity Across 2,658 Human Cancer Genomes. *Cell* (2021) 184:2239–2254 e2239. doi: 10.1016/j.cell.2021.03.009
32. Ziogas DE, Kyrochristos ID, Roukos DH. Discovering Novel Valid Biomarkers and Drugs in Patient-Centric Genomic Trials: The New Epoch of Precision Surgical Oncology. *Drug Discovery Today* (2018) 23:1848–72. doi: 10.1016/j.drudis.2018.07.008
33. Kyrochristos ID, Ziogas DE, Roukos DH. Drug Resistance: Origins, Evolution and Characterization of Genomic Clones and the Tumor Ecosystem to Optimize Precise Individualized Therapy. *Drug Discov Today* (2019) 24:1281–94. doi: 10.1016/j.drudis.2019.04.008
34. Kim TM, Jung SH, An CH, Lee SH, Baek IP, Kim MS, et al. Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic Heterogeneity. *Clin Cancer Res* (2015) 21:4461–72. doi: 10.1158/1078-0432.CCR-14-2413
35. Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, et al. Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. *PloS Genet* (2016) 12:e1005778. doi: 10.1371/journal.pgen.1005778
36. Brannon AR, Vakiani E, Sylvester BE, Scott SN, Mcdermott G, Shah RH, et al. Comparative Sequencing Analysis Reveals High Genomic Concordance Between Matched Primary and Metastatic Colorectal Cancer Lesions. *Genome Biol* (2014) 15:454. doi: 10.1186/s13059-014-0454-7
37. Joung JG, Oh BY, Hong HK, Al-Khalidi H, Al-Alem F, Lee HO, et al. Tumor Heterogeneity Predicts Metastatic Potential in Colorectal Cancer. *Clin Cancer Res* (2017) 23:7209–16. doi: 10.1158/1078-0432.CCR-17-0306
38. Siraj S, Masoodi T, Siraj AK, Azam S, Qadri Z, Ahmed SO, et al. Clonal Evolution and Timing of Metastatic Colorectal Cancer. *Cancers (Basel)* (2020) 12:2938. doi: 10.3390/cancers12102938
39. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative Evidence for Early Metastatic Seeding in Colorectal Cancer. *Nat Genet* (2019) 51:1113–22. doi: 10.1038/s41588-019-0423-x
40. Sabarinathan R, Pich O, Martincorena I, Rubio-Perez C, Juul M, Wala J, et al. The Whole-Genome Panorama of Cancer Drivers. *bioRxiv* (2017) 190330. doi: 10.1101/190330
41. Pal A, Shinde R, Miralles MS, Workman P, De Bono J. Applications of Liquid Biopsy in the Pharmacological Audit Trail for Anticancer Drug Development. *Nat Rev Clin Oncol* (2021) 18:454–67. doi: 10.1038/s41571-021-00489-x
42. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-Wide Cell-Free DNA Fragmentation In Patients With Cancer. *Nature* (2019) 570:385–9. doi: 10.1038/s41586-019-1272-6
43. Rothwell DG, Ayub M, Cook N, Thistlethwaite F, Carter L, Dean E, et al. Utility of ctDNA to Support Patient Selection for Early Phase Clinical Trials: The TARGET study. *Nat Med* (2019) 25:738–43. doi: 10.1038/s41591-019-0380-z
44. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, Williams CB, et al. Molecular Profiling of Cancer Patients Enables Personalized Combination Therapy: The I-PREDICT Study. *Nat Med* (2019) 25:744–50. doi: 10.1038/s41591-019-0407-5
45. Razavi P, Li BT, Brown DN, Jung B, Hubbell E, Shen R, et al. High-Intensity Sequencing Reveals the Sources of Plasma Circulating Cell-Free DNA Variants. *Nat Med* (2019) 25:1928–37. doi: 10.1038/s41591-019-0652-7
46. Swanton C. Take Lessons From Cancer Evolution to The Clinic. *Nature* (2020) 581:382–3. doi: 10.1038/d41586-020-01347-z
47. Waldman AD, Fritz JM, Lenardo MJ. A Guide to Cancer Immunotherapy: From T Cell Basic Science to Clinical Practice. *Nat Rev Immunol* (2020) 20:651–68. doi: 10.1038/s41577-020-0306-5
48. Havel JJ, Chowell D, Chan TA.). The Evolving Landscape Of Biomarkers for Checkpoint Inhibitor Immunotherapy. *Nat Rev Cancer* (2019) 19:133–50. doi: 10.1038/s41568-019-0116-x
49. Stuart T, Satija R. Integrative Single-Cell Analysis. *Nat Rev Genet* (2019) 20:257–72. doi: 10.1038/s41576-019-0093-7
50. Schumann K, Raju SS, Lauber M, Kolb S, Shifrut E, Cortez JT, et al. Functional CRISPR Dissection of Gene Networks Controlling Human Regulatory T Cell Identity. *Nat Immunol* (2020) 21:1456–66. doi: 10.1038/s41590-020-0784-4
51. Wu TD, Madireddi S, De Almeida PE, Banchereau R, Chen YJ, Chitre AS, et al. Peripheral T Cell Expansion Predicts Tumour Infiltration and Clinical Response. *Nature* (2020) 579:274–8. doi: 10.1038/s41586-020-2056-8

Conflict of Interest: AE, AA, KT, IH, ME, MI, GK, and PP are all employed by NIPD Genetics.

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

Publisher's Note: 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 Oncology

ORIGINAL RESEARCH
published: 20 April 2022
doi: 10.3389/fonc.2022.862887

New Studies of the Aberrant Alterations in Fibrillin-1 Methylation During Colorectal Cancer Development

*Ling Lv[1], Jianzhong Ma[2], Lina Wu[3], Chao Zhang[2], Yueping Wang[4] and Guang Wang[5]**

[1] Department of Thoracic Surgery, The First Affiliated Hospital of China Medical University, Shenyang, China, [2] School of Business, Xianda College of Economics & Humanities, Shanghai International Studies University, Shanghai, China, [3] Department of Medical Laboratory, Shengjing Hospital of China Medical University, Shenyang, China, [4] Department of Infection Diseases, The First Affiliated Hospital of Hainan Medical University, Haikou, China, [5] Hepatobiliary Surgery Department, The First Affiliated Hospital, China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Walter Hernán Pavicic, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Renjie Wang, Fudan University, China
Hamid Asadzadeh, Shahid Beheshti University of Medical Sciences, Iran

***Correspondence:**
Guang Wang
wangguang_cmu@163.com

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

Received: *26 January 2022*
Accepted: *04 March 2022*
Published: *20 April 2022*

Citation:
Lv L, Ma J, Wu L, Zhang C, Wang Y and Wang G (2022) New Studies of the Aberrant Alterations in Fibrillin-1 Methylation During Colorectal Cancer Development. Front. Oncol. 12:862887. doi: 10.3389/fonc.2022.862887

Background: Fibrillin-1 (FBN1) methylation risk from control to colorectal cancer (CRC), the variation regularities of FBN1 methylation, and DNA methyltransferase (DNMT) catalyzed with FBN1 methylation had not been reported yet; these were all studied in this paper.

Methods: FBN1 methylation roles were investigated with big data and meta-analysis.

Results: The 6 independent studies were searched including 702 tissue and 448 feces. FBN1 methylation frequencies of CRC, adenoma or polyp, and control in tissue were 79.1%, 69.4%, and 2.7%, respectively; those in feces were 74.6%, 50.7%, and 10.8%, respectively. FBN1 methylation of control samples was used as a standard reference; this study showed that ORs (95% CI) of FBN1 methylation in CRC and control tissues were 124.79 (62.86–248.35); those in feces were detected to be 30.87 (16.48–57.85). FBN1 methylation risk in tissue was higher than that in feces; there was a quadratic equation between the methylation rate of tissue and that of feces. There was another quadratic curve in the variation process of FBN1 methylation; this curve reflected the overall metabolism regularity of DNMT.

Conclusions: The transcriptional inactivation of FBN1 gene might start from normal colonic epithelium; the quadratic curve of FBN1 methylation catalyzed by DNMT can gradually produce powerful strength, accelerate expansion, and eventually lead to CRC. The overall metabolism regularity of DNMT maintains the changing process of FBN1 methylation; it has the changing feature of the same quadratic curve. FBN1 methylation is a promising biomarker. FBN1 methylation risk size in feces reflects that in tissue in non-invasive detection.

Keywords: methylation variation, prevention and diagnosis, epigenetics, biomarker, DNA methyltransferase

INTRODUCTION

Epigenetic alteration is common in cancer occurrence and progression; DNA methylation is an important component of epigenetics. The epigenetic pathway of CpG island methylator phenotype (CIMP) had been used clinically in the diagnosis and screening of colorectal cancer (CRC), but the molecular pathological mechanism of CIMP is still not very clear. DNA methylation affected the expression of genes by interacting with the transcription factors or by changing the chromatin structure. Epigenetic regulators of gene expression were mainly the methylation of CpG islands, histone post-translational modifications (PTMs), and microRNAs (miRNAs) (1). Another reason was that aberrant DNA methylation at the 5-position of cytosine was catalyzed and maintained by DNA methyltransferase (DNMT), and it was associated with not only various cancers by silencing of tumor suppressor genes but also other diseases (2). The higher expression of DNMT was demonstrated in a variety of human malignancies, and tumor progression was facilitated by DNMT-mediated gene inactivation (3). Therefore, it is very important for the aberrant regulation of DNA methylation to be explored regarding gene expression, mediation, DNMT, and tumor development process. We all knew that the process of methylation changing was closely linked to the process of DNMT. During cancer occurrence and development, how did the overall changing process of DNA methylation affect gene expression? How did the overall changing process of DNMT catalyze and maintain DAN methylation and synthesize or degrade physiological activity? These problems were discussed with human methylation experiment data in this paper.

CRC is a common malignancy in the digestive system; its incidence and mortality rose worldwide (4), ranking third in malignant tumors (5). In western developed countries, the CRC mortality rate was 33%, and it is one of the most common causes of malignant tumors. The lifetime risk of CRC is as high as 5% in the American population (6). CRC had the characteristics of hidden onset, long course, and good early diagnosis and prognosis, which makes it suitable for screening. Numerous studies had shown that early screening diagnosis had reduced the incidence and mortality of CRC. The main aim of CRC screening was to find early tumors that were treatable or precancerous lesions that were highly likely to develop into malignant tumors. Colonoscopy was an important way to detect early CRC. However, as an invasive examination requiring complete and thorough intestinal preparation and other shortcomings, it is difficult to be widely accepted. The fecal occult blood testing (FOBT) with the non-invasive method was the current clinically recommended standard for early screening of CRC. However, its sensitivity was poor, generally less than 30%; it cannot meet the requirements of early diagnosis of CRC. Recent studies had shown that DNA methylation changes were closely related to the development of CRC (7) and that they run through the whole CRC development process. Second, the changes in abnormal DNA methylation often occur in the early stage of CRC. Therefore, searching for larger samples became a hot topic in order to find out reliable and good molecular biomarkers of CRC. Previous studies had identified fibrillin-1 (FBN1) as a potential optimal biomarker for early detection of CRC in relatively small population samples (8). However, the role and mechanism changing laws of FBN1 methylation had not been reported using relatively large sample data, for example, meta-analysis. In the present work, our study aimed to find out both the variation regularities and mechanisms changing the characteristics of FBN1 methylation and the relationship between FBN1 methylation in tissue colorectal cell and that in cell-free DNA feces during CRC tumorigenesis and to investigate whether FBN1 methylation acted as an early biomarker in screening of early CRC.

MATERIALS AND METHODS

The databases PubMed, Web of Science, CBA, BENDIPubmed, EMBASE, CNKI, and BAICHAIN had been searched using the systematic search method. The combination keywords were composed of FBN1, fibrillin-1, hypermethylation, methylation, adenoma, polyp, CRC, colorectal cancer, control, normal, tissue, and feces. The systematic search ended on August 4, 2021. The selected articles were also searched manually to identify other relevant independent studies. All published literature was collected in both English and Chinese languages. Based on our discussion, the independent literature that was involved in FBN1 methylation of the case–control study was identified and gathered.

Independent literature that had been searched must satisfy the following criteria: 1) the methods of FBN1 methylation experiment were shown according to methylation-specific PCR (MSP), bisulfite sequencing PCR (BSP), quantitative MSP (qMSP), and other methylation experiment methods. 2) Every literature must possess the study sample size and case–control study in tissue or feces detections. 3) The literature identified had different author names and independent samples. 4) The literature included the first author, publication year, and clinical outcomes. 5) If the data set is published in more than one literature, then only that of the most reasonable literature is included. According to the data extraction criteria above, the disagreement problems were resolved after discussions. Stepwise selection and elimination were run, and a total of 6 independent studies were ultimately received. The data set of FBN1 methylation incidences is summarized in **Table 1** (9–14). The separate information about control persons, adenomas and polyps, and carcinomas in both tissue and feces is provided in **Table 1**.

Meta-analysis software package was used with Review Manager Version 4.2 (RMV4.2). The odds ratios (ORs) and 95% CIs were calculated using the statistical analysis of RMV4.2. The heterogeneity was insignificant if $p > 0.05$; the fixed-effects model in RMV4.2 was adopted; otherwise, the random-effects model in RMV4.2 was adopted. If OR > 1, and the upper and lower limits of the interval 95% CIs were all greater than 1, there is a high risk. The more the lower value of the interval including OR exceeded 1, the riskier it was. Publication bias was assessed with forest plot in RMV4.2. When all discrete points were within the 95% region of a forest plot, and they were centralized symmetrically, there was no

TABLE 1 | The main characteristics and data of some references.

Ref	Author	Nationality	Year	Total	CRC	AP	control	Resource	Detection
(9)	Guro E	Norway	2011	446	142/179	77/111	3/156	Tissue	qMSP
(10)	Qi Guo	China	2013	150	59/75		3/75#	Tissue	MSP
(10)	Qi Guo	China	2013	105	54/75		2/30	Feces	MSP
(11)	Zhonghua	China	2015	20	7/10		0/10#	Tissue	qMSP
(12)	Wen-han	China	2015	178	69/89		3/92	Tissue	MSP
(12)	Wen-han	China	2015	119	63/89		2/30	Feces	MSP
(13)	Chao	China	2016	16	10/10		0/6	Tissue	BSP
(13)	Chao	China	2016	16	10/10		0/6	Feces	BSP
(14)	Heiying	China	2021	279	49/62	36/71	19/146	Feces	qMSP

AP, both adenomas and polyps; qMSP, quantitative methylation-specific PCR; MSP, methylation-specific polymerase chain reaction; BSP, bisulfite sequencing PCR.
#Adjacent non-cancerous.

publication bias in RMV4.2. Otherwise, if there was bias, it was evaluated according to the literature (15). All hypothesis tests concluding the differences of incidence frequency were evaluated with a 2-sided hypothesis test, and $p < 0.01$, based on statistical theory, indicated a significant difference. The curve fitting method, correlation study, and association study of the quadratic equation were calculated with SPSS Statistics 17.0 software. According to the mathematical statistical theory, the correct conclusions were given.

RESULTS

Based on the systematic search method above, we had identified the 6 independent studies on FBN1 methylation in both CRC control study and adenoma or polyp control study, including 702 tissue and 448 fecal samples in **Table 1**. Because our meta-analysis can be satisfied independent of studies being performed under the same conditions, we can analyze the DNA methylation of tissue colorectal cells and that of cell-free DNA in feces. In tissue colorectal cells, total data including 363 CRCs and 339 controls in tissues in **Table 1** were calculated with Meta-analysis software. The calculating results are shown in **Figures 1** and **2**. Because $p = 0.83$ in **Figure 1**, the fixed-effects model was adopted. The analysis indicated that FBN1 methylation risk of CRC was significantly higher than that of controls [OR (95% CI) = 124.94 (62.86–248.35) and $p < 0.00001$; **Figure 1**]. Because 100% (5/5) of the scattered points fell within the 95% confidence region of the funnel plot in **Figure 2**, and the scattered points were symmetrical, there was no publication bias; therefore, the methylation of FBN1 in tissue was associated with CRC. In cell-free DNA in feces, according to the sample data including 236 samples with CRC and 212 samples with control in **Table 1**, the sample data were calculated with Meta-analysis software; this result had shown that because $p = 0.70$ in **Figure 3**, the fixed-effects model was adopted, and that because 100% (4/4) of the scattered points fell within the 95% confidence region of funnel plot in **Figure 4**, the scattered points were relatively symmetrical; the funnel plot in **Figure 4** proved no publication bias. The result had shown that there was a significantly high risk in CRC and control feces studies [OR (95% CI) = 30.87 (16.48–57.85), and $p < 0.00001$, in **Figure 3**].

Based on control feces as a standard reference in **Table 1**, the incidence of control methylation was 23/212 = 10.8%, corresponding to the incidences of both adenoma or polyp and CRC methylation, which were 36/71 = 50.7% and 176/236 = 74.6%. The incidences of 10.8%, 50.7%, and 74.6% of FBN1 methylation from healthy control to CRC though adenoma or polyp in feces were gradually increasing; this increasing risk size was closely related to the histological process of CRC evolution. In the samples of tissue colorectal cells, we used similar methods as above; the incidence of FBN1 methylation in the corresponding control, adenoma or polyp, and CRC was 2%, 69.4%, and 79.1%, respectively. The fecal incidence was taken as a transverse coordinate and represented by Feces M (Feces Methylation); the tissue incidence was taken as an ordinate coordinate and represented by Tissue M (Tissue Methylation). The curve fitting and association study were adopted by SPSS software. The quadratic curve equation (Tissue M) = $-1.984 \times (\text{Feces M})^2 + 2.892 \times (\text{Feces M}) - 0.262$ was obtained, where the correlation coefficient R was 1, and the corresponding Sig was 0. This model had statistical significance through correlation study. The graph of this quadratic curve equation is shown in **Figure 5**. According to both the quadratic curve equation and its graph, our discoveries indicated that feces methylation was highly associated with tissue methylation and might predict the incidence of methylation in tissues.

Let us set the tumor tissue evolution process as the abscissa X, where the normal control tissue is zero, the benign disease tissue is 1, and the CRC tissue is 2. Let us set their corresponding methylation incidence of 23/212, 36/71, and 176/236 as the ordinate Y; the optimal curve fitting equation calculated by SPSS 17.0 showed that the quadratic equation $Y = 0.108 + 0.479X - 0.08X^2$ was significant. Where the correlation coefficient R = 1, the residual error was zero, and a small probability equals 0. Its corresponding figure is shown in **Figure 6**. The quadratic equation reflected the overall changing regulation of FBN1 methylation incidence in CRC lesions site during CRC development. Therefore, this overall changing regulation was gradually enhanced, promoted the transcriptional inactivation to gradually accelerate the expansion, and ultimately led to the occurrence of CRC.

DISCUSSIONS

Methylation of FBN1 associated with CRC had previously been reported (8). However, FBN1 methylation risk from control to

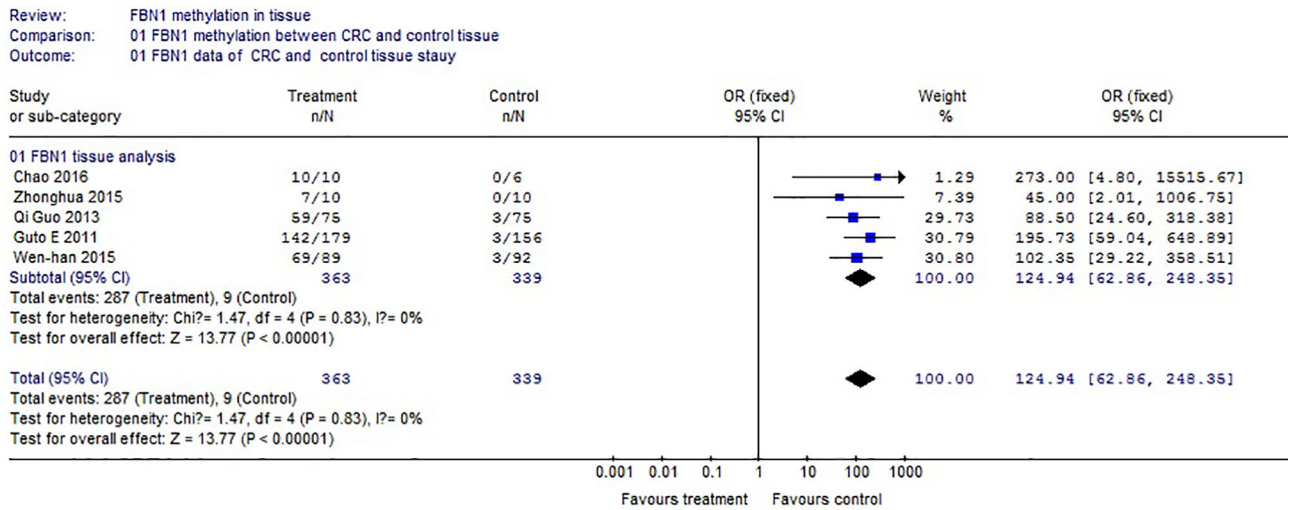


FIGURE 1 | Forest plot of colorectal cancer and control methylation in FBN1 tissue.

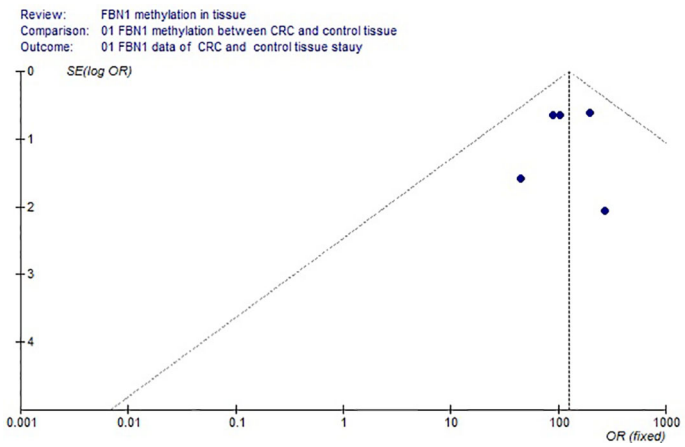


FIGURE 2 | Funnel plot of colorectal cancer and control methylation in tissue.

CRC had not been reported through larger samples with meta-analysis; the risk size of FBN1 methylation still remained unclear. We firstly used meta-analysis to explore the risk size of FBN1 methylation. These results had proved that the OR of the colorectal cell tissues was 124.94 and that the OR of the cell-free DNA in feces was 30.87 in control and CRC studies. From a large sample point of view, the risk size of FBN1 methylation in tissues was up to 124.94; it was higher than the FBN1 methylation risk in previously published articles (8, 10–12). It had clearly proved that FBN1 methylation in tissue is an important and promising biomarker for identifying CRC. In addition, from the perspective of prevention or for people with a fear of colonoscopy, particular attention was paid to ensure no damage during testing. In no-injury feces detection, FBN1 methylation risk size was up to 30.87; this higher risk had illustrated that FBN1 methylation is a no-injury fecal biomarker in clinically early CRC diagnosis and screening. These risk size of FBN1 methylation was closely associated with the occurrence and biological development of CRC.

We had for the first time studied FBN1 methylation risk from controls to CRC through adenoma or polyp. We had found that FBN1 methylation incidence of control, adenoma or polyp patients, and CRC in tissue was 2.7%, 69.4%, and 79.1%, respectively, and that in feces it was 10.8%, 50.7%, and 74.6%, respectively. The results had illustrated that FBN1 methylation risk in tissue was significantly higher than that in feces. In addition, on the basis of corresponding the incidence risk of FBN1 methylation in feces to that in tissues, there was a

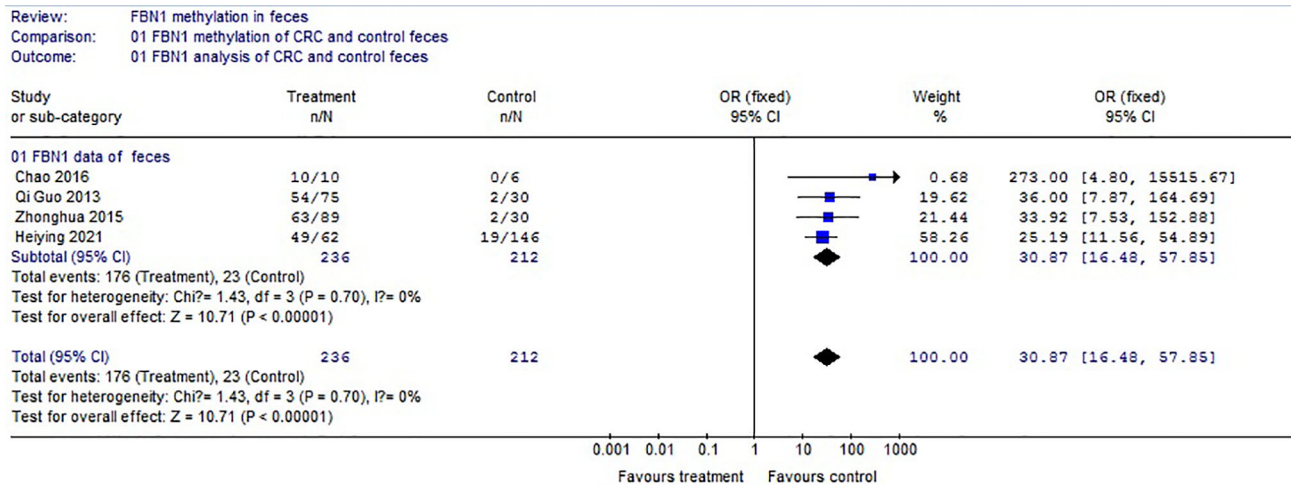


FIGURE 3 | Forest plot of both adenomas or polyps and control methylation in FBN1 feces.

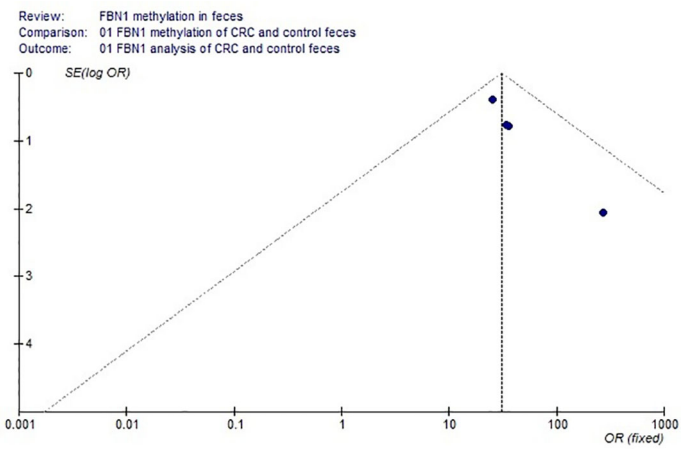


FIGURE 4 | Funnel plot of both adenomas or polyps and control methylation in FBN1 feces.

quadratic curve equation (Tissue M) = $-1.984 \times (\text{Feces M})^2 + 2.892 \times (\text{Feces M}) - 0.262$. The discoveries indicated that feces methylation was highly associated with tissue methylation, the methylation of cell-free DNA in feces reflected that of colorectal cells in tissues, and the risk sizes of FBN1 methylation from normal control to CRC though adenoma or polyp were gradually increased in colorectal tissues and feces detection. Through the test results of the experimental methylation of human feces, using the above equation, we can calculate the FBN1 methylation risk size of the human mass; the accurate assessment of tumor risk was achieved in no-injury detection.

Because FBN1 methylation incidence of individual normal people was 2.7% in tissue and 10.8% in feces as seen in **Table 1**, these FBN1 methylation abnormalities can show that the mechanism of CRC occurrence begins with normal colonic epithelium under some certain environment. Because methylation regulation was one of the important epigenetic regulators, and thus the epigenetic regulation of methylation was closely related to the transcriptional regulation of tumor-related genes (16, 17), the transcriptional inactivation associated with FBN1 methylation might start from the normal colonic epithelium.

CRC develops through an ordered series of events beginning with the transformation of normal colonic epithelium to an adenomatous intermediate and then ultimately adenocarcinoma (18, 19). This variation of CIMP pathways should be followed during tumor evolution progression. Our study revealed that the incidence of FBN1 methylation from control to CRC though adenoma or polyp can gradually produce powerful strength

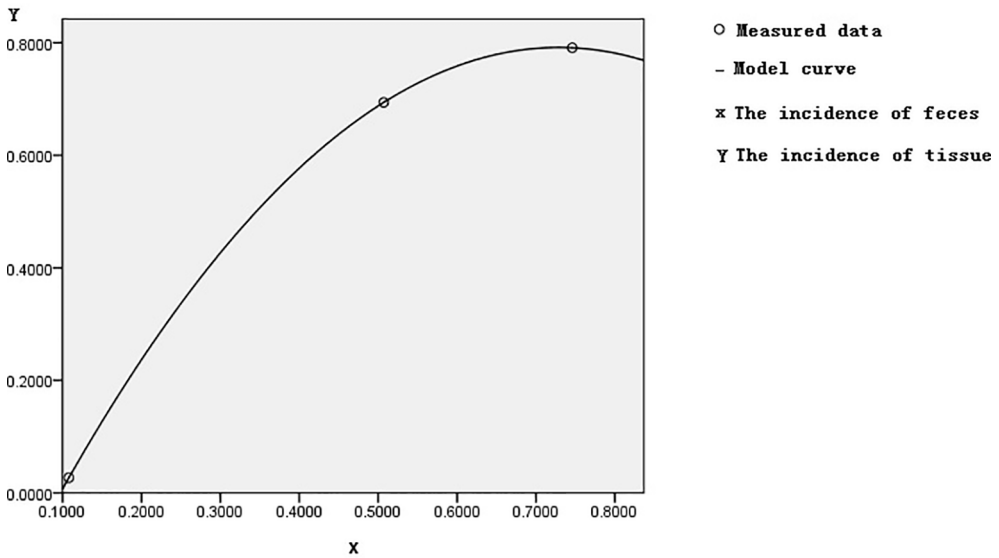


FIGURE 5 | Quadratic graph of the incidence of methylation in tissue and feces.

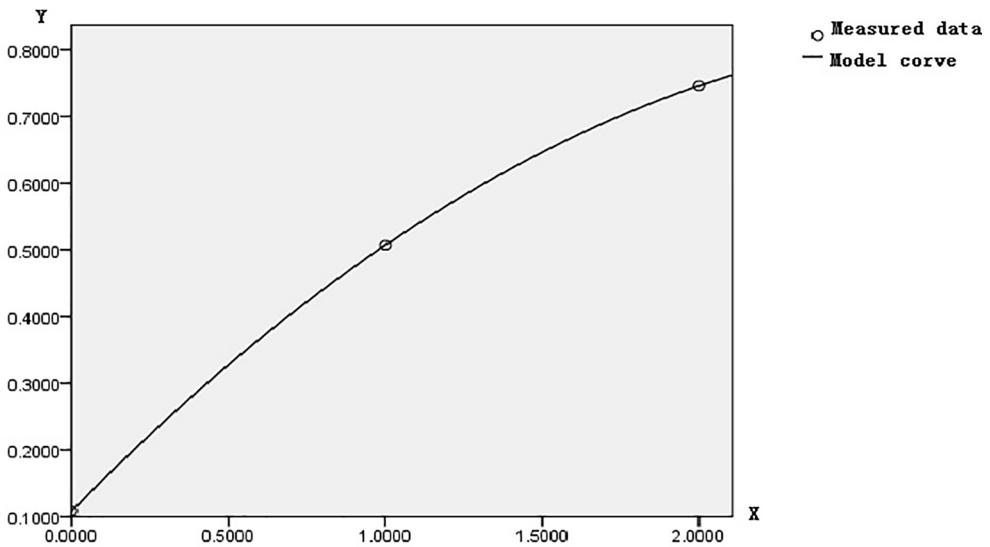


FIGURE 6 | The quadratic graph of methylation incidence in the lesion tissue.

according to the quadratic equation $Y = 0.108 + 0.479X - 0.08X^2$; the transcriptional inactivation associated with FBN1 methylation might start from normal colonic epithelium; the FBN1 methylation incidence related with the quadratic equation promoted the transcriptional inactivation to gradually accelerate the expansion and ultimately lead to the occurrence of CRC.

Because DNA methylation was catalyzed and maintained by DNMT, DNA methylation in tumor development embodied the role of tumor-related DNMT (16, 17). Based on this quadratic curve equation, the metabolizing speed of the DNMT that can catalyze FBN1 methylation might begin slowly, then gradually increase during benign diseases (adenoma or polyp), and then rapidly develop during CRC; this process of DNMT synthesizes or degrades physiological activity and participates in the transformation or progression of human cancers by mediating the methylation of cancer suppressors. DNMT, which affected FBN1 methylation by DNMT, produced the silencing of tumor suppressor miRNA-encoding genes and directly affected carcinogenesis.

The fecal detection incurs no damage; it will have a broad application prospect, especially in clinical routine. The methylation of many genes was shown to be associated with CRC. The combination of several gene methylations will be the current direction in detection. The study of CIMP had enhanced the sensitivity for cancer recurrence monitoring; people are looking for better indicators of CIMP to improve the accuracy of early screening and diagnosis of CRC, which is also the direction of future development (20, 21). However, to achieve the clinical application, a large number of problems will be solved, for example, class, gender, age, and ethnicity, fewer samples, and the impact of tumor staging, and so on (22–24); they all need to be further studied. Because the molecular mechanism and the variation of many factors of CRC are very complicated, and the clinical symptoms are very hard to detect (24, 25), it will take a long time to solve some problems of screening and early diagnosis of CRC.

CONCLUSIONS

The whole process of methylation pathogenesis during CRC development is discovered that the transcriptional inactivation associated with FBN1 methylation might start from the normal colonic epithelium and can gradually enhance, accelerate the expansion, and ultimately lead to the occurrence of CRC. The overall process of DNMT changing also has the feature of the quadratic curve of FBN1 methylation and plays a role in DNMT mechanism. FBN1 methylation is an important biomarker based on the studies of large experimental data. The risk size of fecal methylation can accurately predict that of tissue methylation in non-invasive detection.

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

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

LL, GW, and JM acquired the data and wrote the text. The other authors have made a lot of contributions.

REFERENCES

1. Patsouras MD, Vlachoyiannopoulos PG. Evidence of Epigenetic Alterations in Thrombosis and Coagulation: A Systematic Review. *J Autoimmun* (2019) 104:102347. doi: 10.1016/j.jaut.2019.102347
2. Zhou Z, Li H-Q, Liu F. DNA Methyltransferase Inhibitors and Their Therapeutic Potential. *Curr Top Med Chem* (2018) 18(28):2448–57. doi: 10.2174/1568026619666181120150122
3. Ghoshal K, Bai S. DNA Methyltransferases as Targets for Cancer Therapy. *Drugs Today (Barc)* (2007) 43(6):395–422. doi: 10.1358/dot.2007.43.6.1062666
4. Center MM, Jemal A, Ward E. International Trends in Colorectal Cancer Incidence Rates. *Cancer Epidemiol Biomark Prev* (2009) 18(6):1688–94. doi: 10.1158/1055-9965.EPI-09-0090
5. Kohler BA, Ward E, Mccarthy BJ, Schymura MJ, Ries LAG, Eheman C, et al. Annual Report to the Nation on the Status of Cancer, 1975-2007, Featuring Tumors of the Brain and Other Nervous System. *J Natl Cancer Inst* (2011) 103 (9):714–36. doi: 10.1093/jnci/djr077
6. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. *CA Cancer J Clin* (2010) 60(5):277–300. doi: 10.3322/caac.20073
7. Imperiale TF, Rans DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA Versus Fecal Occult Blood for Eolorectal Cancer Screening in All Average Risk Population. *N Engl J Med* (2004) 35(1):2704–14. doi: 10.1056/NEJMoa033403
8. Zhang Y, Wu Q, Xu L, Wang H, Liu X, Li S, et al. Sensitive Detection of Colorectal Cancer in Peripheral Blood by a Novel Methylation Assay. *Clin Epigenet* (2021) 13(1):90–3. doi: 10.1186/s13148-021-01076-8
9. Lind GE, Danielsen SA, Lothe RA. Identification of an Epigenetic Biomarker Panel With High Sensitivity and Specificity for Colorectal Cancer and Adenomas. *Mol Cancer* (2011) 10:85–7. doi: 10.1186/1476-4598-10-85
10. Guo Q, Song Y, Zhang H, Wu X, Xia P, Dang C. Detection of Hypermethylated Fibrillin-1 in the Stool Samples of Colorectal Cancer Patients. *Med Oncol* (2013) 30(4):695–8. doi: 10.1007/s12032-013-0695-4
11. Ma J, Yang Q, Deng C, Jin H, Gong W, Wang S, et al. Screening and Diagnostic Value of the Molecular Markers of DNA Methylation in Colorectal Neoplasma. *Zhonghua Wei Chang Wai Ke Za Zhi* (2015) 18(11):1149–53. doi: 10.3760/cma.j.issn.1671-0274.2015.11.020
12. Li W-H, Zhang H, Guo Q, Wu X-D, Xu Z-S, Dang C-X, et al. Detection of SNCA and FBN1 Methylation in the Stool as a Biomarker for Colorectal Cancer. *Dis Markers* (2015) 2015:657570. doi: 10.1155/2015/657570
13. Deng C, Ma J, Jin H. Application of a Novel Method to Collect Large Amount of Fecal Mucosa in Screening Colorectal Cancer. *Zhonghua Wei Chang Wai Ke Za Zhi* (2016) 19(6):680–4. doi: 10.3760/cma.j.issn.1671-0274.2016.06.020
14. Jin H, Wang J, Zhang C. The Value of Multi-Targeted Fecal DNA Methylation Detection for Colorectal Cancer Screening in a Chinese Population. *J Cancer* (2021) 12(6):1644–50. doi: 10.7150/jca.47214
15. Sui C, Wang G, Chen Q, Ma J. Variation Risks of SFRP2 Hypermethylation Between Precancerous Disease and Colorectal Cancer. *Tumour Biol* (2014) 35 (10):10457–65. doi: 10.1007/s13277-014-2313-2
16. Zuo S, Chen Y, Xu L, Tang Q, Zou S. Re-Expression of RASSF1A by 5-Aza-CdR Induced Demethylation of the Promoter Region in Human Biliary Tract Carcinoma Cells. *J Huazhong Univ Sci Technol Med Sci* (2007) 27(3):281–4. doi: 10.1007/s11596-007-0316-6
17. Chen Z, Zhang Y. Role of Mammalian DNA Methyltransferases in Development. *Annu Rev Biochem* (2020) 89:135–58. doi: 10.1146/annurev-biochem-103019-102815
18. Pino MS, Chung DC. The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* (2010) 138(6):2059–72. doi: 10.1053/j.gastro.2009.12.065

19. Nguyen LH, Goel A, Chung DC. Pathways of Colorectal Carcinogenesis. *Gastroenterology* (2020) 158(2):291–302. doi: 10.1053/j.gastro.2019.08.059
20. Sun J, Fei F, Zhang M, Li Y, Zhang X, Zhu S, et al. The Role of M SEPT9 in Screening, Diagnosis, and Recurrence Monitoring of Colorectal Cancer. *BMC Cancer* (2019) 19(1):450. doi: 10.1186/s12885-019-5663-8
21. Mojarad EN, Kuppen PJ, Aghdaei HA, Zali MR. The CpG Island Methylator Phenotype (CIMP) in Colorectal Cancer. *Gastroenterol Hepatol Bed Bench* (2013) 6(3):120–8.
22. Glöckner SC, Dhir M, Yi JM, McGarvey KE, Van Neste L, Louwagie J, et al. Methylation of TFPI2 in Stool DNA: A Potential Novel Biomarker for the Detection of Colorectal Cancer. *Cancer Res* (2009) 69(11):4691–9. doi: 10.1158/0008-5472.CAN-08-0142
23. Li DP, Zhang L, Fu JM, Huang H, Sun SM. SCTR Hypermethylation Is a Diagnostic Biomarker in Colorectal Cancer. *Cancer Sci* (2020) 111(12):4558–66. doi: 10.1111/cas.14661
24. Sharma T. Analysis of the Effectiveness of Two Noninvasive Fecal Tests Used to Screen for Colorectal Cancer in Average-Risk Adults. *Public Health* (2020) 182:70–6. doi: 10.1016/j.puhe.2020.01.021
25. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating Tumor DNA Methylation Profiles Enable Early Diagnosis, Prognosis Prediction, and Screening for Colorectal Cancer. *Sci Transl Med* (2020) 12(524):eaax7533. doi: 10.1126/scitranslmed.aax7533

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 Oncology

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fonc.2022.870863

OPEN ACCESS

EDITED BY
Daniela Turchetti,
University of Bologna, Italy

REVIEWED BY
Shirley V. Hodgson,
St George's, University of London, United Kingdom
Yiyang Wu,
Vanderbilt University Medical Center, United States

*CORRESPONDENCE
Alisa Petriina Olkinuora
alisa.olkinuora@helsinki.fi

SPECIALTY SECTION
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Oncology

RECEIVED 07 February 2022
ACCEPTED 03 October 2022
PUBLISHED 28 October 2022

CITATION
Olkinuora AP, Mayordomo AC, Kauppinen AK, Cerliani MB, Coraglio M, Collia ÁK, Gutiérrez A, Alvarez K, Cassana A, Lopéz-Köstner F, Jauk F, García-Rivello H, Ristimäki A, Koskenvuo L, Lepistö A, Nieminen TT, Vaccaro CA, Pavicic WH and Peltomäki P (2022) Mono- and biallelic germline variants of DNA glycosylase genes in colon adenomatous polyposis families from two continents.
Front. Oncol. 12:870863.
doi: 10.3389/fonc.2022.870863

Mono- and biallelic germline variants of DNA glycosylase genes in colon adenomatous polyposis families from two continents

Alisa Petriina Olkinuora[1]*, Andrea Constanza Mayordomo[2,3], Anni Katariina Kauppinen[1], María Belén Cerliani[3], Mariana Coraglio[4], Ávila Karina Collia[4], Alejandro Gutiérrez[4], Karin Alvarez[5], Alessandra Cassana[6], Francisco Lopéz-Köstner[5], Federico Jauk[7], Hernán García-Rivello[7], Ari Ristimäki[8,9], Laura Koskenvuo[10], Anna Lepistö[8,10], Taina Tuulikki Nieminen[1], Carlos Alberto Vaccaro[2,11], Walter Hernán Pavicic[2,11] and Päivi Peltomäki[1]

[1]Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki, Finland, [2]Programa de Cáncer Hereditario (Pro.Can.He.), Hospital Italiano de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina, [3]Instituto Multidisciplinario de Biología Celular (IMBICE), Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CICPBA)-Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)-Universidad Nacional de La Plata (UNLP), Buenos Aires, Argentina, [4]Unidad de Proctología, Hospital de Gastroenterología "Dr. Carlos Bonorino Udaondo", Ciudad Autónoma de Buenos Aires, Argentina, [5]Centro de cáncer, Clínica Universidad de Los Andes, Santiago, Chile, [6]Unidad de Coloproctología, Clínica Las Condes, Santiago de Chile, Chile, [7]Servicio de Anatomía Patológica, Hospital Italiano de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina, [8]Applied Tumor Genomics Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland, [9]Department of Pathology, Helsingin ja Uudenmaan Sairaanhoitopiirin Laboratorio (HUSLAB), Helsinki University Hospital and University of Helsinki, Helsinki, Finland, [10]Department of Gastroenterological Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland, [11]Instituto de Medicina Traslacional e Ingeniería Biomédica (IMTIB), Hospital Italiano de Buenos Aires (HIBA)-Instituto Universitario Hospital Italiano de Buenos Aires (IUHI)-Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires, Argentina

Recently, biallelic germline variants of the DNA glycosylase genes *MUTYH* and *NTHL1* were linked to polyposis susceptibility. Significant fractions remain without a molecular explanation, warranting searches for underlying causes. We used exome sequencing to investigate clinically well-defined adenomatous polyposis cases and families from Finland (N=34), Chile (N=21), and Argentina (N=12), all with known susceptibility genes excluded. Nine index cases (13%) revealed germline variants with proven or possible pathogenicity in the DNA glycosylase genes, involving *NEIL1* (mono- or biallelic) in 3 cases, *MUTYH* (monoallelic) in 3 cases, *NTHL1* (biallelic) in 1 case, and *OGG1* (monoallelic) in 2 cases. *NTHL1* was affected with the well-established, pathogenic c.268C>T, p.(Gln90Ter) variant. A recurrent heterozygous *NEIL1* c.506G>A, p.(Gly169Asp) variant was observed in two families. In a Finnish family, the variant occurred *in trans* with a truncating *NEIL1* variant (c.821delT). In an Argentine family, the

variant co-occurred with a genomic deletion of exons 2 – 11 of *PMS2*. Mutational signatures in tumor tissues complied with biological functions reported for *NEIL1*. Our results suggest that germline variants in DNA glycosylase genes may occur in a non-negligible proportion of unexplained colon polyposis cases and may predispose to tumor development.

KEYWORDS

DNA glycosylase, *NEIL1*, *OGG1*, *NTHL1*, *MUTYH*, polyposis, germline variant, exome sequencing

1 Introduction

Familial Adenomatous Polyposis (FAP; OMIM#175100) is characterized by multiple adenomas in the colorectum and an increased risk of colorectal cancer (CRC). In FAP, more than 100 adenomas are typically present, whereas an attenuated form (AFAP) is characterized by 10 – 100 polyps and a generally milder disease (1). Profuse or attenuated adenomatous polyposis was first associated with germline *APC* (OMIM*611731) variants in families exhibiting autosomal dominant inheritance, but has since been linked to biallelic *MUTYH* variants (MAP for *MUTYH*-associated polyposis; OMIM#608456) in families with autosomal recessive inheritance (2). *MUTYH* encodes a DNA glycosylase that acts on oxidative DNA damage by removing adenine misincorporated opposite 8-oxoG (3).

Up to 11-25% of FAP cases arise *de novo* (4, 5), and a fifth of *de novo* adenomatous polyposis cases are attributable to *APC* mosaicism (6). As much as 20% of cases with clinical features of FAP show no pathogenic variants in *APC*. Moreover, 80% patients with attenuated polyposis are molecularly unexplained (7–9). Recently, new predisposition genes for adenomatous polyposis have been identified, including *POLE* and *POLD1* (PPAP for polymerase proofreading associated polyposis; OMIM# 615083 and # 612591, respectively) (10), *AXIN2* (11), and biallelic *MSH3* (OMIM# 617100) (12) and *MLH3* (13) variants. Moreover, biallelic *NTHL1* (OMIM# 616415) variants have been associated with polyposis and multi-organ cancer predisposition (14). Except for *AXIN2*, all these genes contribute to DNA fidelity, through proofreading DNA after replication (*POLE* and *POLD1*), mismatch repair (*MSH3* and *MLH3*), or base excision repair (*NTHL1*). Like *MUTYH*, *NTHL1* targets oxidative DNA damage; it encodes a DNA glycosylase that repairs pyrimidine-derived oxidation products (3).

Inspired by recent findings of novel polyposis and cancer predisposition genes, we embarked on a study to uncover new molecular factors for unexplained polyposis cases across multiple populations by exome-wide screening. Our efforts revealed several families harboring potentially pathogenic germline variants in DNA glycosylase genes, including mono- and biallelic alterations of *NEIL1* (OMIM *608844; Endonuclease VIII-Like 1).

2 Materials and methods

2.1 Patient cohorts

This investigation was based on 67 index cases with attenuated or profuse adenomatous polyposis (34 from Finland, 21 from Chile, and 12 from Argentina) in which known genetic causes of polyposis had been excluded (*APC*, *POLE*, *POLD1*, *PTEN* and biallelic *MUTYH*; Figure 1). The cases were ascertained through the national polyposis research registries and local hospitals as described below. Most cases (47/66, 71%) exhibited attenuated polyposis. Detailed clinical data are available in Table S1. Patient DNA was extracted from blood or EBV-transformed lymphoblasts as described by Renkonen et al. (15) DNA from formalin-fixed paraffin-embedded (FFPE) samples was extracted as described by Isola et al. (16) Patient RNA was extracted from lymphoblastoid cells using the NucleoSpin RNA extraction kit (Macherey-Nagel, Düren, Germany).

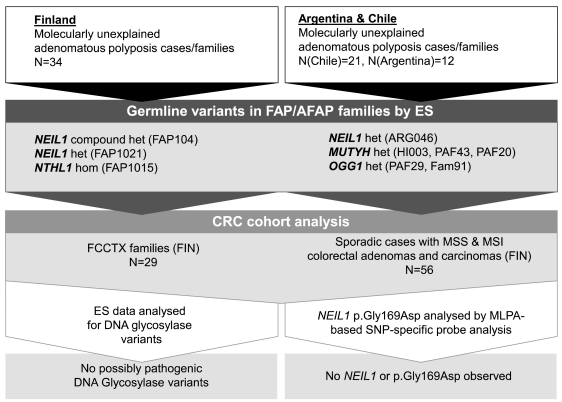
2.1.1 Finnish cohort

Most polyposis cases were attenuated (22/33, 67%; one family could not be classified) and sporadic (20/28, 71%; six families could not be classified) (Table S1). An additional cohort of 29 families representing molecularly unexplained cases with familial colorectal type X (FCCTX) (17) and a series of sporadic cases (56 individuals) with microsatellite-unstable (MSI) CRC (N=13) or microsatellite-stable (MSS) CRC (N=44) or MSS adenomas (N=15) (18) were available for comparison (Figure 1).

2.1.2 South American cohort

Thirty-three (21, Chile; 12, Argentina) unrelated families and index cases without known pathogenic variants in

A



B

Family	Gene Name	HGVSc.	HGVSp.	ClinVar	ACMG/AMP Classification	Frequency (ExAC)
FAP104, ARG046	*NEIL1*	NM_001256552.1:c.506G>A	NP_001243481.1:p.Gly169Asp		VUS - Conflicting evidence	0.001228
FAP1021	*NEIL1*	NM_001256552.1:c.692+2T>C		Benign (single submitter)	Likely benign / VUS	0.006807
FAP104	*NEIL1*	NM_001256552.1:c.821delT	NP_001243481.1:p.Ile274Thrfs*23		Pathogenic	8.236e-06
PAF29	*OGG1*	NM_016821.2:c.137G>A	NP_058214.1:p.Arg46Gln	Conflicting evidence	Likely pathogenic	0.002051
Fam91	*OGG1*	NM_016821.2:c.364G>T	NP_058214.1:p.Glu122Ter		Likely pathogenic	-
FAP1015	*NTHL1*	NM_002528.5:c.268C>T	NP_002519.1:p.Gln90Ter	Conflicting evidence (pathogenic)	Pathogenic	0.001532
HI003	*MUTYH*	NM_001128425.1:c.1147delC	NP_001121897.1:p.Ala385Profs*23	Likely pathogenic	Pathogenic	7.414e-05
PAF20	*MUTYH*	NM_001128425.1:c.1101dupC	NP_001121897.1:p.Arg368Glnfs*164	Likely pathogenic/Pathogenic	Pathogenic	1.647e-05
PAF43	*MUTYH*	NM_001128425.1:c.869G>A	NP_001121897.1:p.Cys290Tyr	Uncertain significance	Likely pathogenic	-

FIGURE 1
(A) Flowchart of the investigation including the Finnish and South American cohorts. Het denotes heterozygosity and hom homozygosity for the variants. **(B)** Detailed information of germline variants discovered in DNA glycosylase genes.

established colon polyposis-associated genes were screened as above. Most of the cases were attenuated (25/33, 76%) and had an apparent dominant mode of inheritance (16/28, 57%; five families could not be classified).

Written informed consent preceded study participation and sample donation. This study was approved by the institutional review board of the Helsinki University Central Hospital (Helsinki, Finland; Valvira/Dnro 10741/06.01.03.01/2015, 14.1.2016) and by the ethics committees of the Hospital de Gastroenterología "Dr. Carlos B. Udaondo" and Hospital Italiano de Buenos Aires (both from CABA, Argentina), and of the Clínica Las Condes (Santiago de Chile, Chile). The collection of archival specimens has been approved by the National Supervisory Authority for Welfare and Health (Valvira/Dnro 10741/06.01.03.01/2015, 14.1.2016).

2.2 Exome sequencing (ES) and germline variant selection

ES was performed at the Institute for Molecular Medicine Finland (FIMM, Helsinki, Finland) on Illumina HiSeq 2000 platform. The sequencing coverage and quality statistics for each sample are summarized in Table S2. Reads were aligned to the human reference genome hg19 using the Burrows-Wheeler Aligner version 0.6.2. Quality control and primary and secondary analysis were carried out as described by Sulonen et al. (19) Tertiary analysis was carried out using VarSeq® software (Golden Helix). Variants with allele frequency <0.003, nonsynonymous (frameshift, stop gained/lost, missense, disrupting donor/acceptor site variants) and predicted pathogenic with at least five of six programs assessing protein function *in silico* (for missense changes) were selected. All variants in DNA glycosylase genes were confirmed by Sanger sequencing with primers listed in Table S3.

2.3 Copy number variant (CNV) analysis

CNV analysis on ES data was carried out using the R package ExomeDepth (v1.1.10) (20). The patient ES data was run against appropriate patient samples with known pathogenic changes using default settings and annotated using common CNV data from the DECIPHER database (https://www.deciphergenomics.org/). All samples had a correlation score >0.99. Only CNVs with a BF score of 10 or above were considered as candidate CNVs.

2.4 Characterization of *NEIL1* variants on DNA, RNA, and protein level

To confirm that the two coding variants detected in *NEIL1* in the index case of FAP104 affected different alleles (i.e., were *in trans*), cDNA was amplified with primers NEIL1_G83D_gcDNA_F and NEIL1_G83D_cDNA_R2 (Table S3) and cloned using the

TOPO® TA Cloning® Kit for Subcloning (Thermo Fisher) according to manufacturer's instructions. Transformed *E. coli* were then grown on selective plates (100 µg Ampicillin) overnight and white colonies were grown in LB (100 µg Ampicillin) overnight. Plasmids were extracted with GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) according to manufacturer's instructions and Sanger sequenced using aforementioned primers as well as primers from Sjöblom et al. (21).

To evaluate allele-specific mRNA expression (ASE) in the lymphoblastoid cells from the index individual of FAP104, a Single Nucleotide Primer Extension (SNuPE) reaction was designed based on the heterozygous *NEIL1* c.506G>A variant. PCR products specific for cDNA (generated with primers NEIL1_G83D_gcDNA_F + NEIL1_G83D_cDNA_R2, Table S3) and gDNA (NEIL1_G83D_gcDNA_F + NEIL1_G83D_gDNA_R) served as templates for primer extensions with NEIL1_G83D_SNuPE_ext as the extension primer and ddA as the stopping nucleotide. The expected extension products were 34 bp (wild-type allele) and 24 bp (variant allele). Allele peak area ratios R<0.6 or R>1.67 indicated ASE (22).

NEIL1 mRNA expression in the lymphoblastoid cells from the index individual of FAP104 and healthy controls was evaluated by quantitative reverse transcription PCR (qRT-PCR) with TaqMan® Gene Expression Assay (Applied Biosystems) for *NEIL1* (Hs00908563_m1) and with glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) as an endogenous reference. The *NEIL1* reaction targeted exons 5 – 6 and covered the two main isoforms. The reactions were normalized against the *NEIL1* expression of healthy controls and the relative quantities were calculated using the ΔΔCT analysis.

To evaluate the stability of NEIL1 protein, lymphoblastoid cells from the index of FAP104 and unrelated healthy controls were treated with MG132 (Selleck Chemicals, Houston, Texas, USA). MG132 is a cell-permeable, proteasome inhibitor which reduces degradation of ubiquitin-conjugated proteins. Briefly, 0.5 x 10^6 cells were incubated on 6 well plates for 8 hours and treated with 10, 30 or 50 µM MG132. Proteins from the cells were extracted in LAEMMLI extraction buffer.

NEIL1 protein expression in the treated and untreated lymphoblastoid cells from the index individual of FAP104 and healthy controls was assessed by Western blotting with the primary NEIL1 rabbit polyclonal antibody (12145-1-AP, RRID:AB_2251228; Proteintech, Rosemont, IL) targeting the NEIL1 short isoform (amino acids 1 – 390). The housekeeping protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a loading control (ab128915, RRID:AB_11143050; Abcam, Cambridge, UK). P53 (#9282 RRID:AB_331476; Cell Signaling Technology, Danvers, Massachusetts, USA) was used as a technical control for MG132 experiments.

2.5 *NEIL1* promoter methylation

A custom assay utilizing Methylation-Specific Multiplex Ligation-Dependent Probe Amplification (MS-MLPA) was designed to evaluate *NEIL1* promoter methylation in constitutional and tumor tissues. The *NEIL1* promoter region was investigated with four MS-MLPA probe pairs (Table S3), of which NEIL1_1 is located just upstream of the area found to be the most informative for methylation by Chaisaingmongkol et al. (23, 24).

2.6 Somatic variant profiling

VarScan2 variant detection algorithm version 2.3.2 was applied to tumor-normal pairs to identify non-synonymous somatic variants from ES data. Annotation of the variants was done using SnpEff version 4.0 with the Ensembl v68 annotation database (https://www.ensembl.org). Variants with a somatic p-value less than 0.01 were selected for somatic mutational signature analysis, which was carried out using the R package MutationalPatterns (25). The signatures were mapped against the 30 single-base substitution (SBS) and 18 insertion-and-deletion (ID) signatures recognized by the COSMIC database (v2 for SBS and v3.1 for ID, respectively, cancer.sanger.ac.uk).

2.7 Analyses for mismatch repair (MMR) and *MUTYH* status

A colorectal tumor from ARG046 was investigated for MMR protein expression by standard immunohistochemical procedures (26). Primary antibodies used were as follows (Roche Ventana, Indiana, USA): Anti-MLH1 (M1; 790-4535, RRID:AB_2336022), anti-MSH6 (44; 790-4455, RRID:AB_2336020), anti-MSH2 (G219-1129; 760-4265, RRID:AB_2336002), and anti-PMS2 (EPR3947; 760-4531, RRID:AB_2336010). DNA from the same tumor was evaluated for *MLH1* promoter methylation by MS-MLPA using the SALSA MLPA ME011-B3 probemix (MRC-Holland, Amsterdam, the Netherlands).

Blood DNAs from the index individuals from our polyposis cohorts were evaluated for large rearrangements in MMR genes and *MUTYH* by multiplex ligation-dependent probe amplification (MLPA) according to the manufacturer's (MRC-Holland, Amsterdam, the Netherlands) instructions. SALSA MLPA P003-D1 and SALSA MLPA P072-D1 were used for *MLH1*/*MSH2* and *MSH6*/*MUTYH*, respectively, whereas *PMS2* was investigated by SALSA MLPA P008-C1. The results from fragment analysis were analyzed by Coffalyser™ (MRC-Holland, Amsterdam, the Netherlands).

2.8 Statistical analyses

Methylation ratios in sporadic tumors vs. matching normal tissues obtained from MS-MLPA analyses (Table S4) were compared using the Wilcoxon matched pairs test. IBM® SPSS® software (IBM SPSS Statistics 27, Armonk, NY: IBM Corp) was used for the analysis.

3 Results

We investigated the exomes of 67 index cases with molecularly unexplained polyposis from two continents, focusing on genes from the DNA glycosylase family. Pathogenic and likely pathogenic germline variants as well as VUSes whose pathogenicity is unknown but that have the potential of being pathogenic are described in Figure 1 and Table S5. All germline variants fulfilling our selection criteria are listed for the DNA glycosylase-associated families in Table S6.

3.1 Germline DNA glycosylase variants found in the Finnish series

The European pathogenic founder variant *NTHL1* c.268C>T, p.(Gln90Ter) (14) was detected in a homozygous state in the index individual from FAP1015 (Figure 1). This individual had attenuated polyposis and was the only member with colorectal tumor manifestations in the family (Table S1). The patient was additionally diagnosed with carcinomas of multiple organs characteristic of the tumor spectrum of *NTHL1*-associated polyposis (27).

NEIL1 variants were identified in FAP104 and FAP1021. The index of FAP104 with profuse polyposis (>200 polyps at 54 years of age) had two *NEIL1* variants (Figure 1, Figure S1); a rare missense variant c.506G>A, p.(Gly169Asp), and a very rare frameshift variant c.821delT, p.(Ile274Thrfs*23), absent in the Finnish population. A subsequent cloning assay revealed that the variants affected different alleles. All three individuals with the *NEIL1* c.506G>A variant had colorectal disease (cancer or polyps) and the same applied to the two individuals with the c.821delT variant (Figure 2).

Conflicting evidence exists regarding the pathogenic significance of *NEIL1* c.506G>A (Figure 1, Table S5). The variant allele frequency in the (global) population (0.001228) is higher than expected for a dominantly inherited disorder when comparing against *NEIL1* variants reported through diagnostics. However, allele frequency of this variant in Finns (0.0001368) is almost ten times lower. Furthermore, previous functional studies conducted on *NEIL1* c.506G>A consistently suggest pathogenicity (see Discussion). *NEIL1* c.821delT is pathogenic according to the ACMG/AMP criteria (Figure 1, Table S5). Suitable biological specimens were available from the index of FAP104 to explore the consequences of the *NEIL1* variants on RNA and protein level. We evaluated the relative mRNA expression from the two *NEIL1* alleles by SNuPE and found that the frameshift variant containing transcripts were approximately twice less abundant than the missense variant containing transcripts in lymphoblastoid cells from the index individual of FAP104 (Figure 3A). By qRT-PCR, the total *NEIL1* mRNA expression was essentially lower than in healthy controls studied for comparison (Figure 3B), suggesting that the ASE seen by SNuPE was more likely to reflect decreased expression from the frameshift allele than increased expression from the missense allele. Interestingly, Western blot analysis revealed a markedly elevated amount of normalized full-length NEIL1 protein compared to healthy controls, and no truncated protein was visible (Figure 3C). The abundant full-length protein likely originated from the missense allele, and no stable protein was apparently generated from the frameshift variant containing allele. In the absence of increased *NEIL1* mRNA expression (Figure 3B), elevated NEIL1 protein in the Western blot was more likely to reflect aberrant protein stabilization than overexpression. The MG132 experiments (see Materials and Methods) did not reveal increased NEIL1 staining after treatment, indicating that regulation of NEIL1 protein expression is not MG132 mediated.

The index of FAP1021 with attenuated polyposis (30 polyps at 72 years of age) had a splice donor variant c.692+2T>C (Figure 1, Figure S1). In the literature, conflicting interpretations of pathogenicity for this splice variant exist (e.g., Dallosso AR et al. (28); Boldinova EO et al. (29)). In the absence of RNA, we were unable to experimentally verify splicing consequences of the variant. Based on available data, the ACMG/AMP classification is likely benign or VUS (Table S5). Available *in silico* software evaluated the splice donor variant highly likely to affect splicing (0.9918, 0.6039, 0.96, and 0.99683 for ADA, RF, SpliceAI, and SPiCE, respectively).

3.2 Germline DNA glycosylase variants found in the South American series

A patient from the Argentine family ARG046 with attenuated mixed polyposis and colorectal carcinoma at the age of 60 years was found to be heterozygous for the previously described *NEIL1* c.506G>A variant (Figure 2). No other possibly pathogenic variants in *NEIL1* were observed in the South American series.

Two families revealed likely deleterious *OGG1* variants. The index of family PAF29 with attenuated polyposis had a heterozygous missense variant of *OGG1*, c.137G>A, p.(Arg46Gln). In the literature, the same variant was described

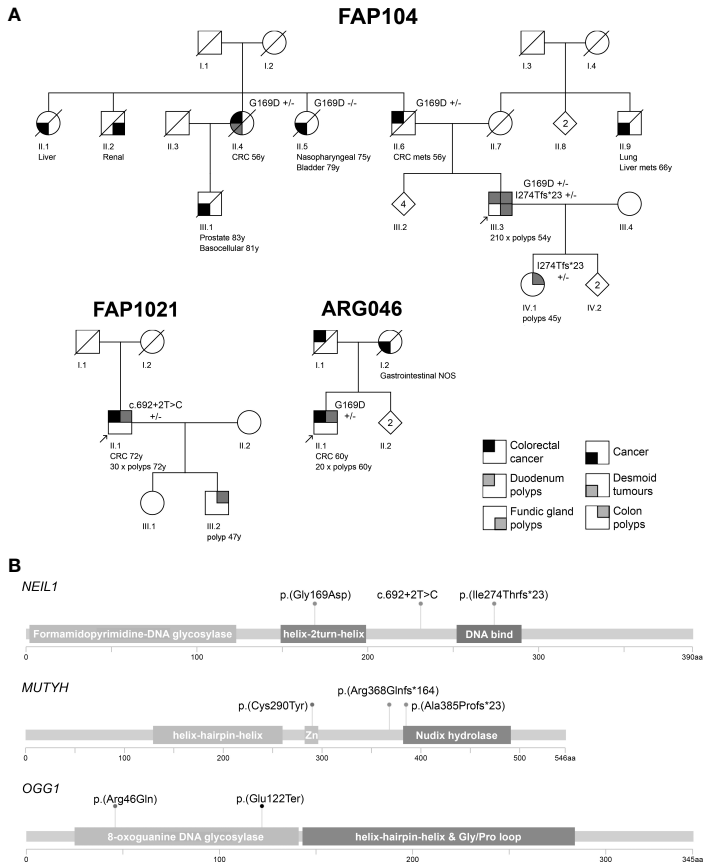


FIGURE 2
(A) Pedigrees of the polyposis families with *NEIL1* variants. Numbers below the symbols are patient identifiers. Arrow denotes the index person. Zygosity of *NEIL1* variants is shown (+/– heterozygous). Tumor manifestations and age at diagnosis (years) are given below the patient symbol. Mets refers to metastasis. Nonessential pedigree features were removed or modified to protect confidentiality. **(B)** Locations of the variants relative to the main functional domains of the DNA glycosylase genes. Zn denotes the metal binding sites in *MUTYH* as listed in the Uniprot database (www.uniprot.org; Q9UIF7). The pedigrees were generated with Pedigree Chart Designer and the lollipop diagrams with MutationMapper.

in a patient with synchronous colorectal cancer at 36 years and adenomas (30). It was shown that the G to A change which affects the last nucleotide of exon 1 disrupts a splice donor sequence, resulting in extinct expression from the variant allele in cDNA from the patient (30). The authors classified the *OGG1* c.137G>A variant pathogenic. Considering all available information, the ACMG/AMP criteria for likely pathogenic are fulfilled (Figure 1, Table S5). A heterozygous c.364G>T, p.(Glu122Ter) nonsense variant in the *OGG1* gene, likely pathogenic by the ACMG/AMP criteria (Figure 1, Table S5), was detected in family 91. The variant was present in the index patient (ID 606) with attenuated polyposis but absent in the index patient's brother (ID 657) with late-onset colorectal carcinoma (Table S1).

Three heterozygous *MUTYH* variants were observed in the South American series (Figure 1, Figure S2). By ES, a frameshift variant of *MUTYH*, c.1101dupC, p.(Arg368Glnfs*164), was present in three individuals (ID 47, 534, and 535) out of four with colorectal adenomas or carcinoma from the Chilean family PAF20 and affected two generations. *MUTYH* c.1147delC, p.(Ala385Profs*23) was observed in the index individual HI003 (no other affected members were known to exist in this family). Both *MUTYH* variants described above are pathogenic by the ACMG/AMP criteria (Table S5), with biallelic involvement linked to MAP. A missense variant, *MUTYH* c.869G>A, p.(Cys290Tyr), classified as likely pathogenic (Table S5), was found in the index individual of family PAF43 (carrier statuses of the remaining family members were unknown). This family showed features of MAP (over 100 polyps in the index individual and an apparent recessive transmission pattern, Figure S2), raising the possibility that the *MUTYH* allele currently considered wildtype might harbor a defect that had escaped detection. However, manual IGV analysis of the gene and MLPA (with *MSH6-MUTYH* and *APC* MLPA kits) for large genomic rearrangements provided no support for biallelic *MUTYH* involvement.

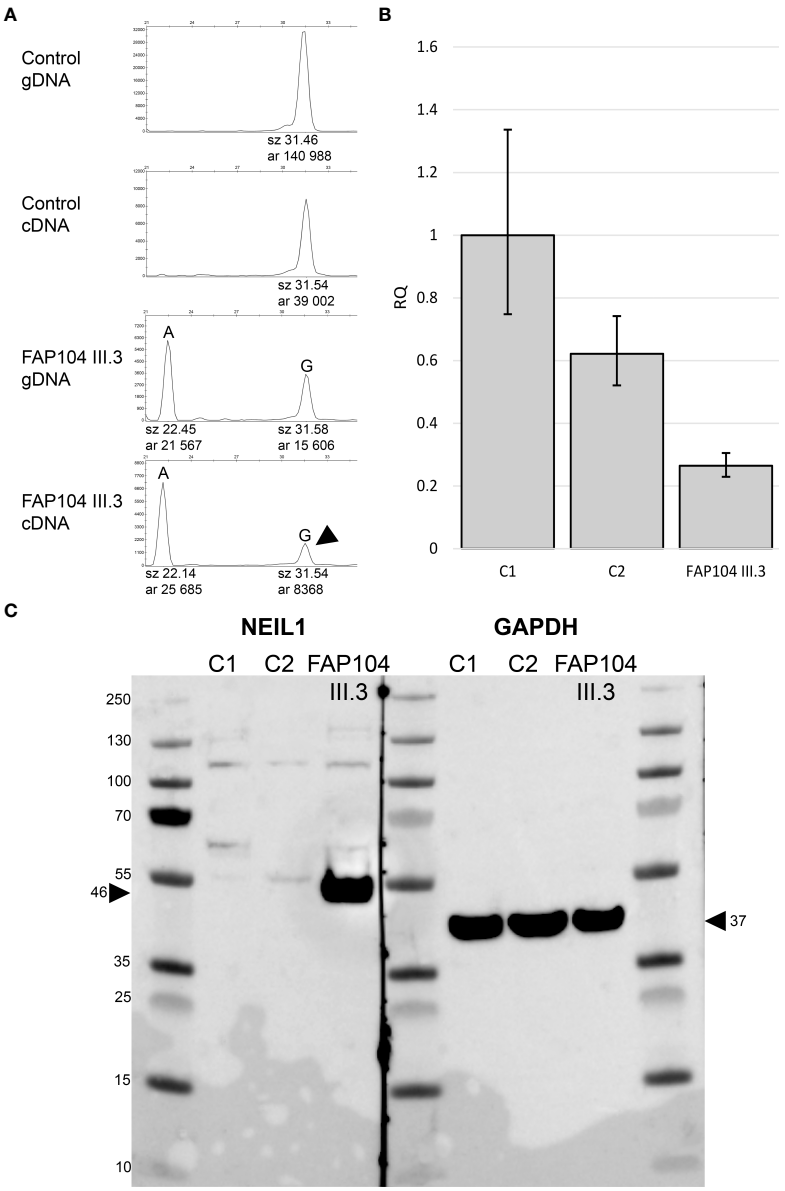


FIGURE 3
Expressional consequences of the *NEIL1* c.506G>A, p.(Gly169Asp) variant. **(A)** ASE analysis based on the *NEIL1* c.506G>A variant. Longer peak (G) represents the wild type sequence. The control individual is homozygous for the wild-type allele. The index individual of FAP104 is heterozygous: Allele A corresponds to the *NEIL1* c.506A missense variant, whereas the G allele is known to have a frameshift variant (c.821delT) in a downstream position. This individual displays ASE with the peak area ratio of 0.45 for G to A in cDNA relative to gDNA. The result indicates that transcripts with G (arrowhead) having the frameshift variant are twice less abundant than transcripts containing the missense variant **(A)**. **(B)** Relative quantity (RQ) from the qRT-PCR experiment targeting the two main isoforms of *NEIL1* using the housekeeping gene *GAPDH* as an endogenous control. Whiskers indicate 95% confidence limits. The index of family FAP104 shows reduced *NEIL1* RNA expression compared to the controls. **(C)** Western blot of two healthy control individuals and the index of FAP104. GAPDH was used as a loading control. FAP104 index displays elevated NEIL1 protein levels compared to the controls (arrowhead). No truncated NEIL1 protein is seen.

3.3 Mutational analyses on tumors from individuals with *NEIL1* variants

DNA was available from a colorectal tubular adenoma from the index of FAP104 (compound heterozygous *NEIL1*; c.506G>A and c.821delT), two desmoid tumors from the paternal aunt of the index of FAP104 (heterozygous *NEIL1* c.506G>A), and a colorectal carcinoma from the index of ARG046 (heterozygous *NEIL1* c.506G>A) for somatic mutational profiling. We first determined the total mutational

loads, since elevated numbers of somatic variants may point to defects in DNA replication or repair (10, 15). The total numbers of somatic nonsynonymous variants were 281 (adenoma), 45 and 57 (desmoids), and 1146 (carcinoma) by VarScan2 analysis. Based on the commonly used threshold of 10 variants/Mb, only the carcinoma of ARG046 was hypermutated (35 somatic variants/Mb).

All somatic variants meeting our selection criteria (VarScan2 p<0.01) are listed in Table S7. No somatic variant or loss of heterozygosity of *NEIL1* was observed in any sample. Thus, there was no evidence of a somatic "second hit" to the remaining wildtype allele in the monoallelic *NEIL1* variant carriers. The adenoma from the index of FAP104 showed a truncating *APC* variant (c.4666dupA, p.Thr1556fs; variant allele frequency (VAF) 23%) and *KRAS* c. 35G>C, p.Gly12Ala (VAF 25%), both representing alterations typical of colorectal tumorigenesis. The two desmoid tumors revealed extensive sharing of somatic variants, suggesting a common origin for the tumors.

As the patterns of somatic variants can offer insights to the underlying biological processes, a mutational signature analysis was conducted on the tumors (Figures 4A, B). VarScan2-based somatic variants were included in this analysis. COSMIC (31) SBS signature 3 (defective homologous recombination) was prominent in all three tumors from FAP104 (Figure 4A). Desmoid tumors from individual II.4 additionally revealed SBS7 (ultraviolet radiation exposure) and a discernible SBS24 linked to aflatoxin-associated mutagenesis (32). Interestingly, the hypermutable colorectal carcinoma from ARG046 showed prominent MMR deficiency-associated signatures SBS6 and SBS26, together with SBS12 (unknown etiology). The ID signature 6 supported defective homologous recombination in tumors from FAP104, whereas ID7 was compatible with deficient MMR in the colorectal carcinoma from ARG046 (Figure 4B).

3.4 *PMS2* genomic deletion found in the index case of ARG046

To resolve the MMR-deficient pattern of somatic alterations in the colorectal carcinoma from ARG046, the tumor was tested for *MLH1* promoter methylation, but no hypermethylation was present. However, immunohistochemical analysis revealed selective absence of PMS2 protein (Figure 4C). Subsequent MLPA analysis of blood DNA showed a heterozygous deletion of *PMS2* exons 2 - 11 (NM_000535.5:c.(23 + 1_24-1)-(2006 + 1_2007-1)del; Figure 4D). No additional cases with large rearrangements of MMR genes were detected when our entire polyposis series was evaluated by MLPA (and no small sequence alterations with possible pathogenicity existed in MMR genes by ES).

3.5 Methylation status of *NEIL1* in the polyposis and control cohorts

As *NEIL1* is commonly hypermethylated in cancer (23, 24), we designed a MS-MLPA kit to determine constitutional and somatic methylation status of our patient samples. Of the four MS-MLPA probe pairs, NEIL1_1 interrogated a region previously shown to be informative for methylation (23, 24) and showed the best discrimination between normal and tumor tissues (Table S4). Blood and normal colonic mucosae even from reference individuals revealed considerable methylation, and examination of blood DNAs from our polyposis cases (with or without *NEIL1* variants) raised no suspicion of constitutional *NEIL1* epimutation in any case. Compared to paired normal tissues, tumors from individuals with *NEIL1* variants occasionally displayed higher methylation dosage ratios, but no significant somatic hypermethylation of the promoter region was evident. Comparing paired tumor and normal tissues from sporadic cases with MSS or MSI carcinomas or adenomas revealed no significant difference by Wilcoxon matched pairs test (Z=-1.03, p=0.133 for MSI carcinomas vs matching normal tissues, and Z=-0.217, p=0.828 for MSS carcinomas vs matching normal tissues by *NEIL1* I probe, respectively).

4 Discussion

The DNA glycosylase family comprises eleven members, of which some (e.g., *MUTYH*) are monofunctional (capable of excising damaged or mispaired bases) and some (e.g., *NTHL1*, *OGG1*, and *NEIL1*) bifunctional (additionally having endonuclease activity to incise the modified strand) (33). The role of DNA glycosylases other than *MUTYH* and *NTHL1* in (colon) tumor susceptibility is unknown and/or associated with conflicting evidence, which encouraged us to undertake the present study. In our exomic screen of 67 index cases from Finnish and South American cohorts, 9 (13%) revealed proven or potentially pathogenic germline variants affecting *NEIL1* (3 cases), *MUTYH* (3 cases), *NTHL1* (1 case), and *OGG1* (2 cases). The findings suggest that germline variants in DNA glycosylase genes may explain a nontrivial proportion of unexplained cases of colorectal polyposis.

In our investigation, *NTHL1* showed biallelic involvement, consistent with the recessive *NTHL1*-associated polyposis syndrome (14). Our *OGG1* variants were monoallelic and suggested dominant transmission with reduced penetrance, which agrees with available literature (30), although a single case with a biallelic truncating *OGG1* variant was recently reported in association with FCCTX (34). Biallelic germline

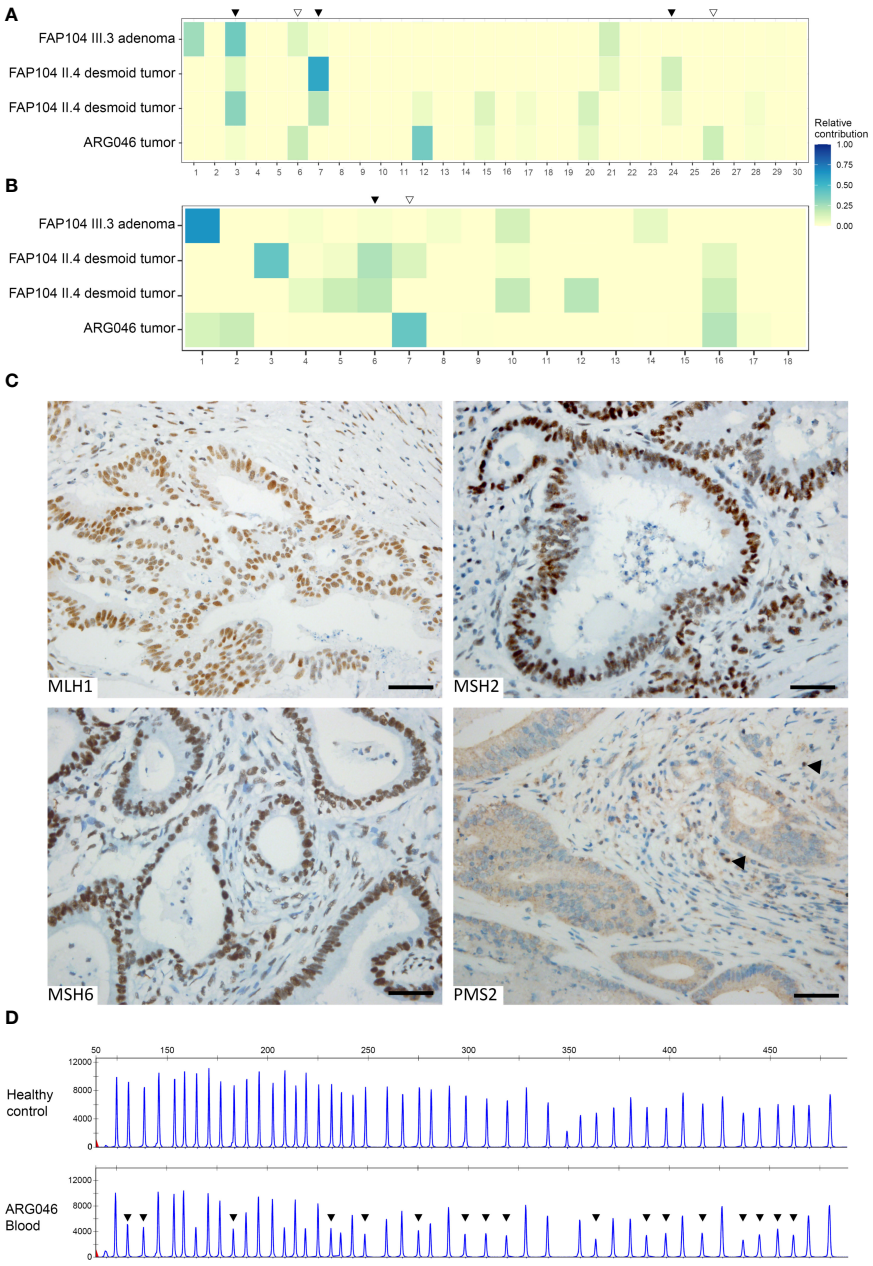


FIGURE 4
Somatic mutational signature analysis of four *NEIL1*-associated tumors **(A, B)**. **(A)** Heatmap indicating the relative contribution of SBS signatures (COSMICv2) to the mutational landscape of each tumor. Black arrowheads indicate *NEIL1*-deficiency associated signatures prominent in tumors from FAP104, whereas open arrowheads represent MSI-signatures present in a colorectal carcinoma from ARG046. **(B)** Heatmap of the relative contributions of ID signatures (COSMICv3.1, GRCh37) to the mutational profiles of the tumors. Subsequent discovery of the *PMS2* alteration in the ARG046 case **(C, D)**. **(C)** Immunohistochemical analysis of the MMR proteins reveals a selective loss of PMS2 in the tumor cells. Normal cells retaining the PMS2 expression are indicated with arrowheads. Scale bar represents 50 µm. **(D)** *PMS2*-MLPA analysis of blood DNA of the ARG046 case as well as a healthy control. Arrowheads indicate reduced emission peaks at exons 2-11. The average probe ratios of exons 2-11 (0.54 ± 0.03) are indicative of a heterozygous deletion.

variants of *MUTYH* underlying the well-defined recessive MAP syndrome (2) were excluded from our series at the outset; the significance of the observed monoallelic *MUTYH* variants will be addressed below. Finally, the transmission pattern of *NEIL1*-associated disease is unclear since no segregation studies for *NEIL1* variants have been reported before. We detected one biallelic and two monoallelic *NEIL1* cases that will be discussed in more detail below.

The *NEIL1* missense variant c.506G>A, p.(Gly169Asp) occurred in two polyposis families, Finnish and Argentine (2/67, 3%). This variant was previously referred to as G83D according to annotation based on the short (390 amino acid) isoform. Forsbring et al. (35) found this variant in two patients among 37 with primary sclerosing cholangitis and cholangiocarcinoma (5%). Biochemical studies have provided consistent evidence that the variant is deleterious. Using 8-oxoG, thymine glycol, and 5-OHU as substrates, the NEIL1 Gly169Asp protein was found to be devoid of DNA glycosylase activity (35–38). Galick et al. (38) additionally showed that the variant NEIL1 protein acted as a dominant negative manner relative to the wild-type protein, being able to bind to damaged DNA but unable to repair it. Roy et al. (37) concluded that in individuals with the Gly169Asp variant, NEIL1 function is likely to be 50% compared to normal levels unless compensatory mechanisms exist. Our Western blot analysis on lymphoblastoid cells from the index of FAP104 with the c.506G>A variant revealed strikingly increased amount of NEIL1 protein, and we hypothesize that the c.506G>A variant is mutagenic due to the accumulation of functionally defective protein. Our result would comply with a possible oncogenic role proposed for *NEIL1* in some studies (39).

The index individual of FAP104 was compound heterozygous for *NEIL1* c.506G>A and c.821delT. We are not aware of the possible existence of any previous reports of biallelic constitutional *NEIL1* involvement in association with human disease. Moreover, in FAP104, all five members who were verified to have either one of the *NEIL1* variants (or both) had a colorectal tumor phenotype. The age at onset of disease (polyposis or cancer) of our heterozygous cases was relatively late with modest numbers of polyps (Figure 2, Table S1), which may indicate reduced penetrance. Apart from colon polyposis, profuse gastric fundic gland polyposis was apparent in the index of FAP104 as well as his daughter, both individuals with the *NEIL1* frameshift variant (Figure 2). Stomach tissue is particularly prone to oxidative damage and some somatic *NEIL1* variants and germline polymorphisms have been found in gastric cancer patients (40) indicating a possible role in stomach polyp formation.

SBS3 and ID6, which are associated with impaired homologous recombination (41), stood out among mutational signatures observed in our *NEIL1*-associated tumors (Figures 4A, B). This is compatible with observations that NEIL1 may participate in the repair of oxidized bases in D-loops (42) and R-loops (43) arising during homologous recombination or transcription. SBS7 which is connected to UV radiation was prominent in the desmoid tumors from a case with the *NEIL1* c.506G>A variant. This is consistent with findings of *Neil1*$^{-/-}$ mice being sensitive to chronic UVB exposure (44). Our desmoid tumors also exhibited SBS24, the so-called aflatoxin signature. McCullough and Lloyd (32) demonstrated that *NEIL1* is a major contributing factor to the repair of AFB_1-N7-dG and AFB_1-Fapy-dG adducts formed by aflatoxin mutagenesis. All in all, mutational signatures observed in tumors from our *NEIL1* cases are well in agreement with the reported biological consequences of defective *NEIL1* function.

Somatic hypermutability and MMR deficiency-associated signatures in a colorectal tumor from an Argentine case with the *NEIL1* c.506G>A variant provided critical clues to discover a large genomic deletion of *PMS2* as a concomitant germline alteration in this patient (Figures 4C, D). In analogy to DNA glycosylase genes, incomplete penetrance characterizes many pathogenic variants of *PMS2* (45). Colonic polyposis commonly accompanies biallelic *PMS2* variants, whereas monoallelic *PMS2* variants typically manifest themselves as (late-onset) colorectal carcinoma (46). The *PMS2* exon 2 – 11 deletion found in our ARG046 case was heterozygous (Figure 3D), and no other *PMS2* sequence variants of suspected pathogenic significance were identified (Table S6). Modifying or additive effects of two or more defective genes may be necessary to explain the observed phenotypes of DNA glycosylase gene variants (30) and *PMS2* variants (46), and base excision repair and MMR defects can potentiate each other's effects (47). It is possible that the late-onset colorectal carcinoma in our ARG046 case mainly reflected the *PMS2* defect, in agreement with available literature (see above), whereas *NEIL1* c.506G>A might be necessary for the patient's polyposis phenotype.

While biallelic germline variants of *MUTYH* cause predisposition to MAP (see Introduction), the clinical phenotype of monoallelic *MUTYH* variants remains unsettled. In our investigation, three families from the South American cohort revealed monoallelic *MUTYH* variants classified as pathogenic (two) and likely pathogenic (one). Among five individuals with monoallelic *MUTYH* variants, four exhibited polyposis with the polyp number ranging from below 20 (in three individuals) to over 100 (in one), and three had late-onset colorectal cancer (Table S1, Figure S2). Our findings together with published reports indicate that individuals with monoallelic *MUTYH* variants may be predisposed to colorectal polyposis of a variable degree and have a moderately increased risk of colorectal cancer (48, 49). The PAF43 index case manifested a phenotype akin to classical MAP, but no second *MUTYH* variant of possible pathogenic significance was identified. Since the ES runs included only about 300 bp flanking sequence, our approach does not exclude possible variants in regulatory regions (including deep intronic splice variants and pseudoexons).

In summary, we describe proven or possibly pathogenic germline variants of DNA glycosylase genes in 9/67 (13%) index cases with colon polyposis. Our study suggests a link between *NEIL1* germline variants and colon polyposis. Because of the relatively limited number of individuals with *NEIL1* variants in this investigation, our findings need to be confirmed in larger multinational cohorts.

Data availability statement

The data analyzed in this study was obtained as described in Materials and Methods. The datasets are not readily available because our IRB approvals do not allow sharing raw exome

sequencing data. However, all variants that fulfilled our selection criteria can be found in the Supplementary Tables. Requests to access these datasets should be directed to the corresponding author, APO, alisa.olkinuora@helsinki.fi.

Ethics statement

The studies involving human participants were reviewed and approved by Helsinki University Central Hospital (Helsinki, Finland; Valvira/Dnro 10741/06.01.03.01/2015, 14.1.2016) Hospital de Gastroenterología "Dr. Carlos B. Udaondo" and Hospital Italiano de Buenos Aires (both from CABA, Argentina) Clínica Las Condes (Santiago de Chile, Chile) by the National Supervisory Authority for Welfare and Health (Valvira/Dnro 10741/06.01.03.01/2015, 14.1.2016). The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, APO, PP, CAV, and WHP. Data curation, FIMM. Formal analysis, APO, AKK, PP, ACM, and MBC. Funding acquisition, PP, CAV, and WHP. Investigation, APO, AKK, AR, PP, ACM, MBC, FJ, HG-R, and WHP. Methodology, APO, AKK, TTN, PP, ACM, MBC, and WHP. Project administration, PP and WHP. Resources, LK, AL, PP, ACM, MBC, MC, AKC, AG, KA, AC, FL-K, CAV, and WHP. Software, APO. Supervision, TTN, PP, and WHP. Validation, APO, AKK, TTN, PP, ACM, MBC, and WHP. Visualization, APO, PP, and WHP. Writing-original draft, APO, PP, and WHP. Writing-review and editing, APO, PP, and WHP. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from Biomedicum Helsinki Foundation (to APO); the Mary and Georg Ehrnrooth's Foundation (to LK); the Jalmari and Rauha Ahokas Foundation (to TTN); the Maud Kuistila Memorial Foundation (to TTN); the Jane and Aatos Erkko Foundation (to PP); the Academy of Finland (grant number 330606 to PP); Cancer Foundation Finland sr (to PP, AR and LK); and Helsinki University Central Hospital Research Funds (to AR); the Sigrid Juselius Foundation (to PP and AR); the HiLIFE Fellows 2017-2020 (to PP); the Agencia Nacional de Promoción Científica y Tecnológica (PICT_2017 -3210 to CAV and -3197 to WHP) and Instituto Nacional del Cáncer Argentina (AF_IV to CAV and WHP). The Doctoral Programme in Biomedicine of Helsinki University offered a paid doctoral student position to APO.

Acknowledgments

Saila Saarinen is thanked for expert laboratory assistance, and Maija Röntynen and Beatriz Alcala-Repo for collecting clinical data.

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/fonc.2022.870863/full#supplementary-material

References

1. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis* (2009) 4:22. doi: 10.1186/1750-1172-4-22

2. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* (2002) 30(2):227–32. doi: 10.1038/ng828

3. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: Mechanisms, mutation, and disease. *FASEB J* (2003) 17(10):1195–214. doi: 10.1096/fj.02-0752rev

4. Aretz S, Uhlhaas S, Caspari R, Mangold E, Pagenstecher C, Propping P, et al. Frequency and parental origin of *de novo* APC mutations in familial adenomatous polyposis. *Eur J Hum Genet* (2004) 12(1):52–8. doi: 10.1038/sj.ejhg.5201088

5. Aretz S, Stienen D, Friedrichs N, Stemmler S, Uhlhaas S, Rahner N, et al. Somatic APC mosaicism: A frequent cause of familial adenomatous polyposis (FAP). *Hum Mutat* (2007) 28(10):985–92. doi: 10.1002/humu.20549

6. Jansen AM, Crobach S, Geurts-Giele WR, van den Akker BE, Garcia MV, Ruano D, et al. Distinct patterns of somatic mosaicism in the APC gene in

neoplasms from patients with unexplained adenomatous polyposis. *Gastroenterology* (2017) 152(3):546–9.e3. doi: 10.1053/j.gastro.2016.10.040

7. Spier I, Horpaopan S, Vogt S, Uhlhaas S, Morak M, Stienen D, et al. Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. *Hum Mutat* (2012) 33(7):1045–50. doi: 10.1002/humu.22082

8. Pavicic W, Nieminen TT, Gylling A, Pursiheimo JP, Laiho A, Gyenesei A, et al. Promoter-specific alterations of APC are a rare cause for mutation-negative familial adenomatous polyposis. *Genes Chromosomes Cancer* (2014) 53(10):857–64. doi: 10.1002/gcc.22197

9. Shirts BH, Salipante SJ, Casadei S, Ryan S, Martin J, Jacobson A, et al. Deep sequencing with intronic capture enables identification of an APC exon 10 inversion in a patient with polyposis. *Genet Med* (2014) 16(10):783–6. doi: 10.1038/gim.2014.30

10. Bellido F, Pineda M, Aiza G, Valdés-Mas R, Navarro M, Puente DA, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: Review of reported cases and recommendations for genetic testing and surveillance. *Genet Med* (2016) 18(4):325–32. doi: 10.1038/gim.2015.75

11. Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, et al. Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* (2004) 74(5):1043–50. doi: 10.1086/386293

12. Adam R, Spier I, Zhao B, Kloth M, Marquez J, Hinrichsen I, et al. Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis. *Am J Hum Genet* (2016) 99(2):337–51. doi: 10.1016/j.ajhg.2016.06.015

13. Olkinuora A, Nieminen TT, Mårtensson E, Rohlin A, Ristimäki A, Koskenvuo L, et al. Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet Med* (2018) 21(8):1868–73. doi: 10.1038/s41436-018-0405-x

14. Weren RD, Ligtenberg MJ, Kets CM, de Voer RM, Verwiel ET, Spruijt L, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat Genet* (2015) 47(6):668–71. doi: 10.1038/ng.3287

15. Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, et al. Altered expression of MLH1, MSH2, and MSH6 in predisposition to hereditary nonpolyposis colorectal cancer. *J Clin Oncol* (2003) 21(19):3629–37. doi: 10.1200/JCO.2003.03.181

16. Isola J, DeVries S, Chu L, Ghazvini S, Waldman F. Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples. *Am J Pathol* (2014) 145(6):1301–8. doi: 10.1016/S0165-4608(96)00076-3

17. Nieminen TT, Abdel-Rahman WM, Ristimäki A, Lappalainen M, Lahermo P, Mecklin JP, et al. BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology* (2011) 141(1):e23–6. doi: 10.1053/j.gastro.2011.03.063

18. Joensuu EI, Abdel-Rahman WM, Ollikainen M, Ruosaari S, Knuutila S, Peltomäki P. Epigenetic signatures of familial cancer are characteristic of tumor type and family category. *Cancer Res* (2008) 68(12):4597–605. doi: 10.1158/0008-5472.CAN-07-6645

19. Sulonen AM, Ellonen P, Almusa H, Lepistö M, Eldfors S, Hannula S, et al. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* (2011) 12(9):R94. doi: 10.1186/gb-2011-12-9-r94

20. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* (2012) 28(21):2747–54. doi: 10.1093/bioinformatics/bts526

21. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* (2006) 314(5797):268–74. doi: 10.1126/science.1133427

22. Nieminen TT, Pavicic W, Porkka N, Kankainen M, Järvinen HJ, Lepistö A, et al. Pseudoexons provide a mechanism for allele-specific expression of APC in familial adenomatous polyposis. *Oncotarget* (2016) 7(43):70685–98. doi: 10.18632/oncotarget.12206

23. Chaisaingmongkol J, Popanda O, Warta R, Dyckhoff G, Herpel E, Geiselhart L, et al. Epigenetic screen of human DNA repair genes identifies aberrant promoter methylation of NEIL1 in head and neck squamous cell carcinoma. *Oncogene* (2012) 31(49):5108–16. doi: 10.1038/onc.2011.660

24. Shinmura K, Kato H, Kawanishi Y, Igarashi H, Goto M, Tao H, et al. Abnormal expressions of DNA glycosylase genes NEIL1, NEIL2, and NEIL3 are associated with somatic mutation loads in human cancer. *Oxid Med Cell Longev* (2016) 2016:1546392. doi: 10.1155/2016/1546392

25. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med* (2018) 10(1):33. doi: 10.1186/s13073-018-0539-0

26. Overbeek LI, Ligtenberg MJ, Willems RW, Hermens RP, Blokx WA, Dubois SV, et al. Interpretation of immunohistochemistry for mismatch repair proteins is only reliable in a specialized setting. *Am J Surg Pathol* (2008) 32(8):1246–51. doi: 10.1097/PAS.0b013e31816401bb

27. Grolleman JE, de Voer RM, Elsayed FA, Nielsen M, Weren RDA, Palles C, et al. Mutational signature analysis reveals NTHL1 deficiency to cause a multi-tumor phenotype. *Cancer Cell* (2019) 35(2):256–66.e5. doi: 10.1016/j.ccell.2018.12.011

28. Dallosso AR, Dolwani S, Jones N, Jones S, Colley J, Maynard J, et al. Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. *Gut* (2008) 57(9):1252–5. doi: 10.1136/gut.2007.145748

29. Boldinova EO, Khairullin RF, Makarova AV, Zharkov DO. Isoforms of base excision repair enzymes produced by alternative splicing. *Int J Mol Sci* (2019) 20 (13):3279. doi: 10.3390/ijms20133279

30. Morak M, Massdorf T, Sykora H, Kerscher M, Holinski-Feder E. First evidence for digenic inheritance in hereditary colorectal cancer by mutations in the base excision repair genes. *Eur J Cancer* (2011) 47(7):1046–55. doi: 10.1016/j.ejca.2010.11.016

31. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* (2013) 500 (7463):415–21. doi: 10.1038/nature12477

32. McCullough AK, Lloyd RS. Mechanisms underlying aflatoxin-associated mutagenesis - implications in carcinogenesis. *DNA Repair (Amst)* (2019) 77:76–86. doi: 10.1016/j.dnarep.2019.03.004

33. Weren RD, Ligtenberg MJ, Geurts van Kessel A, De Voer RM, Hoogerbrugge N, Kuiper RP. NTHL1 and MUTYH polyposis syndromes: Two sides of the same coin? *J Pathol* (2018) 244(2):135–42. doi: 10.1002/path.5002

34. de Oliveira Garcia FA, de Andrade ES, de Campos Reis Galvão H, da Silva Sábato C, Campacci N, de Paula AE, et al. New insights on familial colorectal cancer type X syndrome. *Sci Rep* (2022) 12(1):2846. doi: 10.1038/s41598-022-06782-8

35. Forsbring M, Vik ES, Dalhus B, Karlsen TH, Bergquist A, Schrumpf E, et al. Catalytically impaired hMYH and NEIL1 mutant proteins identified in patients with primary sclerosing cholangitis and cholangiocarcinoma. *Carcinogenesis* (2009) 30(7):1147–54. doi: 10.1093/carcin/bgp118

36. Prakash A, Carroll BL, Sweasy JB, Wallace SS, Doublié S. Genome and cancer single nucleotide polymorphisms of the human NEIL1 DNA glycosylase: Activity, structure, and the effect of editing. *DNA Repair (Amst)* (2014) 14:17–26. doi: 10.1016/j.dnarep.2013.12.003

37. Roy LM, Jaruga P, Wood TG, McCullough AK, Dizdaroglu M, Lloyd RS. Human polymorphic variants of the NEIL1 DNA glycosylase. *J Biol Chem* (2007) 282(21):15790–8. doi: 10.1074/jbc.M610626200

38. Galick HA, Marsden CG, Kathe S, Dragon JA, Volk L, Nemec AA, et al. The NEIL1 G83D germline DNA glycosylase variant induces genomic instability and cellular transformation. *Oncotarget* (2017) 8(49):85883–95. doi: 10.18632/oncotarget.20716

39. Xue W, Liu Y, Xin N, Miao J, Du J, Wang Y, et al. Nei endonuclease VIII-Like1 (NEIL1) inhibits apoptosis of human colorectal cancer cells. *BioMed Res Int* (2020) 2020:5053975. doi: 10.1155/2020/5053975

40. Shinmura K, Tao H, Goto M, Igarashi H, Taniguchi T, Maekawa M, et al. Inactivating mutations of the human base excision repair gene NEIL1 in gastric cancer. *Carcinogenesis* (2004) 25(12):2311–7. doi: 10.1093/carcin/bgh267

41. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature* (2020) 578 (7793):94–101. doi: 10.1038/s41586-020-1943-3

42. Makasheva KA, Endutkin AV, Zharkov DO. Requirements for DNA bubble structure for efficient cleavage by helix-two-turn-helix DNA glycosylases. *Mutagenesis* (2020) 35(1):119–28. doi: 10.1093/mutage/gez047

43. Møllersen L, Rowe AD, Illuzzi JL, Hildrestrand GA, Gerhold KJ, Tveterås L, et al. Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. *Hum Mol Genet* (2012) 21(22):4939–47. doi: 10.1093/hmg/dds337

44. Calkins MJ, Vartanian V, Owen N, Kirkali G, Jaruga P, Dizdaroglu M, et al. Enhanced sensitivity of Neil1(-/-) mice to chronic UVB exposure. *DNA Repair (Amst)* (2016) 48:43–50. doi: 10.1016/j.dnarep.2016.10.010

45. Dominguez-Valentin M, Sampson JR, Seppälä TT, Ten Broeke SW, Plazzer JP, Nakken S, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective

lynch syndrome database. *Genet Med* (2020) 22(1):15–25. doi: 10.1038/s41436-019-0596-9

46. Goodenberger ML, Thomas BC, Riegert-Johnson D, Boland CR, Plon SE, Clendenning M, et al. PMS2 monoallelic mutation carriers: The known unknown. *Genet Med* (2016) 18(1):13–9. doi: 10.1038/gim.2015.27

47. Bridge G, Rashid S, Martin SA. DNA Mismatch repair and oxidative DNA damage: implications for cancer biology and treatment. *Cancers (Basel)* (2014) 6 (3):1597–614. doi: 10.3390/cancers6031597

48. Peterlongo P, Mitra N, Sanchez de Abajo A, de la Hoya M, Bassi C, Bertario L, et al. Increased frequency of disease-causing MYH mutations in colon cancer families. *Carcinogenesis* (2006) 27(11):2243–9. doi: 10.1093/carcin/bgl093

49. Morak M, Laner A, Bacher U, Keiling C, Holinski-Feder E. MUTYH-associated polyposis - variability of the clinical phenotype in patients with biallelic and monoallelic MUTYH mutations and report on novel mutations. *Clin Genet* (2010) 78(4):353–63. doi: 10.1111/j.1399-0004.2010.01478.x

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

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