frontiers | Research Topics

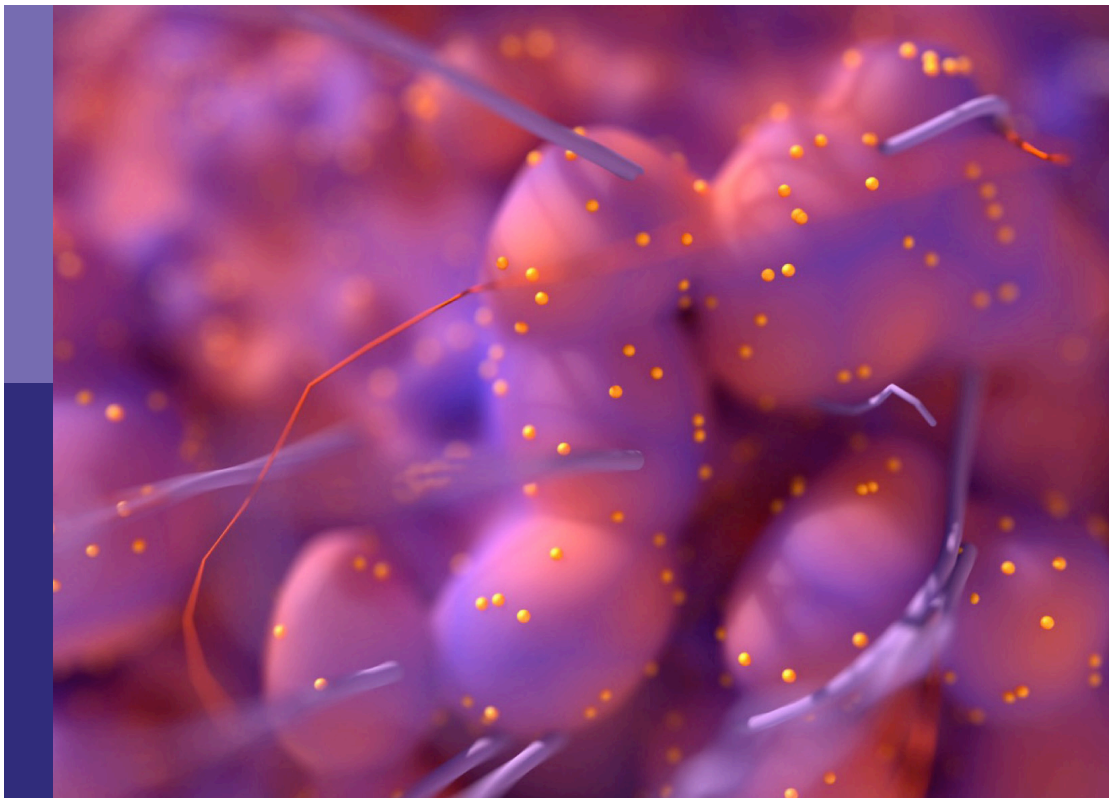
Women in cancer genetics

vol II

2022

Edited by
Giovana Tardin Torrezan and Parvin Mehdipour

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-2208-0
DOI 10.3389/978-2-8325-2208-0

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

Women in cancer genetics vol II: 2022

Topic editors

Giovana Tardin Torrezan — A.C.Camargo Cancer Center, Brazil

Parvin Mehdipour — Tehran University of Medical Sciences, Iran

Citation

Torrezan, G. T., Mehdipour, P., eds. (2023). *Women in cancer genetics vol II: 2022*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2208-0

Table of contents

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 21 April 2022
doi: 10.3389/fonc.2022.880654

Functional Interaction Between the Oncogenic Kinase NEK2 and Sam68 Promotes a Splicing Program Involved in Migration and Invasion in Triple-Negative Breast Cancer

Chiara Naro [1,2], Federica Barbagallo [3†], Cinzia Caggiano [1,2], Monica De Musso [1], Valentina Panzeri [1,2], Silvia Di Agostino [4], Maria Paola Paronetto [5,6] and Claudio Sette [1,2]**

[1] Department of Neuroscience, Section of Human Anatomy, University of the Sacred Hearth, Rome, Italy, [2] Gemelli SCIENCE and TECHNOLOGY PARK (GSTeP)-Organoids Research Core Facility, Fondazione Policlinico Agostino Gemelli IRCCS, Rome, Italy, [3] Department of Experimental Medicine, University of Rome Sapienza, Rome, Italy, [4] Department of Health Sciences, "Magna Graecia" University of Catanzaro, Catanzaro, Italy, [5] Department of Movement, Human and Health Sciences, University of Rome Foro Italico, Rome, Italy, [6] Laboratory of Molecular and Cellular Neurobiology, Fondazione Santa Lucia IRCCS, Rome, Italy

Triple-negative breast cancer (TNBC) represents the most aggressive breast cancer subtype. Poor prognosis in TNBC is partly due to lack of efficacious targeted therapy and high propensity to metastasize. Dysregulation of alternative splicing has recently emerged as a trait of TNBC, suggesting that unveiling the molecular mechanisms underlying its regulation could uncover new druggable cancer vulnerabilities. The oncogenic kinase NEK2 is significantly upregulated in TNBC and contributes to shaping their unique splicing profile. Herein, we found that NEK2 interacts with the RNA binding protein Sam68 in TNBC cells and that NEK2-mediated phosphorylation of Sam68 enhances its splicing activity. Genome-wide transcriptome analyses identified the splicing targets of Sam68 in TNBC cells and revealed a common set of exons that are co-regulated by NEK2. Functional annotation of splicing-regulated genes highlighted cell migration and spreading as biological processes regulated by Sam68. Accordingly, Sam68 depletion reduces TNBC cell migration and invasion, and these effects are potentiated by the concomitant inhibition of NEK2 activity. Our findings indicate that Sam68 and NEK2 functionally cooperate in the regulation of a splicing program that sustains the pro-metastatic features of TNBC cells.

Keywords: triple-negative breast cancer, alternative splicing, transcriptomics, NEK2, SAM68

OPEN ACCESS

Edited by:
Alessandra Montecucco, Italian National Research Council, Italy

Reviewed by:
Stephane Richard, McGill University, Canada
Rebeca Debora Martinez-Contreras, Meritorious Autonomous University of Puebla, Mexico

****Correspondence:***
Claudio Sette
claudio.sette@unicatt.it
Chiara Naro
chiara.naro@unicatt.it

†Present address:
Federica Barbagallo
Università degli Studi di Enna "Kore", Facoltà di Medicina e Chirurgia, Enna, Italy

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

Received: *21 February 2022*
Accepted: *16 March 2022*
Published: *21 April 2022*

Citation:
Naro C, Barbagallo F, Caggiano C, De Musso M, Panzeri V, Di Agostino S, Paronetto MP and Sette C (2022) Functional Interaction Between the Oncogenic Kinase NEK2 and Sam68 Promotes a Splicing Program Involved in Migration and Invasion in Triple-Negative Breast Cancer. Front. Oncol. 12:880654. doi: 10.3389/fonc.2022.880654

INTRODUCTION

Alternative splicing is the molecular process that generates multiple mRNA variants from single eukaryotic genes through variable assortment of their exons (1, 2). This process amplifies the coding potential of genomes and represents a plastic device for the regulation of gene expression. However, errors in alternative splicing regulation are implicated in the pathogenesis of various human

diseases, including cancer (3, 4). Integration of transcriptomic analyses with clinical data have documented that genome-wide alterations in splicing occur in many human cancers, as well as the utility of splice variants as diagnostic or prognostic biomarkers (5–7). In this regard, splicing signatures were shown to distinguish breast cancers (BCs) from normal tissue (7) and to clearly segregate the more aggressive triple-negative BC subtype (TNBC) from the other BCs (8–10).

Oncogenic splicing dysregulation mainly relies on the aberrant expression of specific splicing factors (2, 11). For instance, the oncogenic transcription factor MYC directly induces transcription of several genes encoding for splicing factors (12, 13), including *SRSF1* (14). Importantly, overexpression of the SRSF1 protein was sufficient to induce transformation of mammary epithelial cells and such oncogenic activity was shown to rely, at least in part, on the promotion of splice variants that enhance cell survival, proliferation, and migration (15). Another important layer of splicing regulation relies on the control of the activity of splicing factors through their reversible phosphorylation (16, 17). Splicing-specific kinases, such as the serine arginine protein kinase (SRPK) and the CDC-like kinase (CLK) families, and cell-signaling kinases were reported to regulate both splicing factor expression and activity (16, 17). For instance, phosphorylation by the SRPK1, AURKA, and NEK2 kinases was shown to enhance the splicing activity of SRSF1 in multiple cancer cell types (18–20). Thus, deregulated expression of protein kinases represents another important source of pro-oncogenic splicing alterations.

NEK2 is a mitotic kinase that is frequently upregulated in human cancers, where it contributes to malignancy and drug resistance (21–23). Accordingly, high NEK2 expression was correlated with rapid relapse and poor outcome in multiple cancers (21), including BC (24). In this regard, NEK2 is particularly upregulated in TNBC (25), the BC subtype displaying the poorest prognosis due to high metastatic rate and lack of targeted therapies (26, 27). Although NEK2 oncogenic activity in cancer has been primarily associated to the promotion of genome instability and aneuploidy (22, 28–30), mounting evidence suggests its implication in the pro-tumoral regulation of alternative splicing. Indeed, NEK2 was shown to accumulate in the nucleus of cancer cells (18, 21, 31) and to modulate the activity of splicing factors (18, 32). We recently reported that NEK2 localizes in the nucleus of TNBC cells and exerts a widespread impact on the TNBC-specific transcriptome (25). Part of the splicing changes elicited by NEK2 were mediated by regulation of the expression of the splicing factor RBFOX2, which drives a pro-mesenchymal splicing program in TNBC cells (25). However, a substantial fraction of NEK2-regulated exons and introns were devoid of RBFOX2 binding motifs, suggesting that additional molecular mechanisms contribute to NEK2-mediated splicing regulation in TNBC.

In this study, by searching for additional mediators of NEK2-dependent splicing regulation, we found that NEK2 interacts with and phosphorylates the multifunctional RNA binding protein (RBP) Sam68, thus modulating its splicing activity. Transcriptome analysis of TNBC cells transiently silenced for Sam68 identified the splicing targets of this protein in TNBC cells and revealed a common set of exons that are co-regulated by NEK2, which are enriched in genes related to cell migration. Sam68 depletion in TNBC cells reduces migration and matrix invasion and these effects are enhanced by concomitant inhibition of NEK2 kinase activity. Thus, our study suggests that Sam68 and NEK2 functionally cooperate in the regulation of a splicing program that sustains pro-metastatic features of TNBC cells.

MATERIALS AND METHODS

Cell Culture, Treatment, and Transfection

MDA-MB-231 cells were grown in RPMI 1640 (Lonza), SUM159 cells were grown in DMEM/F12 (Sigma Aldrich), and HEK293T cells were grown in DMEM, all supplemented with 10% FBS, gentamycin, penicillin, and streptomycin. Plasmid transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. For RNA interference, cells were transfected with siRNAs (Sigma-Aldrich) using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instructions and harvested 48 h later for protein and RNA analyses. Sequences of siRNAs are listed in Additional File 1: **Supplementary Table 1**. c-MYC targeting siRNAs were previously described (13).

Plasmid Vectors

The expression vector pcDNA3N_2Myc-NEK2C WT was a generous gift of Prof A. Fry; a catalytically inactive mutant of NEK2C KD ($NEK2_{K37R}$) was created by site-directed mutagenesis of pCDNA3N_2myc-NEK2C WT. Mutagenic oligonucleotides were as follows: forward, 5'AGATATTAGTTTGGAGAGAACTTG ACTATGGC3'; and reverse, with the underlined codon corresponding to residue 37 in wild-type NEK2C. Construction of the pcDNA3N2Myc-NEK2A WT and the kinase-dead inactive mutant NEK2A KD plasmids was previously described (33). The sequence of wild-type and mutant NEK2 were confirmed by sequence analysis. The expression vectors pEGFP NEK2C WT or KD were sub-cloned from pcDNA3N_2Myc into pEGFPc1. pGEX-3X–$NEK2_{273–444}$ encoding the regulatory domain of NEK2 fused to glutathione *S*-transferase (GST) has been described previously (34). CD44 minigene was a kind gift of Prof. Matter.

RNA Extraction, Library Preparation, and RNA-Seq Data Analysis

For RNA-seq analysis, MDA-MB-231 transiently transfected with control (si-CTRL) or SAM68-targeting (si-SAM68) pool of siRNA were harvested 48 h after transfection in triplicate and total RNA was extracted and DNase treated using the RNEasy mini kit (Qiagen) according to the manufacturer's instruction. PolyA plus RNA-seq libraries were constructed according to Illumina's protocol and sequenced using a 100-bp single-end format on an Illumina HiSeq 2000. RNA-Seq data analysis was performed by GenoSplice technology (www.genosplice.com), as previously described (25, 35), using Human FAST DB v2016_1 annotations. Results were considered statistically significant for p-values ≤0.05 and fold-changes ≥1.5.

Bioinformatic Analysis

Analysis of gene expression of transcriptomic data of TNBC patients from the "The Cancer Genome Atlas (TCGA)" database was performed using the UCSC Xena platform (36). Spearman's correlation was used to evaluate association between the expression of *NEK2*, *MYC*, and indicated splicing factors. For gene expression analyses, the patients were divided into two groups according to the first (*NEK2*-low) and fourth (*NEK2*-high) quartile of *NEK2* gene expression. Then, Z-scores of *hnRNPL*, *PTBP1*, and *KHDRBS1* gene expression were calculated in each sample and one-way-ANOVA, with Dunn's multiple comparisons test correction, was performed to evaluate significant differences between the groups (13). Comparison of enriched motif within NEK2-regulated cassette exons with the compendium of RNA-binding motif from (37) were performed using Tomtom Motif comparison tools from the MEME Suite Collection (RRID:SCR_001783) (38, 39). Gene ontology (GO) analysis was performed as previously described, using topGO (RRID:SCR_014798) Bioconductor package, ranking and analyzing ontologies using the elim method (40).

Extraction of RNA, RT-PCR, and Real-Time PCR Analysis

RNA extraction, RT-PCR, and real-time PCR analysis were performed as previously described (25). All primers used are listed in Additional File 1: **Supplementary Table 2**.

Protein Extracts, Co-Immunoprecipitation, and Western Blot Analysis

Total cell extracts were obtained by lysis in 50 mM HEPES, 10 mM $MgCl_2$,100 mM NaCl, 10 mM β-glycerophosphate, 2 mM EGTA, 10% (v/v) glycerol, and 1% (v/v) Triton X-100 for HEK293T and in RIPA buffer for MDA-MB-231, as described (25, 41). Nuclear extracts from MDA-MB-231 cells for co-immunoprecipitation were obtained as described (42). Briefly, cells were lysed in 10 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, phosphatase, and protease inhibitors and incubated on ice for 15 min. Next, following addition of NP-40 to a final concentration of 0.6%, cells were vortexed for 10 s and centrifuged 5 min at 16,000 *g* at 4°C. Supernatant was collected as cytoplasmic extract, and the nuclear pellet was lysed in 20 mM HEPES, pH 7.9, 15 mM $MgCl_2$, 0.42 M KCl, 0.2 mM EDTA, and 25% (v/v) glycerol and agitated for 30 min at 4°C. After centrifugation for 10 min at 16,000 *g* at 4°C, supernatant was collected as nuclear fraction. Lysates were diluted with 20 mM HEPES, pH 7.9, 15 mM $MgCl_2$, 0.2 mM EDTA, and 25% (v/v) glycerol and incubated overnight at 4°C with NEK2 antibody (Santa Cruz, RRID:AB_1126558) or control mouse IgG. Protein-G magnetic beads (Dynabeads, Invitrogen) were added and the sample was incubated at 4°C for 3 h. Beads were washed five times with 20 mM HEPES, pH 7.9, 15 mM $MgCl_2$, 120 mM KCl, 0.2 mM EDTA, and 25% (v/v) glycerol, denatured with SDS-sample buffer, and analyzed by SDS-PAGE. Western blot was performed as previously described (18, 25, 35). Nuclear extracts and immunoprecipitation from HEK293T cells were obtained as previously described (43), using either anti-FLAG antibody (Sigma Aldrich, RRID:AB_259529) or control mouse IgG. Western blot analysis was carried out using the following primary antibodies: anti-NEK2 (Santa Cruz, RRID: AB_1126558), anti-ACTIN (Santa Cruz, RRID:AB_2714189), anti-MYC epitope (Santa Cruz, RRID:AB_2857941), anti-MYC (Cell Signaling, RRID:AB_2151827), anti-GAPDH (1:1000 RRID: AB_627679), anti-HSP90 (Santa Cruz, RRID:AB_675659), anti-hnRNPL (Sigma Aldrich, RRID: AB_261966), anti-FLAG (Sigma Aldrich, RRID:AB_259529), rabbit anti-SAM68 (Santa Cruz, RRID:AB_631869), anti-GFP (Santa Cruz, RRID:AB_627695), anti-ERK2 (Santa Cruz, RRID:AB_2141292), and goat anti-PTBP1 (Santa Cruz, RRID:AB_2253470).

Immunokinase Assays

Anti-MYC antibody (1 μg) (Santa Cruz, RRID:AB_2857941) was incubated for 1 h, with a mixture of protein A/G-Sepharose beads (Sigma-Aldrich) in PBS/0.05% BSA, under constant shaking at 4°C. At the end of the incubation, the beads were washed twice with PBS/0.05% BSA, twice with lysis buffer, and then incubated for 90 min at 4°C with the HEK293T cell extracts (0.5 mg of protein) under constant shaking. Sepharose bead-bound immunocomplexes were rinsed three times with lysis buffer and washed twice with NEK2-kinase buffer (50 mM HEPES, pH 7.5, 5 mM glycerophosphate, 5 mM $MnCl_2$, 5 mM NaF, 0.1 mM sodium orthovanadate, 1 mM DTT, and protease inhibitors). Kinase reactions were carried out in 50 μl for 20 min at 30°C in kinase buffer supplemented with 10 μM [^{32}P]-ATP (0.2 μCi/μl), 4 μM ATP, 1 μg of cAMP-dependent protein kinase inhibitor, and the appropriate substrate (GST-Sam68 N-term or C-term). Reactions were stopped by adding SDS-sample buffer and analyzed by SDS-PAGE and autoradiography.

Wound-Healing and Cell-Invasion Assays

Control or SAM68 silenced MDA-MB-231 were seeded at 100% of confluence into ibidi Culture inserts to create a cell-free gap on the dish. Following two washes with PBS and addition of 1% FBS supplemented medium, inserts were removed and the plate was photographed immediately and every hour for 12 h. Area quantification of the gap was performed with ImageJ software using the MRI Wound Healing tool. For cell invasion assay, MDA-MB-231 cells were seeded into the IncuCyte Clearview 96-well inserts (Sartorius; 1,000 cells/well). Insert membrane had been pre-coated on both sides with 50 μg/ml Matrigel (Corning), diluted in RPMI 1640. Lower chambers were filled with 200 μl of either chemotaxis assay medium (RPMI 1640, supplemented with 10% FBS) or negative control medium (RPMI 1640, without FBS). Images were acquired with IncuCyte SX5 Live-content imaging system every hour for 24 h at 10× magnification. Migrated cells were quantified using the IncuCyte Chemiotaxis migration software (phase-contrast; top/bottom), starting 2 h after initial seeding to allow settlement of cells. In both assays, NEK2-chemical inhibition was achieved by treatment with 3 μM JH295 (44).

Quantification and Statistical Analysis

Statistical analyses for qPCR, densitometric analysis of PCR, and migration and invasion assays were performed in GraphPad Prism (RRID:SCR_002798) according to the statistical tests

described in the figure legends. Number of replicates independently analyzed is indicated by the "*n*" in each figure legend. Results were considered significant if *p*-value ≤ 0.05.

RESULTS

NEK2 Interacts With Splicing Factors in TNBC Cells

We recently found that NEK2 exerts widespread modulation of the alternative splicing program of MDA-MB-231 (GSE140803), a cell line representative of the TNBC subtype (25). A substantial fraction of the NEK2-regulated splicing events were dependent on the ability of the kinase to promote the expression of RBFOX2 (25), a splicing factor involved in the regulation of the epithelial-to-mesenchymal transition (EMT) process (25). However, other exons regulated by NEK2 lacked binding sites for RBFOX2 and were likely regulated by other splicing factors in TNBC cells. Moreover, NEK2 shows an increased localization in the nucleoplasm and chromatin-bound fraction of TNBC cells, suggesting that it might also physically interact with splicing factors and regulate their activity. To test this hypothesis, we searched for splicing factors that can bind the 5 sequence motifs enriched in the NEK2-regulated cassette exons (25). Computational analyses using the Tomtom Motif comparison tool (38, 39) identified 14 splicing factors that might bind to these sequence motifs (**Figure 1A**), including the already characterized RBFOX2 (25). Next, to evaluate which of these factors could functionally interact with NEK2 in TNBC, we assessed whether they are co-expressed with *NEK2* in primary tumors. Query of transcriptomics data from TNBC tumors deposited in The Cancer Genome Atlas (TCGA) database (45) revealed that expression of *hnRNPL*, *PTBP1* (also known as *hnRNP I*), and *KHDRBS1* (also known as *Sam68*) exhibit the highest and most significant positive correlation with *NEK2* expression (**Figure 1B**). Expression of *A1CF*, *KHDRSB3*, and *hnRNPLL* was also positively correlated

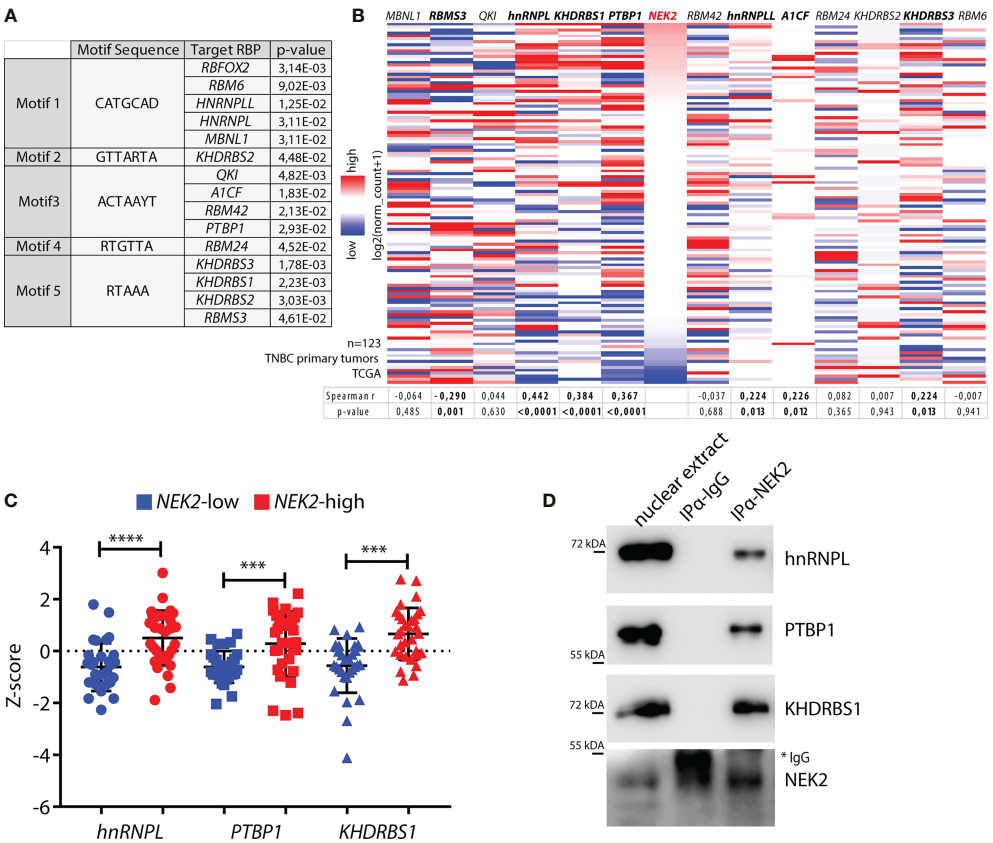


FIGURE 1 | NEK2 is co-expressed and interacts with select splicing factors in TNBC. **(A)** Table showing motifs enriched in sequences of NEK2-regulated cassette exons in MDA-MB-231 cells (GSE140803) and their putative cognate RNA-binding proteins (RBP), identified by the Tomtom motif comparison tool. Only significant results retrieved by the tool are shown (*p*-value ≤ 0.05). **(B)** Heatmap showing expression levels of NEK2 and indicated RBP in primary triple-negative breast cancer (TNBC), according to analysis of transcriptomic data of TCGA project using the UCSC Xena platform. Spearman correlation factors and *p*-value between the expression levels of *NEK2* and every RBP are indicated in the table below the heatmap. **(C)** Expression profile for indicated RBP according to TCGA transcriptomic data of TNBC patients, classified according to Z-score normalization in *NEK2*-low (blue points) and *NEK2*-high (red points) expressing groups. Mean and ± SD are shown in the dot plot. Statistical significance was calculated by one-way ANOVA, with Dunn's multiple comparisons test, ***$p < 0.001$, ****$p < 0.0001$. **(D)** Western blot analysis for indicated RBPs for co-immunoprecipitation assay of NEK2 and control IgG in nuclear extracts from MDA-MB-231 cells. * marks IgG used for immunoprecipitation.

with that of *NEK2*, albeit to a lesser extent, whereas *RBMS3* expression was negatively correlated (**Figure 1B**). On the other hand, expression of *MBNL1*, *QKI*, *RBM24*, *KHDRSB2*, *RBM42*, and *RBM6* was not correlated with that of *NEK2* (**Figure 1B**). Furthermore, Z-score classification of patients for low and high expression of *NEK2* confirmed that *hnRNPL*, *PTBP1*, and *KHDRBS1* levels are significantly higher in the *NEK2*-high group compared to the *NEK2*-low group (**Figure 1C**). Importantly, co-immunoprecipitation experiments using nuclear extracts from MDA-MB-231 revealed that NEK2 physically interacts with hnRNPL, PTBP1, and KHDRBS1 proteins (**Figure 1D**), whereas no interaction was detected for an uncorrelated factor like QKI (**Supplementary Figure 1A**). These findings indicate that the interaction of NEK2 with specific splicing factors could be functionally relevant to modulate the splicing signature of TNBC cells.

Oncogenic Transcription Factor MYC Drives NEK2 Expression in TNBC

Increased nuclear localization of NEK2 in TNBC is driven by its higher expression levels compared to other BC subtypes (25). Since we identified splicing factors that are co-expressed and interact with NEK2 in this tumor subtype, we asked if a common transcription factor could promote their expression. In particular, we focused our attention on the proto-oncogenic transcription factor c-MYC, which is overexpressed in TNBC compared to other BC subtypes (**Supplementary Figure 1B**) (46) and was shown to drive transcription of both *NEK2* (32) and its putative cofactors *PTBP1* (12) and *KHDRBS1* (13) in other tumoral context. Analysis of expression data in the TCGA database revealed a significant upregulation of *hnRNPL*, *PTBP1*, and *KHDRBS1* expression in TNBC compared to other BC subtypes (**Figure 2A**) as previously reported for *NEK2* (25). In addition, we observed that *MYC* expression was positively correlated with the expression of *NEK2* (**Figure 2B**) and of its putative co-factors *hnRNPL*, *PTBP1*, and *KHDRBS1* (**Figure 2C**). By contrast, no significant correlation was observed between *MYC* and *QKI* expression (**Supplementary Figure 1C**), whose expression is not correlated with *NEK2* (**Figure 1B**). These observations suggest that c-MYC could coordinate the expression of NEK2 and its interacting splicing factors in TNBC. To test this hypothesis, we asked whether c-MYC silencing affects the expression of these proteins in TNBC cells. Western blot analyses of extracts from MDA-MB-231 and SUM159 transiently transfected with two different c-MYC siRNAs revealed that c-MYC depletion reduced the expression of NEK2, hnRNPL, PTBP1, and KHDRBS1 in both TNBC cell lines (**Figure 2D**). Collectively, these observations suggest that c-MYC overexpression sustains the concomitant expression of NEK2 and select splicing factors in TNBC cells, thereby favoring their interaction.

Sam68 Is a Direct Substrate of NEK2 Kinase

To functionally test the interaction between NEK2 and splicing factors in TNBC, we focused on KHDRBS1, hereafter named Sam68 (**Figure 3A**), because its activity is extensively modulated by phosphorylation (43, 47–50). Moreover, Sam68 was more dependent on MYC expression than hnRNPL and PTBP1 in both TNBC cell lines tested (**Figure 2D**). First, we confirmed the physical interaction between the proteins by co-immunoprecipitation of transiently transfected FLAG-Sam68 and GFP-NEK2 in HEK293T cells (**Supplementary Figure 2**). Next, we performed *in vitro* kinase assays in the presence of labeled ATP ([γ-32P]ATP) using purified recombinant NEK2 and GST-Sam68. NEK2 readily phosphorylated GST-Sam68, to a similar extent of its known substrate GST-SRSF1 (18), whereas GST alone was not phosphorylated under the assay conditions (**Figure 3B**). Sam68 comprises an hnRNP K homology (KH) RNA binding motif flanked by the QUA1 and QUA2 motifs, which form the GRP33/Sam68/GLD1 (GSG) domain required for dimerization and high affinity RNA binding, and regulatory regions at the N and C terminus that contain sites for protein–protein interactions and post-translational modifications (**Figure 3A**) (47, 48, 51). NEK2 phosphorylates with high efficiency the N-terminal and C-terminal regulatory regions of Sam68, whereas the GSG domain was barely phosphorylated (**Figure 3C**). Moreover, *in vitro* kinase assays using wild type (WT) or kinase-dead (KD) NEK2 immunoprecipitated from transfected HEK293T cells confirmed that the enzymatic activity of NEK2 was directly responsible for phosphorylation of the regulatory regions of Sam68 (**Figures 3D, E**).

Next, we asked if NEK2 phosphorylates Sam68 also in live cells. To this end, HEK293T cells were transfected with plasmids encoding MYC-Sam68 and either WT or KD versions of MYC-tagged NEK2A and GFP-tagged NEK2C, a splice variant of the kinase that is predominantly localized in the nucleus like Sam68 (52). Upon treatment with the protein phosphatase 1 (PP1) and 2A (PP2A) inhibitor Okadaic Acid (OA) to elicit NEK2 activation (33), NEK2 induced a shift in the electrophoretic mobility of Sam68 (**Figure 3F**), which is a hallmark of its phosphorylation in serine/threonine residues (49). Notably, the slower migrating form of Sam68 was observed only in cells co-transfected with WT NEK2, but not with the catalytically inactive KD mutant. Furthermore, NEK2C displayed higher ability to induce Sam68 phosphorylation, further supporting a functional interaction in the nucleus between the proteins. These results identify Sam68 as a novel substrate of NEK2.

NEK2-Mediated Phosphorylation Modulates Sam68 Splicing Activity

Serine/threonine phosphorylation by the mitogen activated protein kinases (MAPKs) was shown to regulate the splicing activity of Sam68 (16, 49, 50, 53). To investigate whether NEK2-dependent phosphorylation also affects Sam68 activity, we employed a reporter minigene that recapitulates the splicing regulation of the *CD44* v5 exon (pET-V5 minigene) (**Figure 4A**), which is a target of Sam68 and is sensitive to its activation by serine/threonine phosphorylation (49, 54). HEK293T cells were co-transfected with plasmids encoding the pET-V5 minigene, MYC-Sam68, and GFP-NEK2C. As expected, sub-optimal

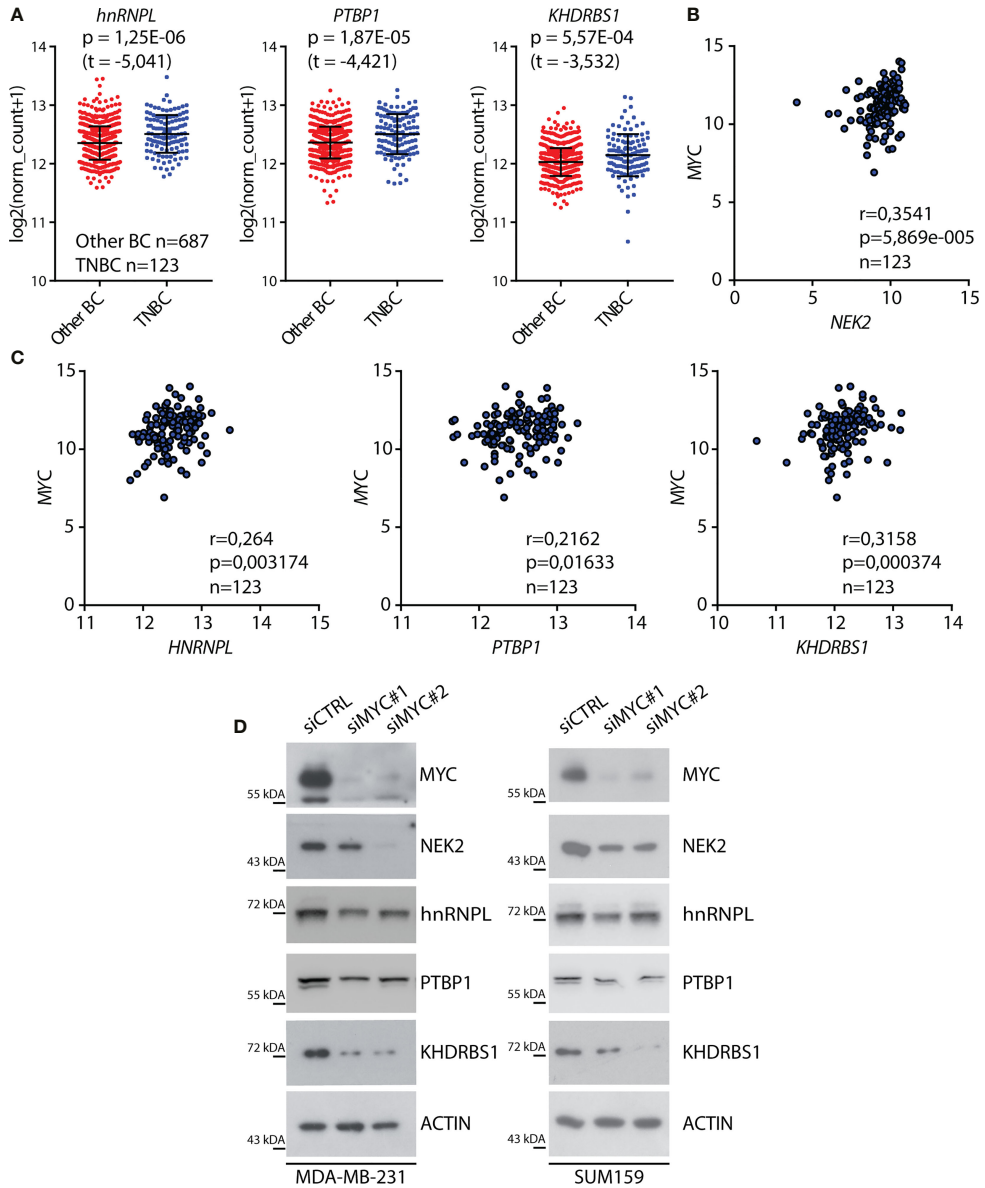


FIGURE 2 | C-MYC regulates the expression of NEK2 in TNBC cells. **(A)** Dot-blot showing expression levels of indicated RNA-binding proteins (RBP) in primary triple-negative breast cancer (TNBC) and other breast cancer (Other BC) subtypes, according to analysis of transcriptomic data of TCGA project using the UCSC Xena platform. Mean and ± SD are shown in the dot plot. Statistical significance was calculated by Welch's *t*-test. **(B, C)** Scatter plots of RNA expression levels of *MYC* and *NEK2* **(B)** or *MYC* and *hnRNPL, PTBP1*, and *KHDRBS1* **(C)** in primary TNBC according to TCGA data. Spearman's correlation coefficient (*r*) and associated *p*-value are shown. **(D)** Representative Western blot analysis for MYC, NEK2, hnRNPL, PTBP1, and KHDRBS1 expression levels in MDA-MB-231 (left panel) and SUM159 (right panel) cells, transiently transfected with indicated siRNAs. ACTIN was evaluated as loading control.

amounts of MYC-Sam68 promoted the inclusion of CD44 variable exon v5 (**Figure 4B**; lane 2). Co-expression of NEK2C significantly enhanced this effect, leading to an almost doubled inclusion of the v5 exon with respect to cells transfected with Sam68 alone (**Figure 4B**; lane 3). Furthermore, u**preg**ulation of NEK2C alone, but not of its cognate KD mutant, was sufficient to promote exon v5 inclusion (**Figure 4C**), suggesting that it might affect the splicing activity of the endogenous Sam68 protein. These results indicate that NEK2-dependent phosphorylation of Sam68 modulates its splicing activity and that their physical interaction might be functionally relevant in TNBC cells.

Sam68 Modulates TNBC Cell Transcriptome

Sam68 is upregulated in breast tumors compared to normal tissue and promotes BC cell proliferation (55). However, although the oncogenic function of Sam68 has been often related to its splicing activity (47, 50, 53), genome-wide

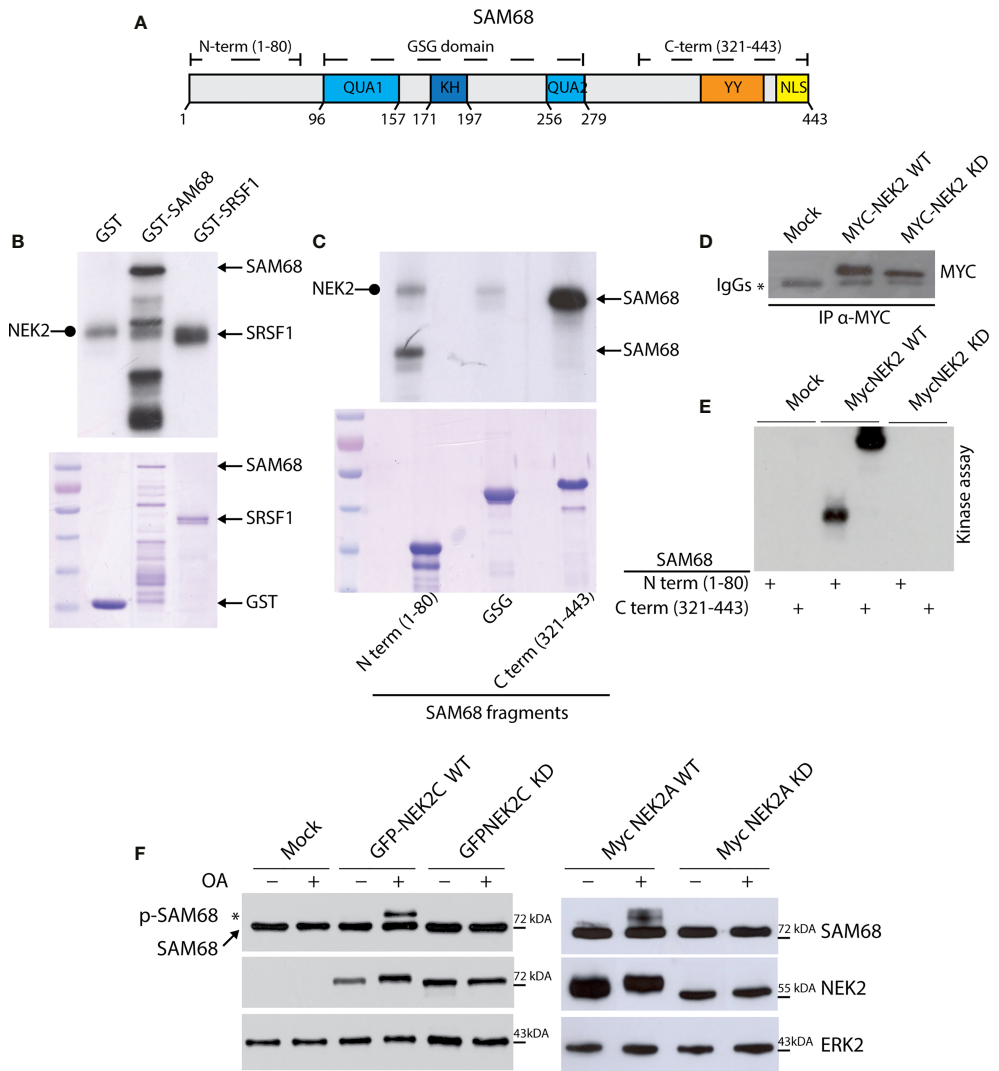


FIGURE 3 | NEK2 phosphorylates SAM68 *in vitro*. **(A)** Schematic representation of protein domain in human full-length SAM68 protein. **(B)** Representative autoradiography for NEK2 kinase assay, performed by incubating an active purified-NEK2 protein with recombinant GST and a full-length SAM68 and SRSF1 as substrates. Coomassie staining was performed as loading control. Rounded tip arrows indicate auto-phosphorylated NEK2. **(C)** Representative autoradiography and Coomassie staining for a kinase assay performed by incubating an active purified-NEK2 recombinant GST N-terminal, GSG domain or C-terminal of SAM68 as substrates. **(D, E)** Representative Western blot analysis **(D)** and autoradiography **(E)** for immunokinase assay, performed by incubating immunoprecipitated wild-type NEK2 (WT) or kinase-dead NEK2 (KD) with recombinant N-terminal or C-terminal GST-Sam68 as substrate. **(F)** Western blot analysis for MYC-Sam68 protein in total extracts from HEK293T cells transfected with expression vectors for GFP-tagged NEK2C wild-type (WT) or kinase-dead (KD), or with MYC-tagged NEK2A variant WT or KD. Tag-specific antibodies were used for recombinant NEK2 detection. ERK2 was evaluated as loading control. Activation of NEK2 was obtained by treating cells with 0.1 μM OA for the last 3 h before collection. * marks the molecular weight shift in SAM68 protein elicited by OA-mediated activation of WT NEK2C (left panel) and WT NEK2A (right panel).

characterization of its splicing targets in BCs or other cancer types is still lacking. To elucidate the splicing signature regulated by Sam68 in TNBC cells, we carried out RNA-sequencing (RNA-seq) analyses of MDA-MB-231 cells that were transiently depleted of Sam68 (**Figure 5A**). Bioinformatics analyses using the reference FAST-DB database (25, 35, 41), revealed a large modulation of the TNBC cell transcriptome by Sam68, with 443 genes regulated at splicing level and 474 at gene expression level upon its depletion (**Figure 5B**; **Supplementary Figure 3A**; Additional File 2: **Supplementary Tables 1, 2**). More than half of the 597 regulated exons (54,7%) are u**preg**ulated in Sam68-depleted cells, suggesting that Sam68 preferentially functions as a splicing repressor in TNBC cells. Classification of the regulated splicing events revealed that exon cassettes (18.4%) and alternative terminal exons (16.1%) are the most regulated patterns (**Figure 5C**). Importantly, RT-PCR analysis of 16 of these splicing events using an independent set of control and Sam68-depleted MDA-MB-231 cells confirmed the RNA-seq

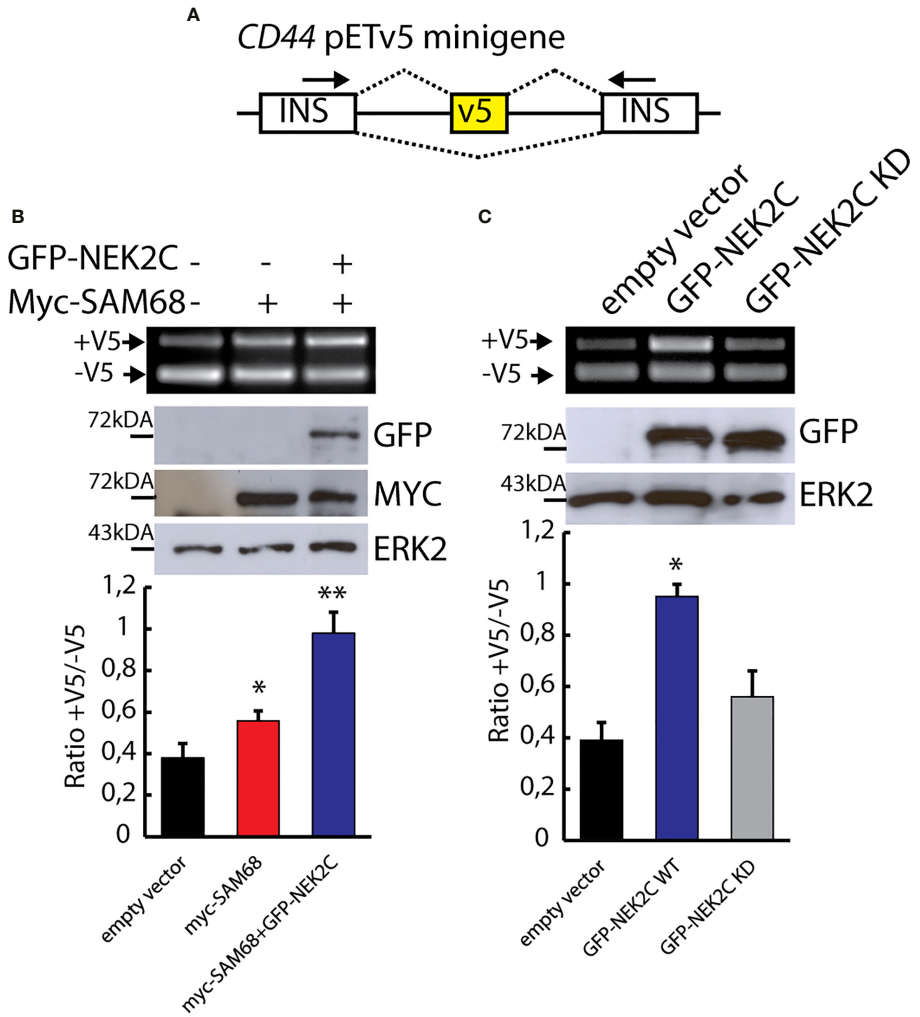


FIGURE 4 | NEK2 phosphorylates Sam68 *in vivo* and modulates its splicing activity. **(A)** Schematic representation of the *CD44* pETV5 minigene. Alternative exon v5 of the *CD44* gene was cloned between two constitutive cassette exons insulin exons 2 and 3. **(B, C)** Representative PCR and Western blot analysis for HEK293T cells transfected with the *CD44* pETV5 minigene and expression vectors for MYC-tagged SAM68 and GFP-tagged NEK2C wild-type (WT) or kinase-dead (KD). Western blot analysis for ERK2 was used as loading control. Densitometric analyses for all experiment were performed and ratio between CD44 (+V5) and CD44 (-V5) is represented by histogram bars (mean ± SD, $n = 3$; *t*-test $^{*}p < 0.05$, $^{**}p < 0.01$).

results (**Figure 5D**; **Supplementary Figure 3B**), thus validating the reliability of the bioinformatics analyses. These results show that Sam68 significantly contributes to the splicing signature of TNBC cells.

Sam68 and NEK2 Co-Regulate Alternative Splicing Events in TNBC Cells

Next, we asked whether NEK2 modulates the splicing activity of the endogenous Sam68 in TNBC cells. To this end, we compared the splicing signatures of Sam68-silenced (**Figures 5B, C**) and NEK2-silenced MDA-MB-231 cells (25). We found a significant overlap between the two datasets, with 95 alternative splicing events that are commonly regulated by Sam68 and NEK2 depletion (**Figure 6A**). Annotation of these events revealed that most of them were modulated in the same direction by silencing of either Sam68 or NEK2 (**Figure 6B**). Nearly half of the splice variants commonly regulated by Sam68 and NEK2 were novel transcripts originating from either unannotated splicing events or selection of an alternative transcription start site, while alternative last exon and exon cassette were the predominant splicing patterns among the remaining events (**Figure 6C**). RT-PCR analysis using RNA from an independent set of samples confirmed that depletion of Sam68 or NEK2 regulated a common pattern of splicing in three of these genes (*ASPH*, *MAPK9*, and *TBC1D23*) in MDA-MB-231 cells (**Figures 6D, E**). Furthermore, RT-PCR analyses revealed that additional Sam68 target exons, like those in *ALCAM*, *CD44*, *GULP1*, and *UGGT2* genes, were also sensitive to NEK2 depletion in MDA-MB-231 (**Figure 6F**), even though they were not highlighted by the bioinformatics analysis (25). These results indicate that Sam68 and NEK2 share common

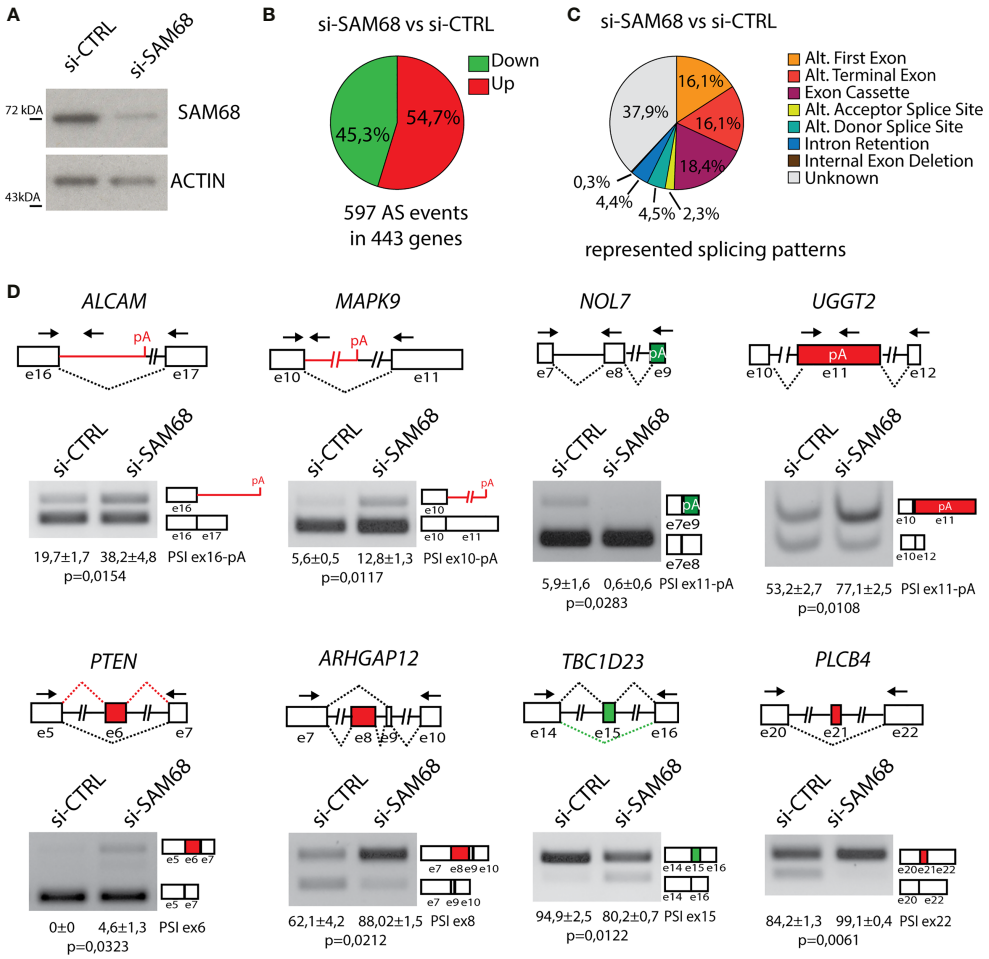


FIGURE 5 | Sam68 regulates alternative splicing in TNBC cells. **(A)** Representative Western blot analysis assessing SAM68 silencing efficiency expression in MDA-MB-231 cells transiently transfected with indicated pool of siRNAs. ACTIN was evaluated as loading control. **(B)** Pie chart showing percentage of upregulated (red) and downregulated (green) exons in the si-SAM68 vs. si-CTRL comparison. **(C)** Pie chart showing percentages of indicated different splicing pattern among regulated splicing events in the si-SAM68 vs. si-CTRL comparison. **(D)** Representative PCR analysis for indicated alternative splicing events in si-SAM68 vs. si-CTRL MDA-MB-231 cells. Schematic representation for each event analyzed is depicted below relative agarose gels. Green and red boxes indicate down- and upregulated exons in si-SAM68 vs. si-CTRL cells. Percentage of splicing inclusion (PSI) of indicated exons was evaluated by densitometric analysis, and results are shown below agarose gels (mean ± SD, $n = 3$, t-test).

splicing targets in TNBC cells and further suggest their functional interaction in splicing regulation in this tumor type.

Sam68 and NEK2 Cooperate in the Regulation of TNBC Cell Migration and Matrix Invasion

GO analysis of the Sam68 splicing-regulated genes highlighted a significant enrichment for terms related to biological processes involved in cell adhesion and migration (**Figure 7A**). Moreover, genes related to the wound-healing process were enriched among the common targets of Sam68 and NEK2 (Supplementary **Figure 4**). These process are frequently deregulated in TNBC and contribute to their aggressive and metastatic phenotype (27, 56). Thus, we asked if Sam68 ablation could affect these pro-metastatic functions. Wound-healing and matrigel-invasion assays revealed that Sam68 depletion caused a significant impairment of the migratory and invasive properties of MDA-MB-231 cells (**Figures 7B–D**). Notably, we also found that the effects elicited by Sam68 knockdown were worsened by concomitant treatment of MDA-MB-231 cells with JH295, an irreversible inhibitor of NEK2-kinase activity (**Figures 7B–D**) (44). Collectively, these observations indicate that the functional interaction with NEK2 potentiates the splicing activity of Sam68 and enhances the motility and invasive properties of TNBC cells.

DISCUSSION

Alternative splicing dysregulation is a common trait of human cancers, which affects multiple cellular processes in the course of

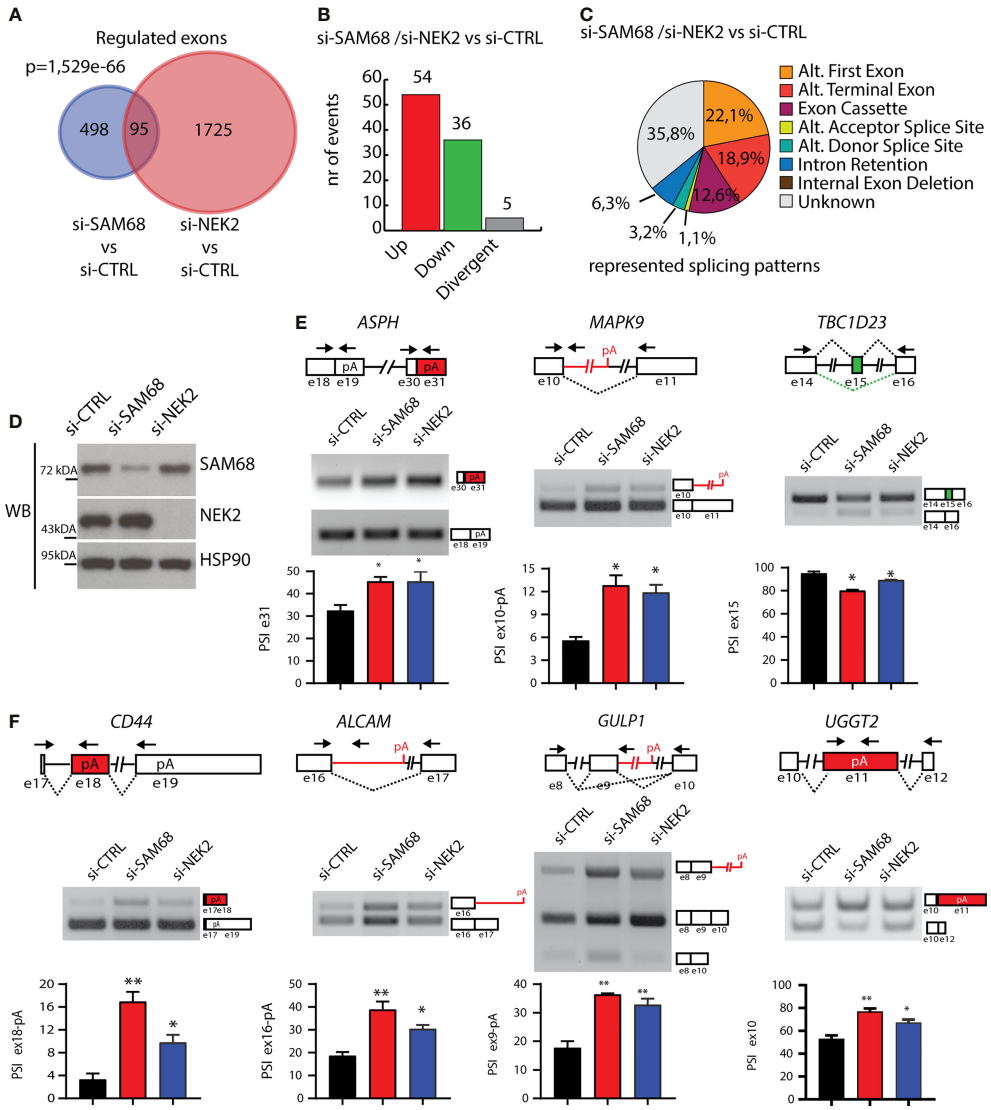


FIGURE 6 | Sam68 and NEK2 co-regulates AS in TNBC cells. **(A)** Venn diagram showing the overlap between regulated alternative exons in MDA-MB-231 cell silenced for either SAM68 (this study) or NEK2 (GSE140803). **(B)** Bar graph showing the number of splicing events either divergently (gray bar) or commonly upregulated (red bar) or downregulated (green bar) in si-SAM68 and si-NEK2 MDA-MB-231 cells compared to control. **(C)** Pie chart showing percentages of indicated splicing patterns among the common splicing events regulated in the si-SAM68/si-NEK2 vs. si-CTRL comparison. **(D)** Representative Western blot analysis assessing SAM68 and NEK2 silencing efficiency in MDA-MB-231 cells transiently transfected with indicated pool of siRNAs. HSP90 was evaluated as loading control. **(E, F)** Representative PCR analysis for indicated alternative splicing events in si-CTRL, si-SAM68, and si-NEK2 MDA-MB-231 cells. Schematic representation for each event analyzed is depicted besides relative agarose gels. Green and red boxes indicate commonly down- and upregulated exons in si-SAM68/si-NEK2 vs. si-CTRL cells. Bar graphs below each agarose gel represent percentage of splicing inclusion (PSI), evaluated by densitometric analysis (mean ± SD, $n = 3$, one-way ANOVA, $^{*}p < 0.05$, $^{**}p < 0.01$).

tumorigenesis (2, 3, 17). Thus, characterization of the molecular mechanisms underlying aberrant splicing offers the opportunity to identify new targets for cancer therapy. This issue is particularly interesting for TNBC, as these tumors currently lack targeted and efficacious therapies, but features a specific splicing signature (8, 10, 25). In this regard, targeting either the expression of specific splicing factors or inhibiting the spliceosome activity was shown to selectively halt TNBC cell proliferation (57). We have previously reported that the mitotic kinase NEK2 is upregulated in TNBC with respect to other BC types and promotes a specific pro-mesenchymal splicing program that confers metastatic features to TNBC (25). Herein, we found that NEK2 interacts with select splicing factors in TNBC cells and, as indicated by its functional interaction with Sam68, could enhance their splicing activity and oncogenic functions.

NEK2 is highly expressed in primary TNBC along with Sam68, hnRNPL, and PTPBP1, whose cognate binding motifs are enriched in NEK2-sensitive cassette exons and physically

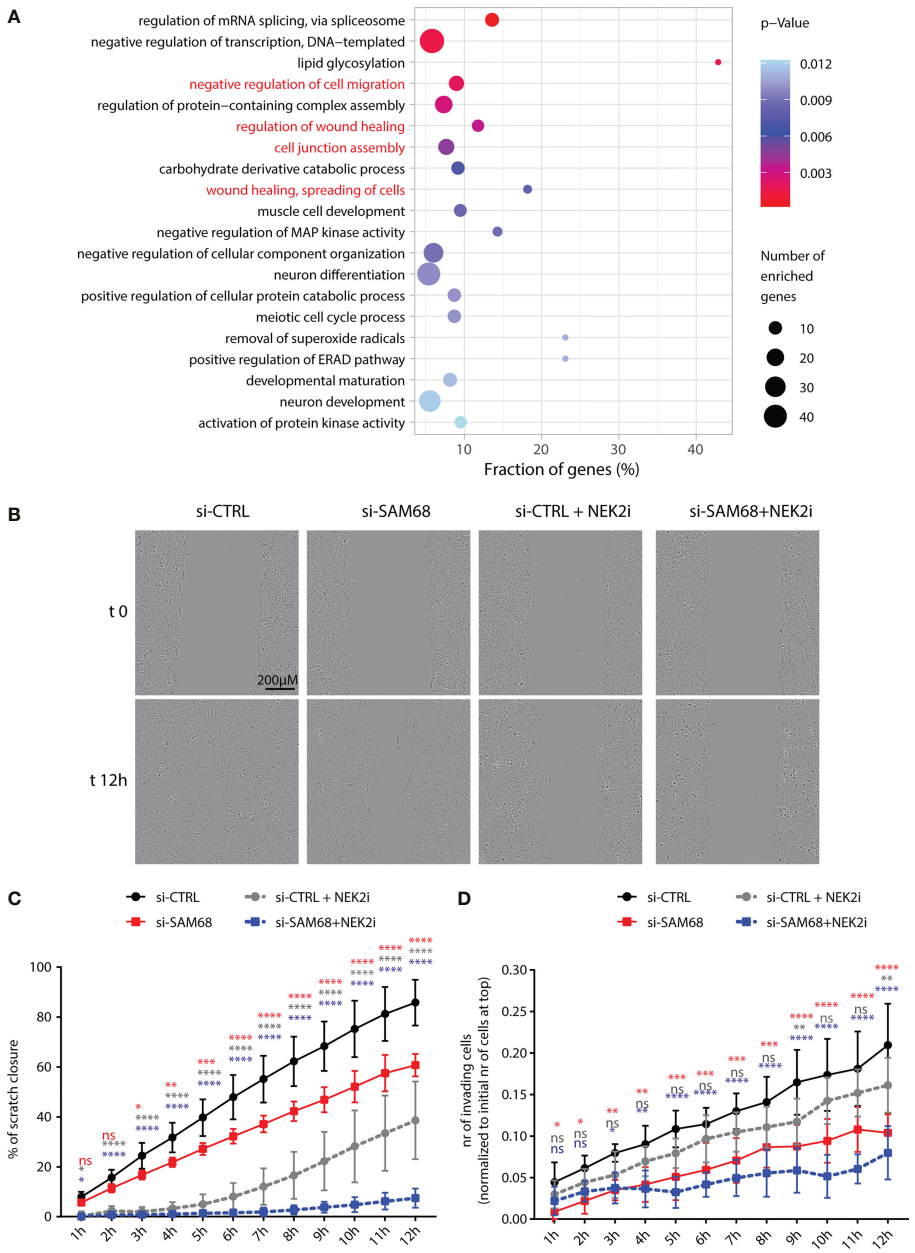


FIGURE 7 | Sam68 and NEK2 co-regulate cell migration in TNBC cells. **(A)** Gene ontology analysis of biological process of AS regulated genes in the comparison between control and Sam68 silenced MDA-MB-231 cells (p-value ≤ 0.05). **(B)** Representative micrograph images, at the initial time point (t0) and 12 h later (t12) of the wound-healing assay performed on control (si-CTRL) or SAM68 silenced (si-SAM68) MDA-MB-231 cells, treated or not with the NEK2 inhibitor (NEK2i) JH295 [3 µM]. **(C)** Line graph showing the percentage of wound closure of si-CTRL and si-SAM68 MDA-MB-231 cells, treated or not with NEK2i (mean ± SD, n = 6, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns = not significant, two-way ANOVA, colors indicate the growth condition to whom si-CTRL cells were compared in the statistical analysis). **(D)** Line graph showing the number of si-CTRL and si-SAM68 MDA-MB-231 cells, treated or not with NEK2i, invading Matrigel-coated transwell of the IncuCyte Clearview 96-well insert system. Number of invading cells on the bottom side of the insert at every hour was normalized to the initial number of cells on the top of the insert at initial seeding (mean ± SD, n = 6, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns= not significant, two-way ANOVA, colors indicate the growth condition to whom si-CTRL cells were compared in the statistical analysis).

interact with this kinase. Moreover, knockdown of Sam68 (this study) and hnRNPL (25) partially recapitulated the splicing changes observed in TNBC cells depleted of NEK2, suggesting that this kinase modulates splicing through functional interaction with splicing factors in TNBC cells. We also found that expression of NEK2 and its interacting splicing factors in primary TNBC correlates with that of MYC, suggesting that this transcription factor coordinates a splicing network that

contributes to the TNBC-specific splicing signature. In support of this hypothesis, MYC depletion in representative TNBC cell lines caused reduced expression of NEK2, Sam68, hnRNPL, and PTBP1 proteins. MYC is a powerful oncogene and is highly expressed in TNBC compared to other BCs (46). Moreover, MYC upregulation was shown to impose a transcriptional stress to cancer cells that increases their dependency on the proper functionality of the splicing machinery (58). Thus, it is tempting to speculate that coordination of the expression of NEK2 and its interacting splicing factors represents a pro-survival mechanism that is selected in MYC-driven TNBC to cope with such transcriptional/splicing stress. In this view, targeting NEK2 expression and/or activity could represent a therapeutic vulnerability for MYC-driven TNBC, as previously shown for inhibition of the spliceosome (58). Remarkably, MYC regulates transcription of other splicing factors (SRSF1 and hnRNPA1) that interact with and are regulated by NEK2 in other tumoral contexts (12, 14, 18, 32). Thus, NEK2 inhibition could represent an exploitable vulnerability also for other types of MYC-driven tumors.

Sam68 is a multifunctional RBP, whose splicing activity exerts a pivotal role for the proper differentiation of neuronal and germ cells (35, 59, 60). Notably, although several studies have shown the oncogenic activity of Sam68 in different human cancers (47, 48), a global analysis of the regulation exerted by this protein on the human transcriptome was still lacking. Herein, genome-wide RNA-seq analysis identified hundreds of splicing events modulated by Sam68 depletion in TNBC cells. Similarly to previous observations in *Sam68* knockout mice (35, 59), exon cassettes and alternative terminal exons were the most affected splicing patterns in MDA-MB-231 cells. Of note, functional annotation of the splicing-regulated genes revealed enrichment for terms relative to neuronal and muscular development, as well as to meiosis, all biological processes that are impaired in *Sam68* knockout mice (35, 59–62). These observations are suggestive of an evolutionary conserved splicing program regulated by Sam68, which is hijacked by cancer cells to sustain oncogenic transformation.

Serine/threonine phosphorylation is one of the major post-translational modifications shown to promote the pro-oncogenic splicing activity of Sam68 (49, 50, 53, 63). Most of these studies identified the MAPK/ERK pathway as responsible for Sam68 phosphorylation and activation (49, 50, 53, 63). By identifying Sam68 as a binding partner and direct substrate of NEK2, we provide evidence for an additional cellular pathway modulating Sam68 phosphorylation and oncogenic splicing activity. Interestingly, activation of the MAPK pathway was shown to promote NEK2 activity in male germ cells (34), suggesting the possible synergy between these kinases in the regulation of Sam68 activity. Given the ubiquitous expression of Sam68 and NEK2 and their frequent upregulation in different cancer types (48, 64), their interaction is likely functional also in other tumors. Thus, our study reveals a new regulatory mechanism of Sam68 function, which adds to the various post-translational modifications, such as tyrosine phosphorylation (43) and acetylation (65), and interactions with regulatory partners, such as the transcriptional cofactors SND1 (54) and FBI-1 (66), that modulate its splicing activity in cancer cells, including TNBC cells.

Our studies have also identified a large number of splice variants regulated by Sam68 and NEK2 that are possibly implicated in the regulation of cancer cell motility and invasiveness. Moreover, combined inhibition of Sam68 expression and NEK2 activity cooperated to suppress TNBC cell migration and matrix invasion, suggesting that modulation of the identified splicing program is functionally relevant. Collectively, these results support the key oncogenic role of NEK2 and suggest that NEK2 targeting approaches represent promising therapeutic tools for TNBC treatment, whose efficacy could be amplified by co-targeting the vulnerability induced by splicing dysregulation in cancer cells.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the online repository: GEO database. accession number: GSE140754.

AUTHOR CONTRIBUTIONS

CN, FB, MP, and CS contributed to conception and design of the study. CN and CS wrote the manuscript. CN, FB, CC, MM, VP, SA, and MP performed the experiments, and analyzed and interpreted data. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

VP was supported by a fellowship from the Associazione Italiana Ricerca sul Cancro (23938). This work was supported by grants from the Associazione Italiana Ricerca sul Cancro (IG23416 and MFAG21899) and Breast Cancer Now (Catalyst Grant n. 2018NovPCC1283). Università Cattolica del Sacro Cuore contributed to the funding of this research project and its publication.

ACKNOWLEDGMENTS

We thank all current and past members of our laboratory for fruitful discussion and helpful suggestions throughout this work and Dr. Diletta Schito for technical assistance. We gratefully acknowledge Professor J. Stevenin for the purification of SR proteins. We wish to thank Dr. Pierre de la Grange (Genosplice, Paris) for RNA-seq analyses and bioinformatics support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.880654/full#supplementary-material

REFERENCES

1. Ule J, Blencowe BJ. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* (2019) 76:329–45. doi: 10.1016/j.molcel.2019.09.017
2. Bonnal SC, López-Oreja I, Valcárcel J. Roles and Mechanisms of Alternative Splicing in Cancer — Implications for Care. *Nat Rev Clin Oncol* (2020) 17:457–74. doi: 10.1038/s41571-020-0350-x
3. Paronetto MP, Passacantilli I, Sette C. Alternative Splicing and Cell Survival: From Tissue Homeostasis to Disease. *Cell Death Differ* (2016) 23:1919–29. doi: 10.1038/cdd.2016.91
4. Daguenet E, Dujardin G, Valcarcel J. The Pathogenicity of Splicing Defects: Mechanistic Insights Into pre-mRNA Processing Inform Novel Therapeutic Approaches. *EMBO Rep* (2015) 16:1640–55. doi: 10.15252/embr.201541116
5. Kahles A, Lehmann K-V, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive Analysis of Alternative Splicing Across Tumors From 8,705 Patients. *Cancer Cell* (2018) 34:211–24.e6. doi: 10.1016/j.ccell.2018.07.001
6. Trincado JL, Sebestyén E, Pagés A, Eyras E. The Prognostic Potential of Alternative Transcript Isoforms Across Human Tumors. *Genome Med* (2016) 8:85. doi: 10.1186/s13073-016-0339-3
7. Sebestyén E, Zawisza M, Eyras E. Detection of Recurrent Alternative Splicing Switches in Tumor Samples Reveals Novel Signatures of Cancer. *Nucleic Acids Res* (2015) 43:1345–56. doi: 10.1093/nar/gku1392
8. Stricker TP, Brown CD, Bandlamudi C, McNerney M, Kittler R, Montoya V, et al. Robust Stratification of Breast Cancer Subtypes Using Differential Patterns of Transcript Isoform Expression. *PloS Genet* (2017) 13:1–19. doi: 10.1371/journal.pgen.1006589
9. Bjørklund SS, Panda A, Kumar S, Seiler M, Robinson D, Gheeya J, et al. Widespread Alternative Exon Usage in Clinically Distinct Subtypes of Invasive Ductal Carcinoma. *Sci Rep* (2017) 7:5568. doi: 10.1038/s41598-017-05537-0
10. Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, et al. RNA Sequencing of Cancer Reveals Novel Splicing Alterations. *Sci Rep* (2013) 3:1689. doi: 10.1038/srep01689
11. Anczuków O, Krainer AR. Splicing-Factor Alterations in Cancers. *RNA* (2016) 22:1285–301. doi: 10.1261/rna.057919.116
12. David CJ, Chen M, Assanah M, Canoll P, Manley JL. HnRNP Proteins Controlled by C-Myc Deregulate Pyruvate Kinase mRNA Splicing in Cancer. *Nature* (2010) 463:364–8. doi: 10.1038/nature08697
13. Caggiano C, Pieraccioli M, Panzeri V, Sette C, Bielli P. C-MYC Empowers Transcription and Productive Splicing of the Oncogenic Splicing Factor Sam68 in Cancer. *Nucleic Acids Res* (2019) 47:6160–71. doi: 10.1093/nar/gkz344
14. Das S, Anczuków O, Akerman M, Krainer AR. Oncogenic Splicing Factor SRSF1 Is a Critical Transcriptional Target of MYC. *Cell Rep* (2012) 1:110–7. doi: 10.1016/j.celrep.2011.12.001
15. Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, et al. The Splicing Factor SRSF1 Regulates Apoptosis and Proliferation to Promote Mammary Epithelial Cell Transformation. *Nat Struct Mol Biol* (2012) 19:220–8. doi: 10.1038/nsmb.2207
16. Naro C, Sette C. Phosphorylation-Mediated Regulation of Alternative Splicing in Cancer. *Int J Cell Biol* (2013) 2013:15. doi: 10.1155/2013/151839
17. Naro C, Bielli P, Sette C. Oncogenic Dysregulation of pre-mRNA Processing by Protein Kinases: Challenges and Therapeutic Opportunities. *FEBS J* (2021) 288:6250–72. doi: 10.1111/febs.16057
18. Naro C, Barbagallo F, Chieffi P, Bourgeois CF, Paronetto MP, Sette C. The Centrosomal Kinase NEK2 Is a Novel Splicing Factor Kinase Involved in Cell Survival. *Nucleic Acids Res* (2014) 42:3218–27. doi: 10.1093/nar/gkt1307
19. Moore MJ, Wang Q, Kennedy CJ, Silver PA. An Alternative Splicing Network Links Cell-Cycle Control to Apoptosis. *Cell* (2010) 142:625–36. doi: 10.1016/j.cell.2010.07.019
20. Goncalves V, Henriques A, Pereira J, Costa AN, Moyer MP, Moita LF, et al. Phosphorylation of SRSF1 by SRPK1 Regulates Alternative Splicing of Tumor-Related Rac1b in Colorectal Cells. *RNA* (2014) 20:474–82. doi: 10.1261/rna.041376.113
21. Zhou W, Yang Y, Xia J, Wang H, Salama ME, Xiong W, et al. NEK2 Induces Drug Resistance Mainly Through Activation of Efflux Drug Pumps and Is Associated With Poor Prognosis in Myeloma and Other Cancers. *Cancer Cell* (2013) 23:48–62. doi: 10.1016/j.ccr.2012.12.001
22. Roberts MS, Sahni JM, Schrock MS, Piemonte KM, Weber-Bonk KL, Seachrist DD, et al. LIN9 and NEK2 Are Core Regulators of Mitotic Fidelity That Can Be Therapeutically Targeted to Overcome Taxane Resistance. *Cancer Res* (2020) 80:1693–706. doi: 10.1158/0008-5472.CAN-19-3466
23. Fang Y, Zhang X. Targeting NEK2 as a Promising Therapeutic Approach for Cancer Treatment. *Cell Cycle* (2016) 15:895–907. doi: 10.1080/15384101.2016.1152430
24. Van Roosmalen W, Le Dévédec SE, Golani O, Smid M, Pulyakhina I, Timmermans AM, et al. Tumor Cell Migration Screen Identifies SRPK1 as Breast Cancer Metastasis Determinant. *J Clin Invest* (2015) 125:1648–64. doi: 10.1172/JCI74440
25. Naro C, De Musso M, Delle Monache F, Panzeri V, de la Grange P, Sette C. The Oncogenic Kinase NEK2 Regulates an RBFOX2-Dependent Pro-Mesenchymal Splicing Program in Triple-Negative Breast Cancer Cells. *J Exp Clin Cancer Res* (2021) 40:397. doi: 10.1186/s13046-021-02210-3
26. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment Landscape of Triple-Negative Breast Cancer — Expanded Options, Evolving Needs. *Nat Rev Clin Oncol* (2021) 9:7–9. doi: 10.1038/s41571-021-00565-2
27. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-Negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease. *Nat Rev Clin Oncol* (2016) 13:674–90. doi: 10.1038/nrclinonc.2016.66
28. Hayward DG, Clarke RB, Faragher AJ, Pillai MR, Hagan IM, Fry AM. The Centrosomal Kinase Nek2 Displays Elevated Levels of Protein Expression in Human Breast Cancer. *Cancer Res* (2004) 64:7370–6. doi: 10.1158/0008-5472.CAN-04-0960
29. Hayward DG, Fry AM. Nek2 Kinase in Chromosome Instability and Cancer. *Cancer Lett* (2006) 237. doi: 10.1016/j.canlet.2005.06.017
30. Cappello P, Blaser H, Gorrini C, Lin DCC, Elia AJ, Wakeham A, et al. Role of Nek2 on Centrosome Duplication and Aneuploidy in Breast Cancer Cells. *Oncogene* (2014) 33:2375–84. doi: 10.1038/onc.2013.183
31. Barbagallo F, Paronetto MP, Franco R, Chieffi P, Dolci S, Fry AM, et al. Increased Expression and Nuclear Localization of the Centrosomal Kinase Nek2 in Human Testicular Seminomas. *J Pathol* (2009) 217:431–41. doi: 10.1002/path.2471
32. Gu Z, Xia J, Xu H, Frech I, Tricot G, Zhan F. NEK2 Promotes Aerobic Glycolysis in Multiple Myeloma Through Regulating Splicing of Pyruvate Kinase. *J Hematol Oncol* (2017) 17:1–11. doi: 10.1186/s13045-017-0392-4
33. Di Agostino S, Fedele M, Chieffi P, Fusco A, Rossi P, Geremia R, et al. Phosphorylation of High-Mobility Group Protein A2 by Nek2 Kinase During the First Meiotic Division in Mouse Spermatocytes. *Mol Biol Cell* (2004) 15:1224–32. doi: 10.1091/mbc.e03-09-0638
34. Di Agostino S, Rossi P, Geremia R, Sette C. The MAPK Pathway Triggers Activation of Nek2 During Chromosome Condensation in Mouse Spermatocytes. *Development* (2002) 129:1715–27. doi: 10.1242/dev.129.7.1715
35. Naro C, Pellegrini L, Jolly A, Farini D, Cesari E, Bielli P, et al. Functional Interaction Between U1snRNP and Sam68 Insures Proper 3′ End Pre-mRNA Processing During Germ Cell Differentiation. *Cell Rep* (2019) 26:2929–2941.e5. doi: 10.1016/j.celrep.2019.02.058
36. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and Interpreting Cancer Genomics Data *via* the Xena Platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8
37. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A Compendium of RNA-Binding Motifs for Decoding Gene Regulation. *Nature* (2013) 499:172–7. doi: 10.1038/nature12311
38. Bailey TL. DREME: Motif Discovery in Transcription Factor ChIP-Seq Data. *Bioinformatics* (2011) 27:1653–9. doi: 10.1093/bioinformatics/btr261
39. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying Similarity Between Motifs. *Genome Biol* (2007) 8(2):R24. doi: 10.1186/gb-2007-8-2-r24
40. Terracciano F, Capone A, Montori A, Rinzivillo M, Partelli S, Panzuto F, et al. MYC Upregulation Confers Resistance to Everolimus and Establishes Vulnerability to Cyclin-Dependent Kinase Inhibitors in Pancreatic Neuroendocrine Neoplasm Cells. *Neuroendocrinology* (2021) 111:739–51. doi: 10.1159/000509865
41. Naro C, Jolly A, Di Persio S, Bielli P, Setterblad N, Alberdi AJ, et al. An Orchestrated Intron Retention Program in Meiosis Controls Timely Usage of Transcripts During Germ Cell Differentiation. *Dev Cell* (2017) 41:82–93. doi: 10.1016/j.devcel.2017.03.003

42. Zheng Y-Z, Xue M-Z, Shen H-J, Li X-G, Ma D, Gong Y, et al. PHF5A Epigenetically Inhibits Apoptosis to Promote Breast Cancer Progression. *Cancer Res* (2018) 78(12):3190–206. doi: 10.1158/0008-5472.CAN-17-3514
43. Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C. The RNA-Binding Protein Sam68 Modulates the Alternative Splicing of Bcl-X. *J Cell Biol* (2007) 176:929–39. doi: 10.1083/jcb.200701005
44. Henise JC, Taunton J. Irreversible Nek2 Kinase Inhibitors With Cellular Activity. *J Med Chem* (2011) 54:4133–46. doi: 10.1021/jm200222m
45. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* (2012) 490:61–70. doi: 10.1038/nature11412
46. Fallah Y, Brundage J, Allegakoen P, Shajahan-Haq AN. MYC-Driven Pathways in Breast Cancer Subtypes. *Biomolecules* (2017) 7(3):53. doi: 10.3390/biom7030053
47. Bielli P, Busà R, Paronetto MP, Sette C. The RNA-Binding Protein Sam68 Is a Multifunctional Player in Human Cancer. *Endocr Relat Cancer* (2011) 18: R91–R102. doi: 10.1530/ERC-11-0041
48. Frisone P, Pradella D, Di Matteo A, Belloni E, Ghigna C, Paronetto MP. SAM68: Signal Transduction and RNA Metabolism in Human Cancer. *BioMed Res Int* (2015) 2015:52895. doi: 10.1155/2015/528954
49. Matter N, Herrlich P, Ko H. Signal-Dependent Regulation of Splicing via Phosphorylation of Sam68. *Nature* (2002) 420:691–5. doi: 10.1038/nature01153
50. Paronetto MP, Cappellari M, Busà R, Pedrotti S, Vitali R, Comstock C, et al. Alternative Splicing of the Cyclin D1 Proto-Oncogene Is Regulated by the RNA-Binding Protein Sam68. *Cancer Res* (2010) 70:229–39. doi: 10.1158/0008-5472.CAN-09-2788
51. Lukong KE, Richard S. Sam68, the KH Domain-Containing superSTAR. *Biochim Biophys Acta - Rev Cancer* (2003) 1653:73–86. doi: 10.1016/j.bbcan.2003.09.001
52. Wu W, Baxter JE, Wattam SL, Hayward DG, Fardilha M, Knebel A, et al. Alternative Splicing Controls Nuclear Translocation of the Cell Cycle-Regulated Nek2 Kinase. *J Biol Chem* (2007) 282:26431–40. doi: 10.1074/jbc.M704969200
53. Valacca C, Bonomi S, Buratti E, Pedrotti S, Baralle FE, Sette C, et al. Sam68 Regulates EMT Through Alternative Splicing-Activated Nonsense-Mediated mRNA Decay of the SF2/ASF Proto-Oncogene. *J Cell Biol* (2010) 191:87–99. doi: 10.1083/jcb.201001073
54. Cappellari M, Bielli P, Paronetto MP, Ciccosanti F, Fimia GM, Saarikettu J, et al. The Transcriptional Co-Activator SND1 Is a Novel Regulator of Alternative Splicing in Prostate Cancer Cells. *Oncogene* (2014) 33:3794–802. doi: 10.1038/onc.2013.360
55. Song L, Wang L, Li Y, Xiong H, Wu J, Li J, et al. Sam68 Up-Regulation Correlates With, and Its Down-Regulation Inhibits, Proliferation and Tumourigenicity of Breast Cancer Cells. *J Pathol* (2010) 222:227–37. doi: 10.1002/path.2751
56. Koedoot E, Fokkelman M, Rogkoti V-M, Smid M, Van De Sandt I, De Bont H, et al. Uncovering the Signaling Landscape Controlling Breast Cancer Cell Migration Identifies Novel Metastasis Driver Genes. *Nat Commun* (2019) 10 (1):2983. doi: 10.1038/s41467-019-11020-3
57. Chan S, Sridhar P, Kirchner R, Lock YJ, Herbert Z, Buonamici S, et al. Basal-A Triple-Negative Breast Cancer Cells Selectively Rely on RNA Splicing for Survival. *Mol Cancer Ther* (2017) 16:2849–61. doi: 10.1158/1535-7163.MCT-17-0461
58. Hsu TY-T, Simon LM, Neill NJ, Marcotte R, Sayad A, Bland CS, et al. The Spliceosome Is a Therapeutic Vulnerability in MYC-Driven Cancer. *Nature* (2015) 525:384–8. doi: 10.1038/nature14985
59. La Rosa P, Bielli P, Compagnucci C, Cesari E, Volpe E, Vecchioli SF, et al. Sam68 Promotes Self-Renewal and Glycolytic Metabolism in Mouse Neural Progenitor Cells by Modulating Aldh1a3 pre-mRNA 3'-End Processing. *Elife* (2016) 5:e20750. doi: 10.7554/eLife.20750
60. Farini D, Cesari E, Weatheritt RJRJ, La Sala G, Naro C, Pagliarini V, et al. A Dynamic Splicing Program Ensures Proper Synaptic Connections in the Developing Cerebellum. *Cell Rep* (2020) 31:107703. doi: 10.1016/j.celrep.2020.107703
61. Paronetto MP, Zalfa F, Botti F, Geremia R, Bagni C, Sette C. The Nuclear RNA-Binding Protein Sam68 Translocates to the Cytoplasm and Associates With the Polysomes in Mouse Spermatocytes. *Mol Biol Cell* (2006) 17:14–24. doi: 10.1091/mbc.E05
62. De Paola E, Forcina L, Pelosi L, Pisu S, La Rosa P, Cesari E, et al. Sam68 Splicing Regulation Contributes to Motor Unit Establishment in the Postnatal Skeletal Muscle. *Life Sci Alliance* (2020) 3. doi: 10.26508/lsa.201900637
63. Locatelli A, Lange CA. Met Receptors Induce Sam68-Dependent Cell Migration by Activation of Alternate Extracellular Signal-Regulated Kinase Family Members. *J Biol Chem* (2011) 286:21062–72. doi: 10.1074/jbc.M110.211409
64. Fang Y, Zhang X. Cell Cycle Targeting NEK2 as a Promising Therapeutic Approach for Cancer Treatment. *Cell Cycle* (2016) 15(7):895-907. doi: 10.1080/15384101.2016.1152430
65. Nakka KK, Chaudhary N, Joshi S, Bhat J, Singh K, Chatterjee S, et al. Nuclear Matrix-Associated Protein SMAR1 Regulates Alternative Splicing *via* HDAC6-Mediated Deacetylation of Sam68. *Proc Natl Acad Sci USA* (2015) 112:E3374–83. doi: 10.1073/pnas.1418603112
66. Bielli P, Busà R, Di Stasi SM, Munoz MJ, Botti F, Kornblihtt AR, et al. The Transcription Factor FBI-1 Inhibits SAM68-Mediated BCL-X Alternative Splicing and Apoptosis. *EMBO Rep* (2014) 15:419–27. doi: 10.1002/embr.201338241

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

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fonc.2022.838911

Worldwide Prevalence and Clinical Characteristics of RAS Mutations in Head and Neck Cancer: A Systematic Review and Meta-Analysis

Ofra Novoplansky[1*†], *Sankar Jagadeeshan*[1†], *Ohad Regev*[2†], *Idan Menashe*[3] *and Moshe Elkabets*[1*]

[1] The Shraga Segal Department of Microbiology, Immunology and Genetics, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel, [2] Joyce & Irving Goldman Medical School, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel, [3] Department of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

OPEN ACCESS

Edited by:
Hua Tan,
National Human Genome Research Institute (NIH), United States

Reviewed by:
Moran Amit,
University of Texas MD Anderson Cancer Center, United States
Chen Li,
Free University of Berlin, Germany

***Correspondence:**
Ofra Novoplansky
novoplan@post.bgu.ac.il
Moshe Elkabets
moshee@bgu.ac.il

[†]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: 18 December 2021
Accepted: 11 April 2022
Published: 06 May 2022

Citation:
Novoplansky O, Jagadeeshan S, Regev O, Menashe I and Elkabets M (2022) Worldwide Prevalence and Clinical Characteristics of RAS Mutations in Head and Neck Cancer: A Systematic Review and Meta-Analysis. Front. Oncol. 12:838911. doi: 10.3389/fonc.2022.838911

In light of the development of RAS inhibitors, a reliable assessment of the prevalence of RAS mutations and their correlation with the clinical features of patients with HNC is crucially needed. This meta-analysis compiles the findings of 149 studies with over 8500 HNC patients and assesses the global prevalence of mutations in the HRAS, KRAS and NRAS genes. The available data were stratified according to geographical region, clinical features, and tumor characteristics, including human papillomavirus (HPV) infection status and tumor stage. In addition, the distribution of codon substitutions in each RAS gene was assessed. The estimated mutation rate is highest for HRAS (7%), followed by KRAS (2.89%) and NRAS (2.20%). HRAS prevalence in South Asia (15.28%) is twice as high as the global estimate. HRAS mutations are more prevalent in oral cavity and salivary gland tumors. In contrast, KRAS mutations are found more frequently in sinonasal tumors, and NRAS mutations are found chiefly in tumors of the nasopharynx. OR analyses show a significant association between HRAS mutations and a high tumor stage (OR=3.63). In addition, there is a significant association between HPV-positive status and KRAS mutations (OR=2.09). This study highlights RAS as a potential therapeutic target in certain subsets of HNC patients.

Keywords: HRAS, KRAS, NRAS, head and neck cancer, meta-analysis, clinical characteristics

1 INTRODUCTION

Head and neck cancer (HNC) includes neoplasms that arise in the oral cavity, pharynx, larynx, sinuses, nasal cavity, and salivary glands (1). The main risk factors associated with HNC include tobacco smoking, alcohol abuse, and human papillomavirus (HPV) infection. Other risk factors include exposure to wood and leather dust, Epstein Barr Virus (EBV) infection, and betel nut chewing (2). In recent decades intensive research has confirmed that HNC is exceptionally heterogeneous at the molecular

level, and there is no single genetic alteration or a unique dysregulated molecular pathway responsible for its development and progression (3–5). This heterogeneity may explain the limited efficiency of current systemic therapies for HNC, which emphasizes the need to study specific and less common genetic alterations that may affect disease characteristics and clinical outcomes in HNC patients.

RAS GTPase family proteins are crucial players in many signaling networks, controlling cell proliferation, differentiation, and survival (6). The RAS family members, HRAS, KRAS, and NRAS, share significant sequence homology and largely overlapping functions (7). Mutations in RAS family members are well-established drivers of cancer. Gain-of-function mutations in RAS genes are found in ∼19% of human cancers, most clustering in three hotspots at codons 12, 13, and 61 (8). The immense effort invested in the development of RAS inhibitors has led to several breakthroughs in recent years, allowing for the targeted treatment of patients with alterations in these RAS genes (9, 10), including HNC patients (11–13).

Many studies reported on the frequency of HRAS, KRAS, and NRAS mutational status in HPV-positive and HPV-negative HNC patients. Even though mutations in the members of the RAS gene family are seemingly rare in the general HNC patient population, these studies vary in their assessments on the prevalence of mutations in RAS genes. Therefore, the purpose of the current study was to conduct the first systematic review and meta-analysis evaluating the prevalence of mutations in RAS genes in HNC. By collecting data on over 8500 patients from 149 studies, we were able to reveal differences in the prevalence of RAS mutations between geographical regions, anatomical sites, stage of disease, and HPV status. Moreover, in light of the clinical development of codon-specific RAS inhibitors (namely, G12C and G12D), we have included a comprehensive analysis of codon substitution in RAS mutations.

2 METHODS

This systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist (14).

2.1 Study Design

We evaluated the prevalence of mutations in HRAS, KRAS, and NRAS genes in HNC patients.

2.2 Search Strategy

A systematic review of the literature was conducted by searching the PubMed, Embase, Web of Science, and Cochrane Central Register of Controlled Trials databases in June 2021 for studies published since 1 January 2000. The strings used in the systemic search of databases are detailed in the Supplementary Methods section. The bibliographies of retrieved studies and systematic reviews identified in the search were screened for relevant references. Publicly available databases were screened for unpublished data.

2.3 Selection Criteria

The inclusion criteria for the meta-analysis were that the study had to include a mutational analysis of at least one of the target genes (HRAS, KRAS, or NRAS) and a report the prevalence and frequency of mutations as an outcome measure. Exclusion criteria were defined as: 1) Studies displaying results from patients with tumors other than HNCs or mutations other than those in the target genes; 2) studies that did not report data related to the prevalence or frequency of mutations in the target genes; 3) studies that did not evaluate target genes for somatic mutations; 4) studies published before 1 January 2000; 5) studies that were conducted using cell lines or animal models; 6) studies of pediatric populations; 7) review articles, letters, personal opinions, book chapters, or conference abstracts; 8) studies containing data included in other studies or studies in which it was not possible to determine whether duplicate data were included; and 9) studies enrolling fewer than ten patients.

2.4 Data Extraction

Two researchers (SJ, ON) screened the studies at the title and abstract level, followed by a full-text review. Disagreements over inclusion were resolved by consensus adjudication, and studies were extracted into a standardized extraction database. Extracted variables included study cohort size, number of mutated cases for each RAS family gene, primary tumor location, tumor grade or stage, geographical origin of studied patients, mutation assessment method, mutated codon, HPV status, and biopsy type, if reported.

2.5 Evaluation of Quality and Risk of Bias

Our study selection process excluded individual case reports and cohorts of less than ten patients due to the risk of bias. All papers considered after initial screening were reviewed and scored for risk of bias according to the Joanna Briggs Institute Critical Appraisal Checklist for Studies Reporting Prevalence Data (15). Studies that did not evaluate all three RAS family members were considered more prone to risk of bias and were not included in the general prevalence analysis. In addition, publication bias and heterogeneity were assessed by visual inspection of funnel plots and *via* Egger's regression test (16) (**Figure S1 in the Supplementary Material**).

2.6 Statistical Analysis

Pooled prevalence rates, pooled odds ratios (ORs), and forest plots were generated using the R Meta and MetaFor Packages (17, 18). The Cochrane Q chi-squared test and the inconsistency index statistic (I^2) were used to examine the heterogeneity across studies. Fixed-effects models were used to assess the pooled prevalence of genes for results with low heterogeneity ($I^2 \leq 50\%$). Otherwise, random-effects models were used for the analyses. A sensitivity analysis using a "leave-one-out" paradigm from the built-in function in MetaFor, as proposed by Wang et al. (19), was used to assess each study's effect on the overall pooled prevalence and detected outliers (19). First, the pooled overall prevalence of mutations in the three different target genes (KRAS, HRAS, NRAS) was calculated with a corresponding 95% confidence interval (95% CI). Next,

subgroup analyses were performed according to geographical region, mutated codon position and anatomical site. Finally, we assessed the association between the RAS gene mutational status and HPV status or tumor grade using the R MetaBin function.

3 RESULTS

3.1 Study Selection

The flow diagram shown in **Figure 1** depicts the search strategy and study selection process. A total of 867 studies were retrieved from four electronic databases and a bibliography screen. After the removal of duplicates, 375 studies were considered potentially eligible for evaluation, but 217 did not meet all the inclusion criteria, leaving a final sample of 158 studies. Nine additional studies were excluded due to the high risk of bias. To reduce the risk of bias, only papers with the highest grade (n = 85) were included for pooled analyses of the overall mutation prevalence. The literature references for the studies included in the meta-analysis are listed in **Table S1 in the Supplementary Material**.

3.2 Study Characteristics

Detailed characteristics of the studies are provided in **Table S2 in the Supplementary Material**. Of the 149 studies included in the analysis, 112, 130, and 93 contained data pertaining to gene alterations in HRAS, KRAS, and NRAS, respectively, and 85 presented analyses of all three RAS family genes. In total, 148 of the included studies were cohort studies, while one was a phase 1 clinical trial. Forty-seven studies used targeted next-generation sequencing, 46 utilized Sanger sequencing, 23 employed whole-exome sequencing, 9 conducted Mass Array analysis, 4 used whole-genome analyses, and 20 employed other or mixed analysis methods. The anatomical location of the tumors in the included study cohorts are detailed in **Table S3 in the Supplementary Material**. The studies were conducted in 29 different countries. Four studies included mixed populations from various geographical regions.

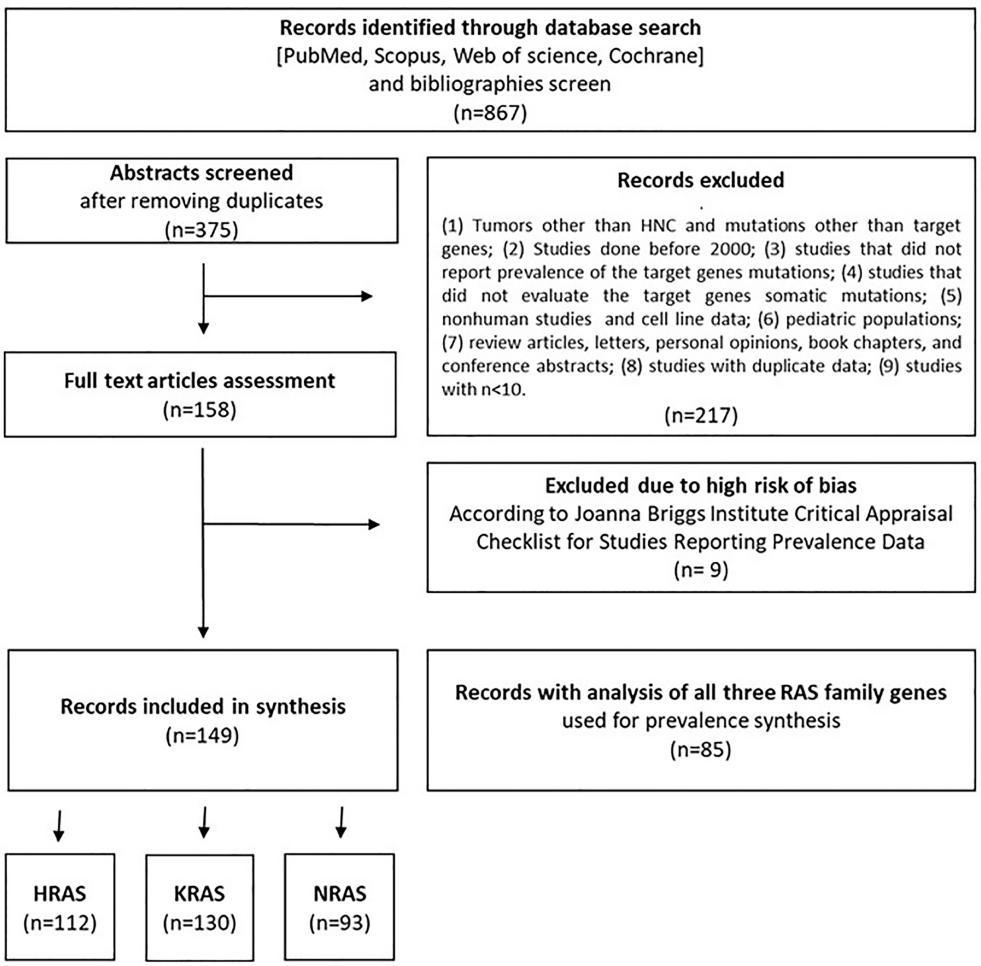


FIGURE 1 | Flow diagram of the literature search process and selection criteria.

3.3 Risk of Bias Within Studies

Nine studies were classified as having a high risk of bias and were therefore excluded from this meta-analysis. Eleven studies were classified as having a moderate risk of bias due to a small cohort size, while 57 studies were classified as having a moderate risk of bias, since they analyzed only one of the three RAS family target genes. The remaining 85 studies were classified as having a low risk of bias and were used in the general prevalence analysis. All low and moderate risk studies were used in prevalence analyses pertaining to tumor anatomical sites, mutated codons, and the association between RAS mutations and patient clinical features.

3.4 Prevalence of RAS Mutations

3.4.1 HRAS Mutations

Mutations in HRAS were identified in 564 tumors from 8501 patients. The mean prevalence of HRAS mutations was 7% (95% CI, 5.38-9.06, p <0.01, I^2 = 87%) (**Figure 2A**). Geographical region-specific analyses revealed significant differences in these rates in different regions of the world (Q = 22.51, Pv <0.0001). The mean frequency of HRAS mutations in South Asia was 15.28%, with this rate being higher than the rates in other geographical regions, including East Asia (5.07%), Europe (4.65%), and North America (6.87) (**Figure 3A**, **Figure S2 in the Supplementary Material**).

3.4.2 KRAS mutations

Mutations in KRAS were identified in 188 tumors from 8631 patients. The mean prevalence of KRAS mutations was 2.89% (95% CI, 2.19-3.80, p <0.0.1, I^2 = 67%) (**Figure 2B**), with no significant differences in prevalence between analyzed geographical regions (Q = 1.41, Pv = 0.7) (**Figure 3B**, **Figure S2 in the Supplementary Material**).

3.4.3 NRAS Mutations

Mutations in NRAS were identified in 113 tumors from 8512 patients. The mean prevalence of NRAS mutations was 2.20% (95% CI, 1.86-2.59, p <0.01, I^2 = 29%) (**Figure 2C**). No significant differences in these rates were observed between the different parts of the world (Q = 3.32, Pv = 0.34) (**Figure S2 in the Supplementary Material**).

3.5 Hot Spot Mutations and Amino Acid Substitutions

3.5.1 HRAS Mutations

In an analysis of all cases with HRAS mutations, 27%, 18%, and 36% were situated in codons 12, 13, and 61, respectively (**Figure 4A**). Mutations in codon 12 were mostly G12S point mutations (56.3%), while those in codon 13 were primarily G13R point mutations (46.8%). Lastly, mutations found in Q61 were primarily Q61R (49.2%), Q61K (26.4%), and Q61L (22.2%) point mutations (**Figure 4B**).

3.5.2 KRAS Mutations

In an analysis of all cases with KRAS mutations, 56%, 19%, and 0.8% were situated in codons 12, 13, and 61, respectively (**Figure 4A**). Among the codon 12 mutations, the most common amino acid substitution was G12D (51%), followed by G12V (16.3%) and G12C (12.9%) (**Figure 4C**).

3.5.3 NRAS Mutations

NRAS mutations were more evenly distributed among the different codons, with 29%, 13%, and 23% being situated in codons 12, 13, and 61, respectively. (**Figure 4A**). Analysis of amino acid substitutions was not feasible for NRAS mutations due to the low number of cases.

3.6 Difference in Prevalence of RAS Mutations Between Anatomical Sites

As HNC includes tumors that arise from a wide range of anatomical sites and sub-sites, an analysis of the frequency of mutations in the three RAS genes was performed for seven major anatomical areas. A summary of these analyses is presented in **Figure 5A** and **Figure S3 in the Supplementary Material**.

3.6.1 HRAS mutations

A significant difference in the prevalence of HRAS mutations was detected between anatomical sites (Q = 67.96, Pv <0.0001): HRAS mutations were found more frequently in tumors of the salivary glands (10.37%; 95% CI, 7.18-14.06) and oral cavity (7.36%; 95% CI, 5.39-9.76) than in tumors of the sinonasal cavity (1.2%; 95% CI, 0.2-3), oropharynx (2.6%; 95% CI, 1.12-4.56), nasopharynx (0.68%; 95% CI, 0-4.06), larynx (2.76%; 95% CI, 0.99-5.38), or hypopharynx (0.12%; 95% CI, 0-0.04). Salivary gland tumors exhibited a higher frequency of mutations in codon 61 (67%), while in tumors of the oral cavity, mutations in codon 12 were the most frequent (50%) (**Figure 5B**, left side).

3.6.2 KRAS mutations

A trend towards more frequent KRAS mutations was observed for tumors of the sinonasal cavity (5.67%; 95% CI, 1.33-12.74) as compared to tumors of the salivary glands (0.98%; 95% CI, 0.33-1.96), oral cavity (0.7%; 95% CI, 0.17-1.59), oropharynx (1.49%; 95% CI, 0.6-2.77), nasopharynx (0.83%; 95% CI, 0.29-1.63), larynx (1.43%; 95% CI, 0.34-3.25), or hypopharynx (0.84%; 95% CI, 0-3.18). However, these differences were not robust (Q = 8.5, Pv = 0.29). Mutations in codon 12 were the most frequent across all anatomic sites, followed by those in codon 13. Mutations in codon 61 were primarily detected in tumors of the oropharynx (17%) (**Figure 5B**, middle).

3.6.3 NRAS mutations

A significant difference between anatomical sites was also seen for NRAS mutations (Q = 18.37, Pv = 0.01), with a rate of 1.85% (95% CI, 0.92-3.1) in the nasopharynx compared to lower rates in tumors of the salivary glands (0.51%; 95% CI, 0.11-1.22), oral cavity (0.3%; 95% CI, 0.11-0.58), sinonasal cavity (0.28%; 95% CI, 0-1.65), oropharynx (0.65%; 95% CI, 0.28-1.16), larynx (0.16%; 95% CI, 0-0.68), or hypopharynx (0%; 95% CI, 0-0.85).

We note that the analyses of the mutated position in specific anatomical sites, and of the specific amino acid substitution, should be interpreted with caution owing to the limited number of mutated cases.

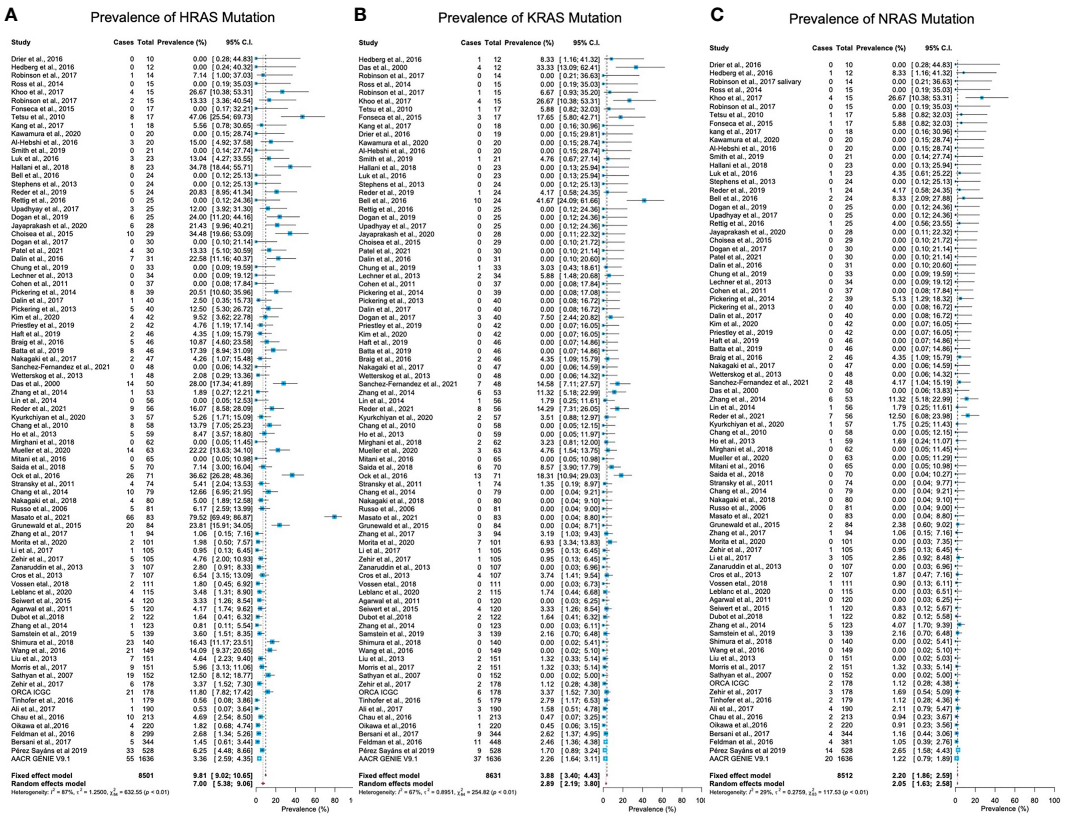


FIGURE 2 | Prevalence of RAS mutations in head and neck cancer. **(A)** The prevalence of HRAS mutations is 7% (95% CI, 5.38-9.06, p <0.01, I^2 = 87%). **(B)** The prevalence of KRAS mutations is 2.89% (95% CI, 2.19-3.80, p <0.0.1, I^2 = 67%). **(C)** The prevalence of NRAS mutations is 2.20% (95% CI, 1.86-2.59, p <0.01, I^2 = 29%). CI: Confidence interval. I^2: Inconsistency index.

3.7 Association Between RAS Mutations and Disease Stage/Grade

Tumor grade and stage are well-studied prognostic factors for HNC (20). In total, 44 studies reported details of the tumor stage or grade of patients along with the mutation status. Tumors with a stage or grade of 1 and 2 were defined as low-grade tumors, while those with a stage or grade of 3 and 4 were categorized as high-grade tumors. An OR analysis revealed a significant association between HRAS mutation and advanced stage (OR = 3.63; 95% CI, 1.53-8.64) (**Figure 6A**). KRAS (OR = 2.41; 95% CI, 0.85-6.86) and NRAS (OR = 1.52; 95% CI, 0.68-3.41) mutations were both associated with an OR>1, but the association did not reach statistical significance (**Figure S4 in the Supplementary Material**).

3.8 Association Between RAS Mutations and HPV Status

Of the 38 cohort studies that reported the HPV status of HNC patients, only 25 provided specific patient data, and of these, 17 included both HPV-negative and HPV-positive patients, thus allowing an OR analysis. This analysis revealed a significant association between HPV-positive status and KRAS mutations, with an OR of 2.09 (95% CI, 1.01-4.31) (**Figure 6B**), but no significant correlation between HPV-positive status and HRAS or NRAS mutations (**Figure S5 in the Supplementary Material**).

4 DISCUSSION

After years of extensive research, new strategies that target the RAS-MAPK pathway are now opening new therapeutic options for affected patients (9). The meta-analysis presented here compiles findings from the past two decades and provides updated insight into the global prevalence of mutations in RAS family genes, underscoring their potential as therapeutic targets in HNC patients.

The prevalence of mutations was highest for the HRAS gene, followed by KRAS and NRAS. This finding aligns with previous reports of the higher frequency of HRAS mutations in HNC as compared to its frequency in other cancer types in which KRAS mutations are most prevalent, followed by NRAS mutations (8). The results of our prevalence analysis diverge slightly from the results of The Cancer Genome Atlas project (21, 22), which has carried out one of the most significant studies on an HNC patient population. These slight differences may be due to the more heterogeneous population of patients from diverse geographical regions, disease stages, and detection methods included in our analysis.

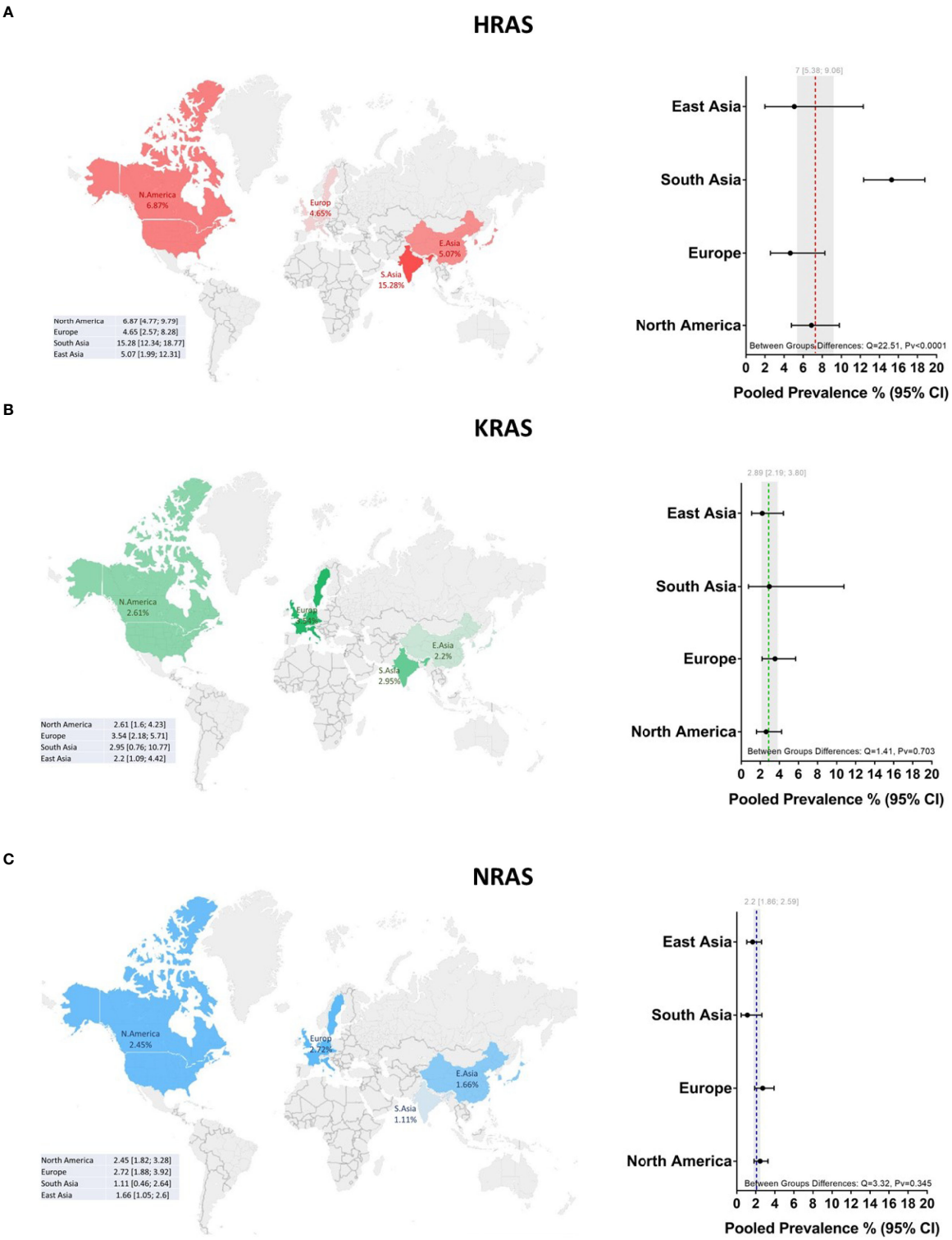


FIGURE 3 | Global prevalence of RAS mutations in head and neck cancer. Cohort studies were grouped according to the geographical origins of the patients. On the map of the world are shown the frequencies [%] of **(A)** HRAS, **(B)** KRAS, and **(C)** NRAS mutations in East Asia, South Asia, Europe, and North America. Dotted lines and gray shading correspond to the overall prevalence and the 95% CI. CI, Confidence interval; I^2, Inconsistency index.

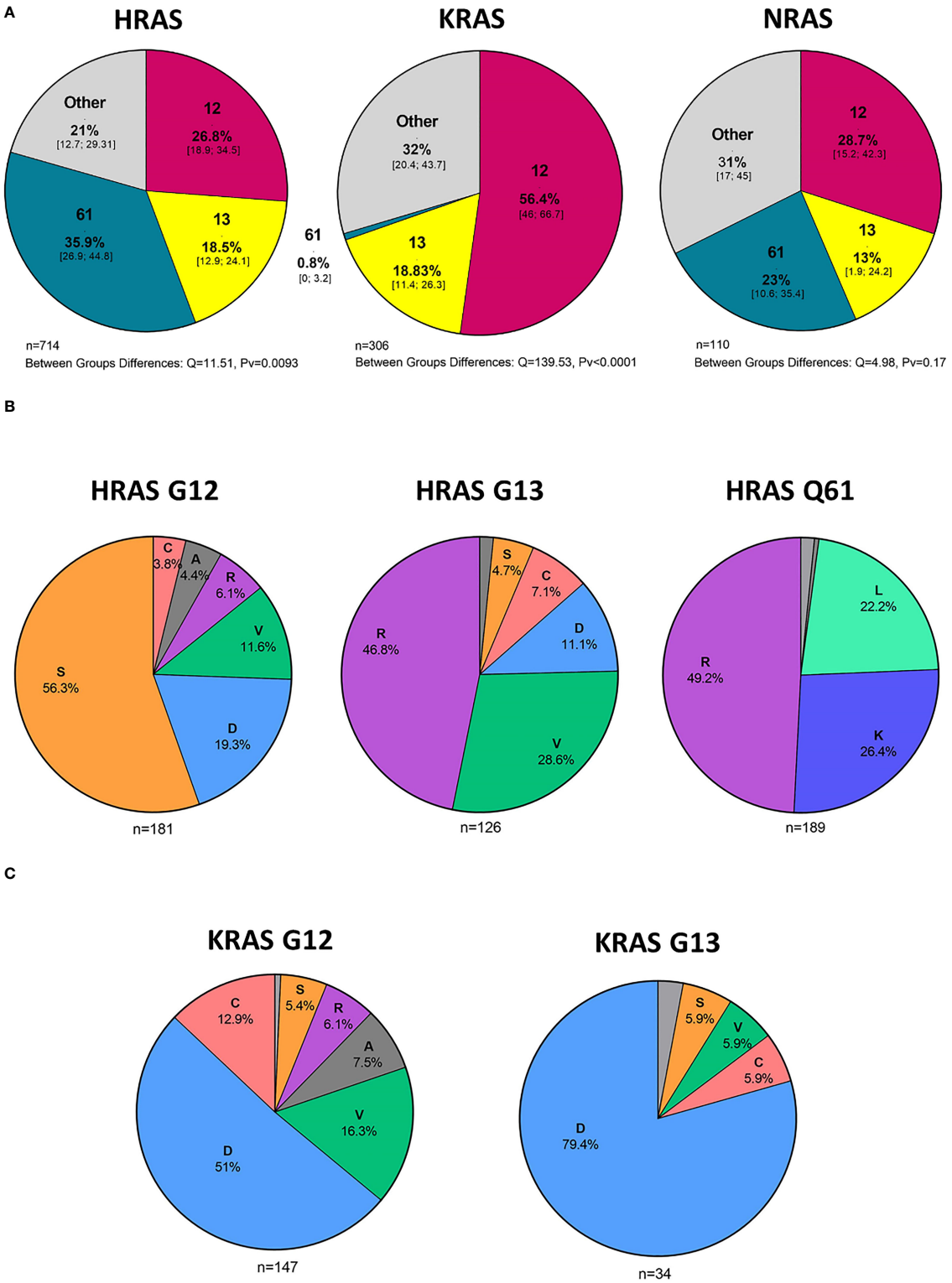


FIGURE 4 | Hot spot mutations and amino acid substitutions. **(A)** Mutated codons [%] in cases with HRAS, KRAS, and NRAS mutations and 95% CI. **(B)** Amino acid substitutions [%] in cases with KRAS G12 and G13 mutations. **(C)** Amino acid substitutions [%] in cases with HRAS G12, G13, and Q61 mutations. D - aspartic acid, C, cysteine; V, valine; S, serine; R, arginine; A, alanine; K, lysine; L, leucine; CI, Confidence interval.

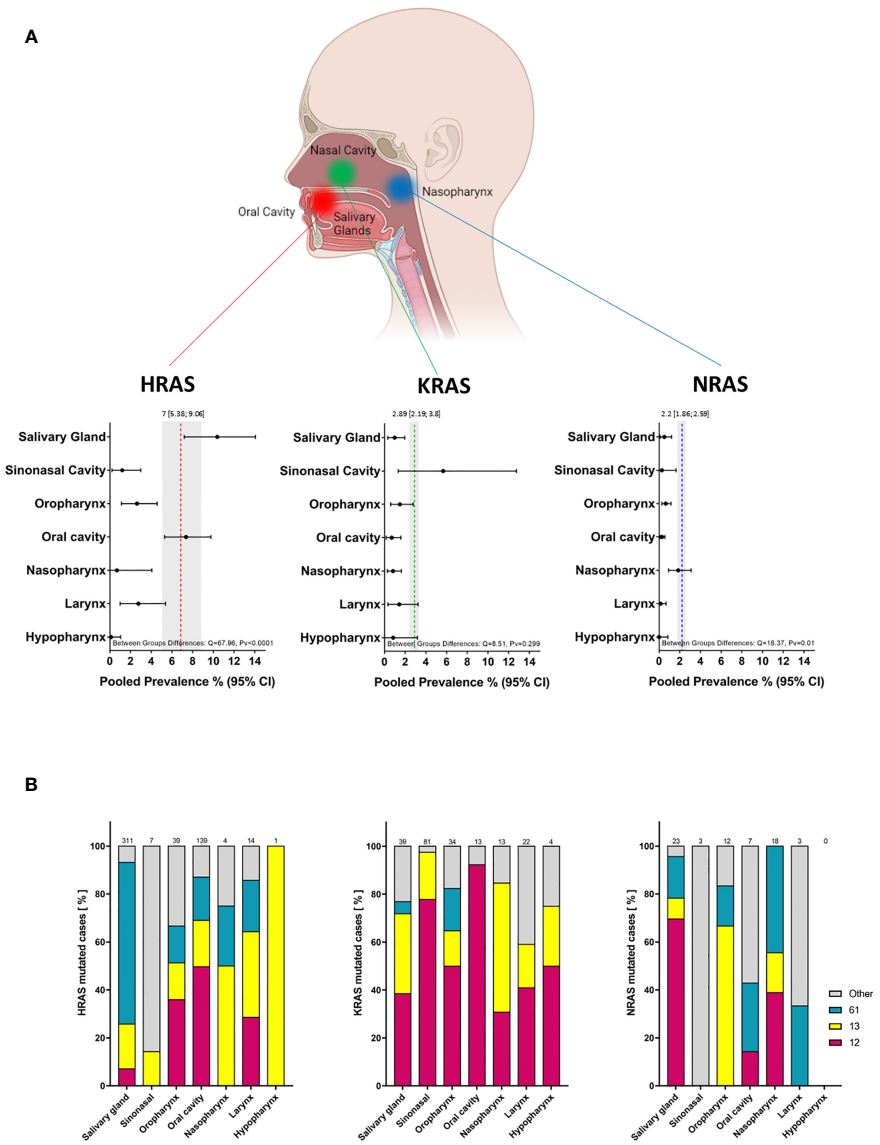


FIGURE 5 | Prevalence of RAS mutations and locations of mutated codons according to tumor anatomical site. **(A)** Prevalence of RAS mutations according to tumor anatomical site. Dotted lines and gray shading correspond to the overall prevalence and the 95% CI. **(B)** Mutated codon locations [%] according to tumor anatomical site. CI, Confidence interval; I^2, Inconsistency index.

Our analyses revealed differences in the prevalences of RAS mutations according to the anatomical site, which may account for some of the heterogeneity between cohorts in the overall prevalence analysis. HRAS mutations were more prevalent in oral cavity and salivary gland tumors. In contrast, KRAS mutations were more frequent in sinonasal tumors, and NRAS mutations were found chiefly in tumors of the nasopharynx. These findings emphasize the importance of taking the anatomical site of the tumor into consideration so as to achieve a more accurate assessment of RAS mutation frequencies. The variation in frequencies between tissue types may be due to differences in baseline expression and activity of RAS in different anatomical sites, which may, in turn, affect cellular reprogramming and tumor formation (8). Another explanation might be differences in the quality and quantity of exposure to risk factors (23).

Our data reveal a significantly higher prevalence of HRAS mutations in South Asia, corroborating previous studies on oral cancer in India (24–27). Those studies identified region-specific risk factors, such as smoking bidis (cigarettes wrapped in a tendu or temburni leaf) (28, 29), chewing betel nuts (30), and oral hygiene (31), that contribute, separately or synergistically, to the development of tumors, specifically within the oral cavity (32–35). Indeed, in our database, 86% of the patients from

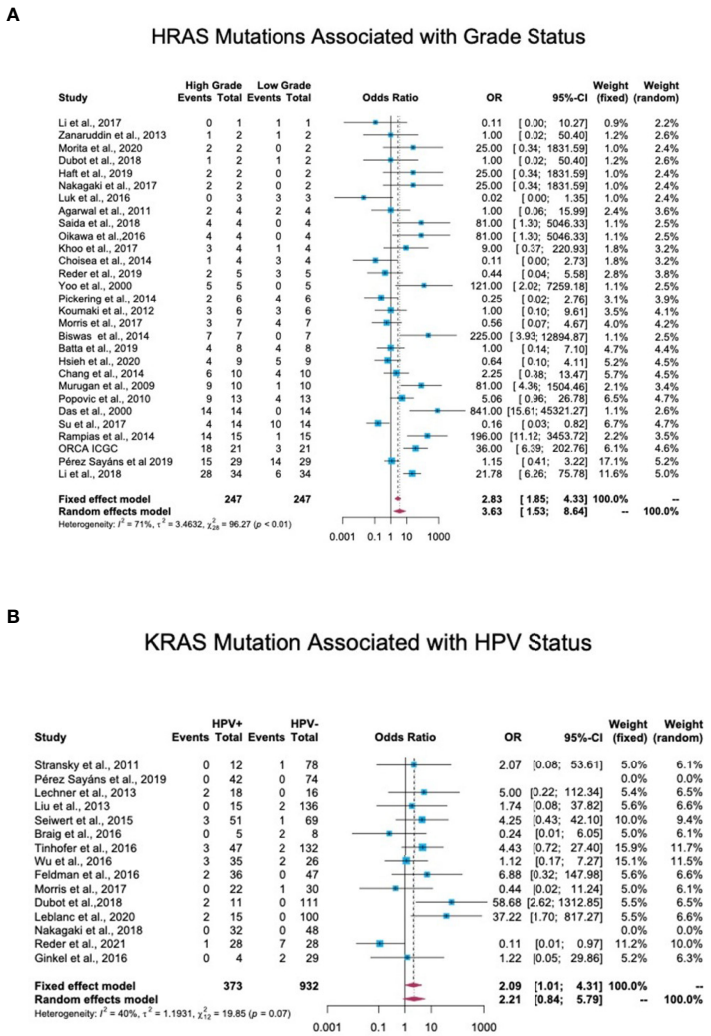


FIGURE 6 | Association between RAS mutations and patients' clinical features. **(A)** Association between HRAS mutations and tumor grade. In total, 44 studies reported details regarding tumor stage or grade and mutation status. Stage/grade 1 and 2 tumors were categorized as low-grade, while stage/grade 3 and 4 tumors were categorized as high-grade. An OR analysis exhibited a significant association between HRAS mutation status and advanced stage (OR = 3.63; 95% CI, 1.53-8.64). **(B)** Association between KRAS mutations and human papillomavirus (HPV) infection status. In total, 17 studies reported the detection of RAS mutations in both HPV-positive and HPV-negative patients. An OR analysis revealed a significant association between KRAS mutation status and HPV infection (OR = 2.09; 95% CI, 1.01-4.31). CI, Confidence interval; I^2, Inconsistency index; OR, Odds ratio.

South Asia suffered from oral cancer, as opposed to primary tumors in other sites. As noted above, the HRAS mutation frequency is higher in oral cancer worldwide. Thus, further studies are needed to determine whether exposure to such risk factors directly causes mutations in HRAS or whether these factors increase the odds of tumors developing in the oral cavity, in which the prevalence of HRAS mutations is high.

We found that the most frequent amino acid substitution in codon 12 of KRAS was G12D (51%), followed by G12V (16.3%) and G12C (12.9%). This finding provides an indication of the size of the population that could benefit from treatment with mutant-specific inhibitors, i.e., G12C and G12D KRAS inhibitors, that are in various stages of development.

A considerable percentage of HRAS mutations were present in codon 61, particularly in salivary gland tumors. To date, only limited studies have been performed to elucidate the etiology of these specific alterations, but recent analyses of patients with salivary gland cancer have indicated the diagnostic significance of these mutations (36). These findings may help evaluate the size of the subpopulations that could benefit from a particular treatment.

Data regarding the association between RAS gene mutations and prognosis in HNC are contradictory. Some studies link RAS mutations with stage and disease recurrence (37–40), while others predict better prognosis and overall survival (41–43). Our meta-analysis found that mutations in HRAS are significantly associated with high stage/grade scores, emphasizing the importance of

considering RAS mutational status when assessing patient prognosis. KRAS and NRAS mutations also exhibited a trend towards being associated with high stage/grade scores. The lower number of cases with KRAS or particularly of NRAS mutations that were available for OR analysis may account for the observed lack of statistical significance.

An association between RAS mutations and HPV status in HNC has been suggested (38, 44). In keeping with these studies, our data reveal a significant association between HPV-positive status and KRAS mutations. Studies on HPV-related cancers, mainly cervical cancer, demonstrate a similar association (45–47). Notably, KRAS mutations, HRAS mutations, and HPV infection were mutually exclusive in benign neoplasms of the head and neck (48). These findings suggest that RAS mutations in the context of HPV infection contribute to carcinogenesis.

Several inhibitors of the RAS-MAPK pathway are currently under evaluation as therapeutics for various cancers [(reviewed in (9)]. Therefore, knowledge regarding the prevalence of RAS family mutations and associated characteristics in HNC may enable researchers to better assess the need for and the potential of trials with molecularly relevant targeted therapeutics.

4.1 Limitations

Certain methodological limitations of this review should be considered. First, even after selecting only those studies with a low 'risk of bias score,' the heterogeneity between studies remained high. We believe that this is due to the heterogeneous nature of HNC, which includes a wide range of anatomical sites and etiologies. We attempted to address this issue by conducting additional sub-group analyses, which consistently revealed significant differences between groups. Due to differences in the categorization of sub-anatomic sites between reports and the lack of a minimal number of cases needed for sufficient statistical power, we could not perform analysis on sub-anatomical sites within the seven major anatomical sites. A second limitation of this analysis derives from the differences in the sequencing methods used in the various studies, which may have influenced overall pooled results by interfering with the accuracy and precision of pooled estimates. Third, we could not provide an analysis on RAS mutations association with exposure to risk factors due to insufficient patient-specific data and a lack of standardized categories of risk factors. Such data could potentially have strengthened the observed associations in this study and provided additional insights.

5 CONCLUSIONS

This study highlights RAS as a potential therapeutic target in certain subsets of HNC patients. The findings underscore the differences in the prevalence rates of HRAS, KRAS, and NRAS according to tumor anatomical site and geographical region. The analysis also demonstrates that RAS mutations are associated with tumor stage and HPV status.

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.

AUTHOR CONTRIBUTIONS

Conceptualization: ON, SJ, ME. data curation: ON, SJ. Formal analysis: ON, SJ, OR, IM. Methodology: ON, SJ, OR, IM. Resources: ME. Visualization: ON, SJ, OR, ME. Writing-original draft: ON, OR. Writing-review and editing: ON, SJ, OR, ME, IM. All listed authors have made substantial, contributions and have approved the submitted version. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by the Israel Science Foundation (ISF, 302/21 and 700/16) (to ME), the Israel Cancer Research Foundation (ICRF, 17-1693-RCDA) (to ME), United States-Israel Binational Science Foundation (BSF, 2017323) (to ME).

ACKNOWLEDGMENTS

Fellowships: Eileen & Louis Dubrovsky Doctoral Cancer Fellowship Endowment Fund, BGU fellow to ON, and a PBC post-doctoral fellowship from Israeli Council for Higher Education to SJ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.838911/full#supplementary-material

REFERENCES

1. Chow LQM. Head and Neck Cancer. *N Engl J Med* (2020) 382:60–72. doi: 10.1056/NEJMRA1715715
2. Boras VV, Fučić A, Baranović S, Blivajs I, Milenović M, Bišof V, et al. Environmental and Behavioural Head and Neck Cancer Risk Factors. *Cent Eur J Public Health* (2019) 27:106–9. doi: 10.21101/cejph.a5565
3. Leemans CR, Snijders PJFF, Brakenhoff RH. The Molecular Landscape of Head and Neck Cancer (2018). (Accessed June 6, 2019).
4. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. *Science (80-)* (2011) 333:1157–60. doi: 10.1126/science.1208130
5. Lawrence MS, Sougnez C, Lichtenstein L, Cibulskis K, Lander E, Gabriel SB, et al. Comprehensive Genomic Characterization of Head and Neck

Squamous Cell Carcinomas. *Nature* (2015) 517:576–82. doi: 10.1038/NATURE14129
6. Simanshu DK, Nissley DV, McCormick F. RAS Proteins and Their Regulators in Human Disease. *Cell* (2017) 170:17. doi: 10.1016/J.CELL.2017.06.009
7. Mo SP, Coulson JM, Prior IA. RAS Variant Signalling. *Biochem Soc Trans* (2018) 46:1325–32. doi: 10.1042/BST20180173
8. Prior IA, Hood FE, Hartley JL. The Frequency of Ras Mutations in Cancer. *Cancer Res* (2020) 80:2669–974. doi: 10.1158/0008-5472.CAN-19-3682
9. Chen K, Zhang Y, Qian L, Wang P. Emerging Strategies to Target RAS Signaling in Human Cancer Therapy. *J Hematol Oncol* (2021) 14:1–23. doi: 10.1186/s13045-021-01127-w
10. Skoulidis F, Li BT, Dy GK, Price TJ, Falchook GS, Wolf J, et al. Sotorasib for Lung Cancers With KRAS P.G12C Mutation. *N Engl J Med* (2021) 384:2371–81. doi: 10.1056/nejmoa2103695
11. Hong DS, Fakih MG, Strickler JH, Desai J, Durm GA, Shapiro GI, et al. KRAS G12C Inhibition With Sotorasib in Advanced Solid Tumors. *N Engl J Med* (2020) 383:1207–17. doi: 10.1056/nejmoa1917239
12. Hanna GJ, Guenette JP, Chau NG, Sayehli CM, Wilhelm C, Metcalf R, et al. Tipifarnib in Recurrent, Metastatic HRAS-Mutant Salivary Gland Cancer. *Cancer* (2020) 126:3972–81. doi: 10.1002/cncr.33036
13. Ho AL, Brana I, Haddad R, Bauman J, Bible K, Oosting S, et al. Tipifarnib in Head and Neck Squamous Cell Carcinoma With HRAS Mutations. *J Clin Oncol* (2021) 39:1856–64. doi: 10.1200/JCO.20.02903
14. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PloS Med* (2009) 6:e1000097. doi: 10.1371/JOURNAL.PMED.1000097
15. Aromataris E, Munn Z eds. JBI Manual for Evidence Synthesis. *JBI* (2020). Available at: https://synthesismanual.jbi.global. doi: 10.46658/jbimes-20-01
16. Egger M, Smith GD, Schneider M, Minder C. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *Br Med J* (1997) 315:629–34. doi: 10.1136/bmj.315.7109.629
17. Viechtbauer W. Conducting Meta-Analyses in R With the Metafor. *J Stat Softw* (2010) 36:1–48. doi: 10.18637/jss.v036.i03
18. Balduzzi S, Rücker G, Schwarzer G. How to Perform a Meta-Analysis With R: A Practical Tutorial. *Evid Based Ment Health* (2019) 22:153–60. doi: 10.1136/ebmental-2019-300117
19. Wang N. *How to Conduct a Meta-Analysis of Proportions in R : A Comprehensive Tutorial Conducting Meta-Analyses of Proportions in R*. (2018). doi: 10.13140/RG.2.2.27199.00161.
20. Hashim D, Genden E, Posner M, Hashibe M, Boffetta P. Head and Neck Cancer Prevention: From Primary Prevention to Impact of Clinicians on Reducing Burden. *Ann Oncol* (2019) 30:744. doi: 10.1093/ANNONC/MDZ084
21. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The Cbio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095
22. Sayáns MP, Petronacci CMC, Pouso AIL, Iruegas EP, Carrión AB, Peñaranda JMS, et al. Comprehensive Genomic Review of TCGA Head and Neck Squamous Cell Carcinomas (HNSCC). *J Clin Med* (2019) 8:1896. doi: 10.3390/jcm8111896
23. Simard EP, Torre LA, Jemal A. International Trends in Head and Neck Cancer Incidence Rates: Differences by Country, Sex and Anatomic Site. *Oral Oncol* (2014) 50:387–403. doi: 10.1016/J.ORALONCOLOGY.2014.01.016
24. Das N, Majumder J, Dasgupta UB. Ras Gene Mutations in Oral Cancer in Eastern India. *Oral Oncol* (2000) 36:76–80. doi: 10.1016/S1368-8375(99)00058-5
25. Saranath D, Chang SE, Bhoite LT, Panchal RG, Kerr IB, Mehta AR, et al. High Frequency Mutation in Codons 12 and 61 of H-Ras Oncogene in Chewing Tobacco-Related Human Oral Carcinoma in India. *Br J Cancer* (1991) 63:573–8. doi: 10.1038/bjc.1991.133
26. Paterson IC, Eveson JW, Prime SS. Molecular Changes in Oral Cancer may Reflect Aetiology and Ethnic Origin. *Eur J Cancer Part B Oral Oncol* (1996) 32:150–3. doi: 10.1016/0964-1955(95)00065-8
27. Chang SE, Bhatia P, Johnson NW, Morgan PR, McCormick F, Young B, et al. Ras Mutations in United Kingdom Examples of Oral Malignancies Are Infrequent. *Int J Cancer* (1991) 48:409–12. doi: 10.1002/ijc.2910480318
28. Rahman M, Sakamoto J, Fukui T. Bidi Smoking and Oral Cancer: A Meta-Analysis. *Int J Cancer* (2003) 106:600–4. doi: 10.1002/IJC.11265
29. Jayalekshmi PA, Gangadharan P, Akiba S, Koriyama C, Nair RRK. Oral Cavity Cancer Risk in Relation to Tobacco Chewing and Bidi Smoking Among Men in Karunagappally, Kerala, India: Karunagappally Cohort Study. *Cancer Sci* (2011) 102:460–7. doi: 10.1111/j.1349-7006.2010.01785.x
30. Garg A, Chaturvedi P, Gupta PC. A Review of the Systemic Adverse Effects of Areca Nut or Betel Nut. *Indian J Med Paediatr Oncol* (2014) 35:3–9. doi: 10.4103/0971-5851.133702
31. Gupta B, Bray F, Kumar N, Johnson NW. Associations Between Oral Hygiene Habits, Diet, Tobacco and Alcohol and Risk of Oral Cancer: A Case–Control Study From India. *Cancer Epidemiol* (2017) 51:7–14. doi: 10.1016/J.CANEP.2017.09.003
32. Hsieh L, Wang P, Chen I, Liao C, Wang H, Chen M, et al. Characteristics of Mutations in the P53 Gene in Oral Squamous Cell Carcinoma Associated With Betel Quid Chewing and Cigarette Smoking in Taiwanese G : C to A : T Transitions Were the Predominant Mutations Still Showed an Independent Effect on G : C to A . *Carcinogenesis* (2001) 22:1497–503. doi: 10.1093/carcin/22.9.1497
33. Chang Y, Tai K, Chou M, Tseng T. Synergistic Effects of Peroxynitrite on Arecoline-Induced Cytotoxicity in Human Buccal Mucosal Fibroblasts. *Toxicol Lett* (2000) 118:61–8. doi: 10.1016/S0378-4274(00)00262-9
34. Cancer I and Consortium G. Mutational Landscape of Gingivo-Buccal Oral Squamous Cell Carcinoma Reveals New Recurrently-Mutated Genes and Molecular Subgroups. *Nat Commun* (2013) 4:2873. doi: 10.1038/ncomms3873
35. Wu IC, Lu CY, Kuo FC, Tsai SM, Lee KW, Kuo WR, et al. Interaction Between Cigarette, Alcohol and Betel Nut Use on Esophageal Cancer Risk in Taiwan. *Eur J Clin Invest* (2006) 36:236–41. doi: 10.1111/J.1365-2362.2006.01621.X
36. Urano M, Nakaguro M, Yamamoto Y, Hirai H, Tanigawa M, Saigusa N, et al. Diagnostic Significance of HRAS Mutations in Epithelial-Myoepithelial Carcinomas Exhibiting a Broad Histopathologic Spectrum. *Am J Surg Pathol* (2019) 43:984–94. doi: 10.1097/PAS.0000000000001258
37. Zhang Z-CC, Fu S, Wang F, Wang H-YY, Zeng Y-XX, Shao J-YY. Oncogene Mutational Profile in Nasopharyngeal Carcinoma. *Onco Targets Ther* (2014) 7:457–67. doi: 10.2147/OTT.S58791
38. Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, et al. Integrative and Comparative Genomic Analysis of HPV-Positive and HPV-Negative Head and Neck Squamous Cell Carcinomas. *Clin Cancer Res* (2015) 21:632. doi: 10.1158/1078-0432.CCR-13-3310
39. Wang HM, Liao CT, Yen TC, Chen SJ, Lee LY, Hsieh CH, et al. Clues Toward Precision Medicine in Oral Squamous Cell Carcinoma: Utility of Next-Generation Sequencing for the Prognostic Stratification of High-Risk Patients Harboring Neck Lymph Node Extracapsular Extension. *Oncotarget* (2016) 7:63082–92. doi: 10.18632/oncotarget.11762
40. Reder H, Wagner S, Wuerdemann N, Langer C, Sandmann S, Braeuninger A, et al. Mutation Patterns in Recurrent and/or Metastatic Oropharyngeal Squamous Cell Carcinomas in Relation to Human Papillomavirus Status. *Cancer Med* (2021) 10:1347–56. doi: 10.1002/cam4.3741
41. Sathyan KM, Nalinakumari KR, Kannan S. H-Ras Mutation Modulates the Expression of Major Cell Cycle Regulatory Proteins and Disease Prognosis in Oral Carcinoma. *Mod Pathol* (2007) 20:1141–8. doi: 10.1038/modpathol.3800948
42. Szablewski V, Solassol J, Poizat F, Larrieux M, Crampette L, Mange A, et al. EGFR Expression and KRAS and BRAF Mutational Status in Intestinal-Type Sinonasal Adenocarcinoma. *Int J Mol Sci* (2013) 14:5170–81. doi: 10.3390/IJMS14035170
43. Ruíz-Godoy RLM, García-Cuellar CM, Herrera González NE, Suchil BL, Pérez-Cárdenas E, Sánchez-Pérez Y, et al. Mutational Analysis of K-Ras and Ras Protein Expression in Larynx Squamous Cell Carcinoma. *J Exp Clin Cancer Res* (2006) 25:73–8.
44. Tinhofer I, Budach V, Saki M, Konschak R, Niehr F, Jöhrens K, et al. Targeted Next-Generation Sequencing of Locally Advanced Squamous Cell Carcinomas of the Head and Neck Reveals Druggable Targets for Improving Adjuvant Chemoradiation. *Eur J Cancer* (2016) 57:78–86. doi: 10.1016/J.EJCA.2016.01.003
45. Hoover AC, Strand GL, Nowicki PN, Anderson ME, Vermeer PD, Klingelhutz AJ, et al. Impaired PTPN13 Phosphatase Activity in Spontaneous or HPV-Induced Squamous Cell Carcinomas Potentiates Oncogene Signaling Through the MAP Kinase Pathway. *Oncogene* (2009) 28:3960–70. doi: 10.1038/onc.2009.251
46. Mammas IN, Zafiropoulos A, Sifakis S, Sourvinos G, Spandidos DA. Human Papillomavirus (HPV) Typing in Relation to Ras Oncogene mRNA

Expression in HPV-Associated Human Squamous Cervical Neoplasia. *Int J Biol Markers* (2005) 20:257–63. doi: 10.1177/172460080502000409
47. Narisawa-Saito M, Inagawa Y, Yoshimatsu Y, Haga K, Tanaka K, Egawa N, et al. A Critical Role of MYC for Transformation of Human Cells by HPV16 E6E7 and Oncogenic HRAS. *Carcinogenesis* (2012) 33:910–7. doi: 10.1093/carcin/bgs104
48. Sasaki E, Masago K, Fujita S, Hanai N, Yatabe Y. Frequent KRAS and HRAS Mutations in Squamous Cell Papillomas of the Head and Neck. *J Pathol Clin Res* (2020) 6:154–9. doi: 10.1002/cjp2.157

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

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fonc.2022.939969

Next-Generation Sequencing Analysis of Gastric Cancer Identifies the Leukemia Inhibitory Factor Receptor as a Driving Factor in Gastric Cancer Progression and as a Predictor of Poor Prognosis

Cristina Di Giorgio[1†], *Silvia Marchianò*[1†], *Elisabetta Marino*[2†], *Michele Biagioli*[1], *Rosalinda Roselli*[3], *Martina Bordoni*[1], *Rachele Bellini*[1], *Ginevra Urbani*[1], *Angela Zampella*[3], *Eleonora Distrutti*[2], *Annibale Donini*[1], *Luigina Graziosi*[2†] *and Stefano Fiorucci*[1*]

[1] Department of Medicine and Surgery, University of Perugia, Perugia, Italy, [2] Azienda Ospedaliera Santa Maria della Misericordia, Perugia, Italy, [3] Department of Pharmacy, University of Naples Federico II, Naples, Italy

Gastric cancer (GC) is the third cause of cancer-related mortality worldwide. Nevertheless, because GC screening programs are not cost-effective, most patients receive diagnosis in the advanced stages, when surgical options are limited. Peritoneal dissemination occurs in approximately one-third of patients with GC at the diagnosis and is a strong predictor of poor outcome. Despite the clinical relevance, biological and molecular mechanisms underlying the development of peritoneal metastasis in GC remain poorly defined. Here, we report results of a high-throughput sequencing of transcriptome expression in paired samples of non-neoplastic and neoplastic gastric samples from 31 patients with GC with or without peritoneal carcinomatosis. The RNA-seq analysis led to the discovery of a group of highly upregulated or downregulated genes, including the leukemia inhibitory factor receptor (LIFR) and one cut domain family member 2 (ONECUT2) that were differentially modulated in patients with peritoneal disease in comparison with patients without peritoneal involvement. Both LIFR and ONECUT2 predicted survival at univariate statistical analysis. LIFR and its major ligand LIF belong to the interleukin-6 (IL-6) cytokine family and have a central role in immune system regulation, carcinogenesis, and dissemination in several human cancers. To confirm the mechanistic role of the LIF/LIFR pathway in promoting GC progression, GC cell lines were challenged *in vitro* with LIF and a LIFR inhibitor. Among several GC cell lines, MKN45 cells displayed the higher expression of the receptor, and their exposure to LIF promotes a concentration-dependent proliferation and epithelial–mesenchymal transition (EMT), as shown by

OPEN ACCESS

Edited by:
Daniele Vergara, University of Salento, Italy

Reviewed by:
Caroline Aquino Moreira-Nunes, Federal University of Ceara, Brazil
Lorenzo Antonuzzo, University of Florence, Italy

***Correspondence:**
Stefano Fiorucci
stefano.fiorucci.@unipg.it

†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: *09 May 2022*
Accepted: *30 May 2022*
Published: *30 June 2022*

Citation:
Di Giorgio C, Marchianò S, Marino E, Biagioli M, Roselli R, Bordoni M, Bellini R, Urbani G, Zampella A, Distrutti E, Donini A, Graziosi L and Fiorucci S (2022) Next-Generation Sequencing Analysis of Gastric Cancer Identifies the Leukemia Inhibitory Factor Receptor as a Driving Factor in Gastric Cancer Progression and as a Predictor of Poor Prognosis. Front. Oncol. 12:939969. doi: 10.3389/fonc.2022.939969

modulation of relative expression of E-cadherin/vimentin along with JAK and STAT3 phosphorylation and acquisition of a migratory phenotype. Furthermore, exposure to LIF promoted the adhesion of MKN45 cells to the peritoneum in an *ex vivo* assay. These effects were reversed by the pharmacological blockade of LIFR signaling. Together, these data suggest that LIFR might have a major role in promoting disease progression and peritoneal dissemination in patients with GC and that development of LIF/LIFR inhibitors might have a role in the treatment of GC.

Keywords: gastric cancer (GC), metastases, peritoneal carcinomatosis (PC), trascriptome analysis, biomarker, LIF/LIFR axis, EC359

INTRODUCTION

Gastric adenocarcinoma (GAC) is the fifth most common cancer but the third leading cause of cancer-related death (1–3) worldwide (1, 2), with a 5-year survival rate of ≈ 30% (3). The GC is a phenotypically and genotypically heterogeneous disease driven by multiple causative factors, including environmental factors and diet, Helicobacter (H.) pylori or Epstein–Barr virus (EBV) infection and host genetics (4, 5). According to the classical Lauren's classification, GC is subdivided into three main histological subtypes: diffuse, intestinal, and mixed (6, 7). The diffuse subtype is generally more aggressive and predicts treatment resistance and poor prognosis (8). In contrast to the diffuse type, the intestinal GC is frequently associated with intestinal metaplasia, and H. pylori infection and its prevalence have faced a constant reduction in the last three decades in line with a progressive decrease of P. pylori infection in Western countries (9). Although the Lauren histological classification has been widely used over the past decades, its clinical significance remains limited because it does not reflect the molecular heterogeneity of the disease, which has been progressively elucidated by the diffuse application of next-generation sequencing (NGS) technologies to GC (10–13).

The Cancer Genome Atlas and the Asian Cancer Research Group have identified four distinct subtypes of GAC based on genetic and epigenetic signatures: EBV+, microsatellite instability, genome stability, and chromosomal instability (12, 14). These molecular patterns have been partially validated for clinical use, but there is still a need to better define negative or positive prognostic factors that will predict treatment efficacy.

Currently, radical chirurgical resection is the only therapeutic strategy that offers an effective cure for patients with GC (15). However, very often oncological curative surgery is prevented as most patients are diagnosed in an advanced stage with extensive lymph nodes involvement and distant metastases with limited survival rates. Thus, whereas Stage IIIC resected tumors are associated with 5-year survival rate of 18%, the survival rates for stage IA and IB tumors treated with surgery are 94% and 88%, respectively.

Metastasis is a multistep process (16, 17). A critical event in the formation of metastases is the epithelial–mesenchymal transition (EMT) process in which polarized epithelial cells undergo a process of de-differentiation, characterized by phenotypic changes that are supported by the profound reshaping of EMT biomarkers, including the downregulation of E-cadherin and the upregulation of N-cadherin or vimentin, along with the acquisition of migratory properties (18, 19) and a mesenchymal phenotype. Peritoneal metastases occur in approximately 30% of patients with GC at the time of diagnosis (20), and their presence impacts dramatically on patients survival (21). Furthermore, the peritoneal cavity is a common site of relapse of GC after treatment (22, 23). The poor response of peritoneal carcinomatosis (PC) to existing treatments highlights the need to better understand the promoting mechanisms and to identify molecular biomarkers that will predict development of PC in GC. Recently, NGS studies have shown that the leukemia inhibitory factor (LIF) is one of the highest expressed gene in various solid tumors, including stomach (24, 25), pancreas (26), colon (27), liver (28), and breast (29). Of relevance, LIF/LIF receptor (LIFR) overexpression in these tumors seems to predict a poor prognosis. LIF belongs to the interleukin-6 (IL-6) family of cytokines, promotes EMT, and is envisioned as a potential therapeutic target in many cancers (30). In target cells, LIF signaling is mediated by the formation of a heterodimeric complex assembled by the LIFRβ with the glycoprotein (GP) 130 subunit of IL-6 receptor. The GP130 subunit of the receptor is shared with other members of the IL-6 family of cytokines, whereas LIFRβ is shared only with oncostatin M, cardiotrophin-1, ciliary neurotrophic growth factor, and cardiotrophin-like cytokine. The downstream signaling of the LIF/LIFR pathway involves a JAK (Janus Kinase)-induced STAT3 (Signal Transducer And Activator Of Transcription 3) phosphorylation, AKT (Akt kinase), and mTor (mammalian target of rapamycin) (31–33). Furthermore, LIF is commonly upregulated in carboplatin and paclitaxel resistant cells, suggesting that LIF/LIFR overexpression might contribute to cancer chemoresistance (34). Nevertheless, the role of LIF in GC remains unclear, and some data suggest that LIF overexpression could be protective (35, 36).

In this paper, we report the transcriptome sequencing (RNA-seq) of paired samples of gastric mucosa and adenocarcinoma samples of patients with GC with or without PC and identified LIFR as one of the highest expressed genes in the GC. LIFR expression is a predictor of PC and poor prognosis. In addition, by using *in vitro* cancer cells and pharmacological approaches,

we demonstrate that inhibition of LIF/LIFR signaling might have utility in the treatment of GC.

MATERIALS AND METHODS

Patients and Specimens

Gastric carcinoma tissues were obtained from 31 patients undergoing surgical resection at the Department of Surgery at the Perugia University Hospital (Italy). Patients included in this series were from a larger cohort of patients with GC who underwent surgery for GC in the years 2014–2017. Patients were selected on the basis of availability of all clinical and histology data and at least 5-year follow-up in 2022, as well as paired tissue samples from normal and primary neoplastic tissues. None of them received chemotherapy or radiation before surgery. Specimen collection was freshly carried out during surgery by a biologist, and paired samples from of normal mucosa sample and neoplastic tissues were collected. Samples were transported to the Gastroenterology laboratory in RNA later and then snap-frozen at −80°C until use. Permission to collect post-surgical samples was granted to Prof. Stefano Fiorucci by the Ethics Committee of Umbria (CEAS), permit FI00001, no. 2266/2014 granted on February 19, 2014, and by University of Perugia Bioethics Committee, permit FIO0003, no. 36348 granted on May 6, 2020. An informed written consent was obtained by each patient before surgery.

AmpliSeq Transcriptome

High-quality RNA was extracted from tumor gastric mucosa and healthy mucosa using the PureLink™ RNA Mini Kit (Thermo Fisher Scientific), according to the manufacturer's instructions. RNA quality and quantity were assessed with the Qubit® RNA HS Assay Kit and a Qubit 3.0 fluorometer followed by agarose gel electrophoresis. Libraries were generated using the Ion AmpliSeq™ Transcriptome Human Gene Expression Core Panel and Chef-Ready Kit (Thermo Fisher Scientific), according the manufacturer's instructions. Briefly, 10 ng of RNA was reverse-transcribed with SuperScript™ Vilo™ cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA) before library preparation on the Ion Chef™ instrument (Thermo Fisher Scientific, Waltham, MA). The resulting cDNA was amplified to prepare barcoded libraries using the Ion Code™ PCR Plate, and the Ion AmpliSeq™ Transcriptome Mouse Gene Expression Core Panel (Thermo Fisher Scientific, Waltham, MA), Chef-Ready Kit, according to the manufacturer's instructions. Barcoded libraries were combined to a final concentration of 100 pM and used to prepare Template-Positive Ion Sphere™ (Thermo Fisher Scientific, Waltham, MA) Particles to load on Ion 540™ Chips, using the Ion 540™ Kit-Chef (Thermo Fisher Scientific, Waltham, MA). Sequencing was performed on an Ion S5™ Sequencer with Torrent Suite™ Software v6 (Thermo Fisher Scientific). The analyses were performed with a range of fold change of <−2 and >+2 and a p-value of <0.05, using Transcriptome Analysis Console Software (version 4.0.2), certified for AmpliSeq analysis (ThermoFisher). The transcriptomic data have been deposited as dataset on Mendeley data repository (Mendeley Data, doi: 10.17632/9t86hd78sj.1).

Gastric Cancer Cell Lines

Human gastric cell lines MKN74, MKN45, and KATO III were from the Japanase Collection of Research Bioresources, Human Science Resources Bank (Osaka, Japan). These cells were grown in RPMI 1640 (Sigma-Merk LIFe Science S.r.l. Milan, Italy) medium supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin, in a humidified 5% CO_2 atmosphere, 37°C. Cells, free from Mycoplasma contamination, confirmed by the use of Mycoplasma PCR Detection (Sigma) were regularly passaged to maintain exponential growth and used from early passages (<10 passages after thawing). To perform all experiments, cells were plated, serum-starved for 24 h, and stimulated for 8, 24, and 48 h.

Real-Time PCR

The RNA was extracted from patient biopsies using the Trizol reagent (Invitrogen) and from cell lines using and Direct-zol™ RNA MiniPrep w/Zymo-Spin™ IIC Columns (Zymo Research, Irvine, CA, USA), according to the manufacturer's protocol. After purification from genomic DNA by DNase I treatment (ThermoFisher Scientific, Waltham, MA USA), 2 µg of RNA from each sample was reverse-transcribed using the FastGene Scriptase Basic Kit (Nippon Genetics, Mariaweilerstraße, Düren, Germania) in a 20-µl reaction volume. Finally, 50 ng of cDNA was amp LIFied in a 20-µl solution containing 200 nM of each primer and 10 µl of the SYBR Select Master Mix (ThermoFisher Scientific). All reactions were performed in triplicate, and the thermal cycling conditions were as follows: 3 min at 95°C, followed by 40 cycles of 95°C for 15 s, 56°C for 20 s, and 72°C for 30 s, using a Step One Plus machine (Applied Biosystem). The relative mRNA expression was calculated accordingly to the $2^{-\Delta Ct}$ method. Primers used in this study were designed using the PRIMER3 (http://frodo.wi.mit.edu/primer3/) software using the NCBI (National Center for Biotechnology Information) database. RT-PCR (Reverse transcriptase-polymerase chain reaction) primers used in this study for human sample and human cell lines were as follows [forward (for) and reverse (rev)]: Cmyc (for: TCGGATTCTCT GCTCTCCTC; rev: TTTTCCACAGAAACAACATCG), E-cadherin (for: GAATGACAACAAGCCCGAAT; rev: TGAGGA TGGTGTAAGCGATG), Snail1 (for: ACCCACACTGGCGAG AAG; rev: TGACATCTGAGTGGGTCTGG), and vimentin (for: TCAGAGAGAGGAAGCCGAAA; rev: ATTCCACTTTGCGT TCAAGG).

Immunohistochemistry and Immunocytochemistry

Immunohistochemistry (IHC) was performed on paraffin-embedded human stomach. In brief, Ag retrieval was achieved by incubation of the slides for 90 min in the hot (95°C) sodium citrate buffer (pH 6.0) and 30 min of cooling at room temperature. Immunostaining technique was carried out using the commercial kit Elabscience®2-step plus Poly-HRP Anti-Rabbit/Mouse IgG Detection System (with DAB Solution) (Houston, TX 77079, USA.) Anti-LIFR Rabbit Polyclonal

Antibody (Ab) (ab235908; Abcam, Cambridge, UK) was incubated overnight at 4°C. Subsequently, sections were incubated with Polyperoxidase anti-Mouse/Rabbit IgG and then with DAB Working Solution, both supplied by the kit. Slides were counterstained with hematoxylin, dehydrated through ethanol and xylene, and coverslipped using a xylene-based mounting medium.

Slides were observed under microscope and the photos were obtained with the Nikon DS-Ri2 camera, with magnification of ×20, ×40, and ×100. Immunocytochemistry (ICC) was performed on MKN45, untreated or treated with LIF (10 ng/ml; 14890-H02H, SinoBiological, Düsseldorfer, 65760 Eschborn, Germany). Cells were plate on slides using cytospined. The spots obtained were fixed in 4% formalin for 20 min and then submitted at the same procedure of immunostaining with the commercial kit Elabscience®2-step plus Poly-HRP Anti-Rabbit/Mouse IgG Detection System (with DAB Solution) (Houston, TX 77079, USA). After incubation with LIFR primary Ab and secondary Ab supplied by the kit, cells were counterstained with hematoxylin and then observed under microscope with magnification of ×100.

Cell Proliferation Assay

The cell viability assay was done using the CellTiter 96 Aqueous One Solution Cell Proliferation Assay (Promega, Milano, Italy), a colorimetric method for accessing the number of viable cells in proliferation. The MTS assay protocol is based on the reduction of the MTS [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] tetrazolium by cells into a colored formazan product that is soluble in cell culture media. Briefly, on day 0, MKN45 cells were seeded in RPMI 1640 complete medium at 36×10^3 cells/100-µl well into 96-well tissue culture plate. On day 1, cells were serum-starved for 24 h, and on day 3, cells were primed with the LIFR major ligand, LIF (0.5, 5, 10, 50, and 100 ng/ml), or only with vehicle. In another experimental setting, on day 3, cells were triggered with LIF (10 ng/ml) plus LIFR antagonist, EC359 (25, 50, 100, and 1,000 nM) (MedChemExpress, NJ 08852, USA), and cell proliferation was assessed as mentioned above. Absorbance was measured using a 96-well reader spectrophotometer (490 nm). In these experiments, each experimental setting was replicated 10 folds. For analysis, the well background readings with the medium alone were subtracted from the samples readouts.

Flow Cytometry

The intracellular flow cytometry staining for Ki-67 was performed using the following reagents: Ki-67 Monoclonal Antibody (SolA15), Alexa Fluor™ 488 (eBioscience™, San Diego, CA, United States), and DAPI (4',6-diamidin-2-fenilindolo) to characterize the cell cycle phases G0-G1, S-G2-M, and the apoptosis rate. Briefly, MK45 cells were seeded in six-well tissue culture plate (cell density 700×10^3 per well) in 100 µl of RPMI 1640 medium supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin and streptomycin at 37°C and 5% CO_2. Cells were serum-starved for 24 h and then incubated with LIF (10 and 50 ng/ml) or vehicle for 48 h. In another experimental setting, cells were first challenged with LIF (10 ng/ml) alone or in combination with LIFR antagonists EC359 25 nM. Before intracellular IC-FACS (Immun cells-Fluorescence-Activated Cell Sorting) staining cells were fixed for 30 min in the dark using IC Fixation Buffer (eBioscience™) and then permeabilized using Permeabilization Buffer (10X) (eBioscience™). Flow cytometry analyses were carried out using a three-laser standard configuration ATTUNE NxT (LIFe Technologies, Carlsbad, CA). Data were analyzed using FlowJo software (TreeStar) and the gates set using a fluorescence minus one (FMO) control strategy. FMO controls are samples that include all conjugated Abs present in the test samples except one. The channel in which the conjugated Ab is missing is the one for which the FMO provides a gating control.

Western Blot Analysis

MK45 cells were seeded in six-well tissue culture plate (cell density, 400×10^3 per well) in 100 µl of RPMI 1640 medium supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin and streptomycin at 37°C and 5% CO_2. Cells were serum-starved for 24 h and then incubated with LIF (10 ng/ml) and EC359 (25 and 100 nM), alone or in combination, for 48 h. Total lysates were prepared by homogenization of MKN45 cells in Ripa buffer containing phosphatase and protease inhibitors. Protein extracts were electrophoresed on 12% acrylamide Tris-Glycine gel (Invitrogen), blotted to nitrocellulose membrane, and then incubated overnight with primary Abs against Jak1 (1:500; sc-7228, Santa Cruz Biotechnology), phospho-Jak1 (1:1,000; GTX25493, GeneTex), STAT3 (1:500; sc-8019, Santa Cruz Biotechnology), phosho-Stat3 (1:1,000; GTX118000, GeneTex), and Gapdh (1:1,000; bs2188R, Bioss Antibodies). Primary Abs were detected with the HRP (horseradish peroxidase)-labeled secondary Abs. Proteins were visualized by Immobilon Western Chemiluminescent Reagent (MilliporeSigma) and iBright Imaging Systems (Invitrogen). Quantitative densitometry analysis was performed using ImageJ software. The degree of JAK1 and STAT3 phosphorylation was calculated as the ratio between the densitometry readings of p-Jak1/Jak1 and p-STAT3/STAT3, respectively.

Wound Healing Assay

MKN45 cells were seeded in RPMI 1640 complete medium at 800×10^3 cells per well into 24-well plate and used at 70%–80% confluence rate (37). On the day 1, the cell monolayers were gently scraped vertically with a new 0.2-ml pipette tip across the center of the well; during the scratch, the medium was not removed to avoid cell death. After scratching, the well was gently washed twice with PBS (Phosphate buffered saline) (Euroclone, Milan, Italy) to remove the detached cells and cell debris, and finally, fresh medium containing LIF (10 ng/ml) and EC359 (100 nM), alone or in combination, was added into each well. Immediately after scratch creation, the 24-well plate was placed under a phase-contrast microscope, and the first image of the scratch was acquired (T = 0 h) using an OPTIKAM Pro Cool 5 – 4083.CL5 camera. Cells were grown for additional 48 h, and images were taken at 24 and 48 h. The gap distance between scarps borders was quantified by assessing that area between the

two margins of the scratchs. All experiments were performed in triplicate.

Cell Adhesion to Peritoneum

For these experiments, MKN45 cells were grown in a complete RPMI medium and on day 2, starved, and left untreated or incubated with LIF (10 ng/ml) and EC359 (100 nM), alone or in combination for 48 h. On day 5, mouse parietal peritoneum sections (~1.6 cm^2) were placed in a 24-well culture plate, which had been filled with 1.0 ml of RPMI 1640 medium supplemented with 5% of FBS (38) and incubated with MKN45 cells. For this purpose, GC cells were first detached, fluorescently labeled with BCECF-AM (2',7'-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein) (10 μM) at 37°C for 30 min, and washed twice with PBS and after trypan blue staining; a suspension of living cells (5×10^5 cells/ml in RPMI 1640) were seeded on the peritoneum in a 24-well plate; and the plate was incubated at 37°C for 60 min. After a gentle washing with PBS, the cells adherent to the peritoneum were lysed with 1.0 ml of Tris (50 mM) plus 1% SDS (Sodium dodecyl sulfate). Fluorescence intensity was measured with a fluorescence spectrophotometer (Ex = 490 nm and Em = 520 nm). Experiments were carried out in quintuplicate.

Statistical Analysis

Patients' descriptive analysis was generated, and their differences were investigated using Student's t-test for quantitative data; normality test according to D'Agostino-Pearson was performed, and when not passed, quantitative data were compared using the Mann–Whitney test. For qualitative data, we used either the Fisher's exact test or the Chi-square test. Overall survival analyses were carried out with the Kaplan–Meier method, and differences were evaluated using log-rank test. Only variables that achieved statistical significance in the univariate analysis were subsequently evaluated in the multivariate analysis using Cox's proportional hazard regression model. ROC (receiver operating characteristic curve and Area Under the Curve) curves and AUC have also been calculated with the help of statistical software. A p-value of less than 0.05 was considered statistically significant. All statistical analyses were performed using the MedCalc Statistical Software version 14.8.1 (MedCalc Software, Ostend, Belgium), Prism 7.2 GraphPad, and SPSS, IBM version 23.

In vitro statistical analysis was carried out using the ANOVA followed by the nonparametric Mann–Whitney U-test or a two-tailed unpaired Student's t-test comparisons (* $p < 0.05$) using the Prism 6.0 software (GraphPad San Diego, CA, USA).

RESULTS

Patients

This study includes RNA-seq analysis of paired gastric samples from 31 patients with GC undergoing surgery at the Perugia University Hospital (2013–2019). Peritoneal metastasis dissemination was verified at surgery either macroscopically (P+) or microscopically (Cy+). This led to the identification of 19 patients with no peritoneal involvement (P0 and Cy0) and 12 who had peritoneal involvement (P+ or Cy+) at surgery. **Table 1** shows demographic characteristics, primary tumor features and surgical approaches followed in these patients. Patients were then followed up to 5 years after surgery, and, as shown in **Figure 1**, median survival time was 41 months and the 5-year overall survival rate was 35.7%. As shown in **Figure 1**, patients with peritoneal involvement have a significantly worse prognosis, whereas patients without peritoneal involvement had a median survival of 53 months (5-year survival rate, 49.2%), and the median survival time was 14.5 months in patients with PC (5-year survival rate, 25%).

Gene Expression Profile

The RNA transcription profile by AmpliSeq Transcriptome analysis (RNA-seq) of the two patient cohorts was carried out on paired samples of GCs and their matched normal tissues. The principal component analysis (PCA) of transcriptome shown in **Figure 2** highlighted the dissimilarities between GC samples obtained from patients with and without PC, showing only a partial overlap of the two groups. These results were confirmed by Venn diagram analysis of differentially expressed transcripts. As shown in **Figure 2**, this analysis allowed the identification of 341 transcripts belonging to AC+C subsets that were differentially modulated only in the cancer tissues with patients with peritoneal involvement. Specifically, 79 genes were upregulated and 262 downregulated (**Figure 2**). The *per-pathway* analysis of these differentially expressed genes using the TAC software (Affymetrix) demonstrated that the most modulated pathways in GC tumoral tissue belong to the EMT pathways, receptors and metabolism, inflammation, and signaling clusters (**Figure 2**). Analysis of differentially expressed (most upregulated and downregulated genes) in two cohorts of patients with GC (with or without PC) demonstrated that the top three upregulated genes were osteoglycin (Ong), LIFR, and secreted frizzled related protein 2 (Sfrp2); whereas the top three downregulated were fatty acid–binding protein 1 (Fabp1), one cut homeobox 2 transcriptional factor (Onecut2), and Ig superfamily protein glycoprotein A33 (Gpa33) genes (**Figure 2**). Whereas all six genes showed some degrees of correlation with patient survival (**Figure 3**), only the relative expression of Onecut2 and LIFR was statistically correlated with reduced patient survival at univariate analysis ($P < 0.05$). However, because, in comparison with normal mucosa, the expression of Onecut2 mRNA (39) was upregulated in the bulk tumor of patients with GC without peritoneal involvement but downregulated in those showing PC, we have focused our attention on LIFR.

LIFR Expression Is Increased in Mucosa of Patient With GC With Peritoneal Carcinomatosis

To explore the role of LIFR and LIF in GC, we have then assessed LIFR expression in 31 tumor samples from patients with GC and compared them to the corresponding non-neoplastic mucosa.

TABLE 1 | Clinical and pathological characterization of patients population at baseline.

Clinical pathological characteristics	Cy0 and P0 (n = 19)	Cy+ or P+ (n = 13)	P
Age*	76.3 ± 5.3	71± 12.7	N.S
Gender			
Male	10 (52.6%)	8 (61.5%)	N.S
Female	9 (47.4%)	5 (38.5%)	
N/L**	2.7 (1.2-22)	29 (1.6-5.7)	N.S
P/L**	136.4 (58.11-342.9)	144.1 (89.71-255-9)	N.S
L/M**	2.45 (0.97-6.68)	2.92 (1.14.03)	N.S
Surgery***			
Subtotal			
Gastrectomy GaGastrectomy	7 (36.9%)	8 (61.5%)	N.S
Total			
Gastrectomy	11 (57.9%)	5 (38.5%)	
Lymphoadenectomy			
Level*:**			
D1	2 (10.5%)	2 (15.5%)	N.S
D2	11 (57.9%)	9 (69,2%)	
D2+	6 (31.6%)	1 (7,7%)	
Lauren Hystotype*:**			
Intestinal	13 (68.4%)	6 (46.2%)	N.S
Diffuse	4 (21.1%)	6 (46,2%)	
Mixed	1 (5.3%)	1 (7.6%)	
Signet Ring Cell:			
Yes	1 (5.3%)	2 (15.4%)	N.S
No	18 (94.7%)	11 (84.6%)	
pT			
2	3 (15.8%)	1 (7.6%)	
3	9 (47.4%)	3 (23.1%)	N.S
4a	5 (26.3%)	7 (53.8%)	
4b	2 (10.5%)	2 (15.5%)	
pN			
0	5 (26.3%)	1 (7.6%)	
1	1 (5.3%)	1 (7.6%)	
2	3 (15.8%)	4 (30.8%)	N.S
3a	5 (26.3%)	2 (15.5%)	
3b	5 (26.3%)	5 (38.5%)	
Stage			
I	1 (5.4%)	0	
II	4 (21.0%)	0	
IIIa	4 (21.0%)	0	< 0.0001
IIIb	6(31.6)	0	
IIIc	4 (21.0%)	0	
IV	0	13 (100%)	
Lymphonodal Harvasted	46 (8-126)	34 (15-44)	0.05
Ln ratio	0.18 (0-0.85)	0.20 (0-0.75)	n.s
Veno-Lymp. Invasion			
Yes	17 (89.5%)	11 (84.6%)	
No	2 (10.5%)	1 (76%)	n.s
Periner. Invasion			
Yes	13 (68.4%)	10 (76.9%)	n.s
No	2 (10.5%)	3 (23.1%)	

*Patients were subdivided according to the presence or not of peritoneal disease (Cyoand Povs. Cy+ or P+). * Mean and SD; **Median values and range; ***data of Lauren classification were missed in one patient from each group. ns, not statistic.*

The results of this experiment demonstrated that LIFR expression in GC tissues was similar to that detected in paired samples of non-neoplastic mucosa (**Figure 4**). However, when patients with GC with or without peritoneal disease were compared, we found that LIFR expression was significantly increased in patients with PC (P-value of <0.05) (**Figure 4**). These findings were confirmed by LIFR IHC staining on GC biopsies. As shown in **Figure 4**, LIFR expression was detected as a faint signal in gastric glands on the normal mucosa, but the signal increased dramatically in the cancer tissues, showing a strong localization on the cell membrane of cancer cells (arrow), whereas some scattered signals were also detected in the tumor matrix (**Figure 4F**). Furthermore, to investigate the role of LIF/LIFR signaling, LIF mRNA expression level was assessed in paired samples of neoplastic and non-neoplastic mucosa of these patients, and, as shown in **Figure 4**, mRNA LIF expression showed a trend, although not significant, toward reduction in GC samples compared with non-neoplastic mucosa.

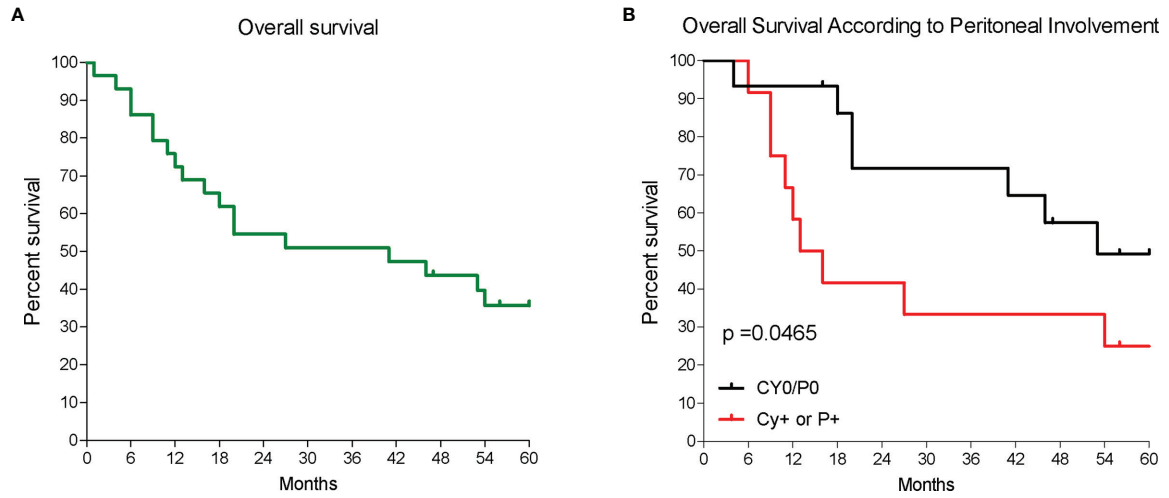


FIGURE 1 | Patients survival. **(A)** Overall survival of a cohort of patients with GC and **(B)** overall survival of 31 patients according to the presence of peritoneal disease either macroscopically or microscopically; $p < 0.05$.

LIF and LIFR Expression in GC Cell Lines

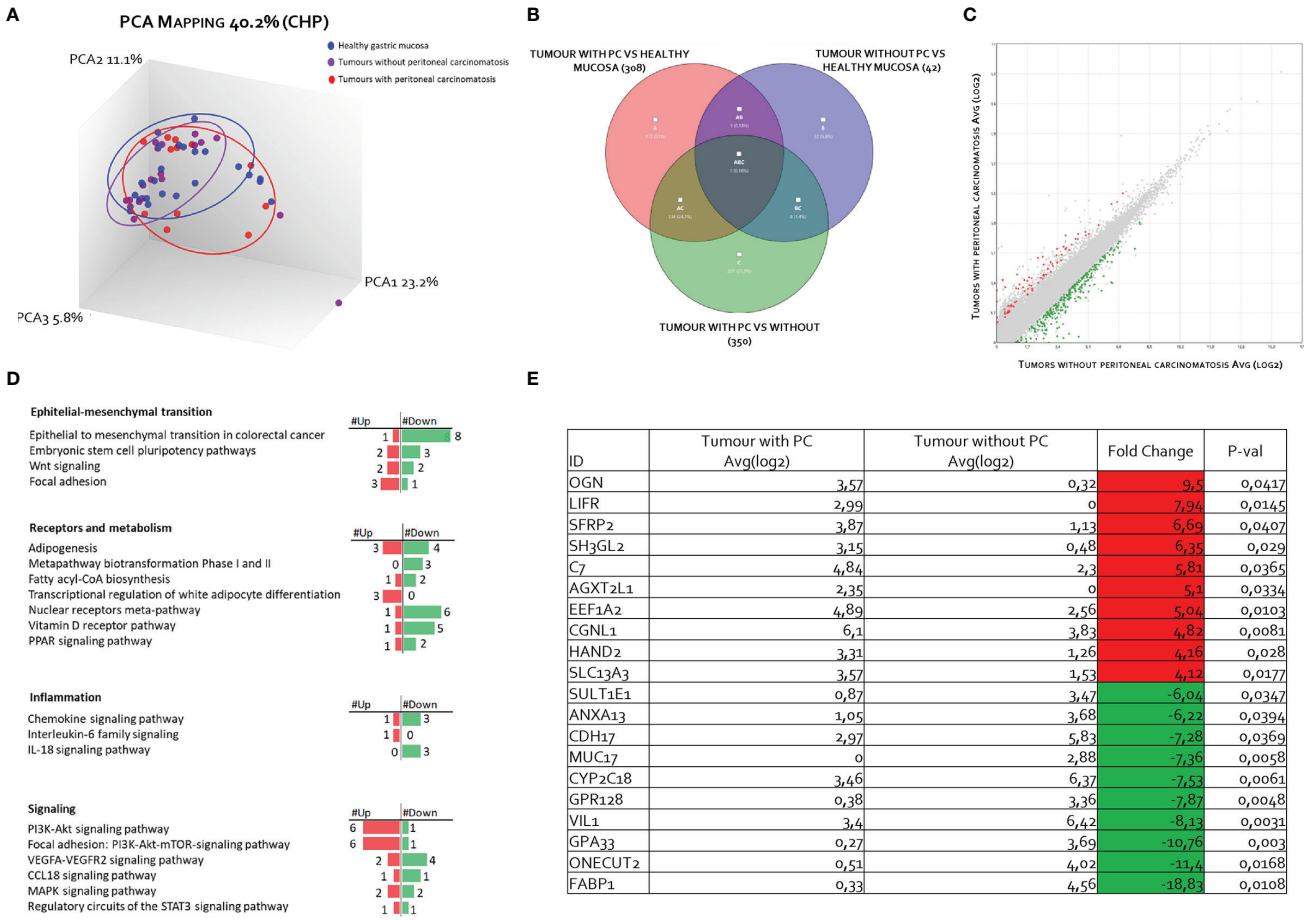
Because the abovementioned data demonstrated that LIFR expression increases in patients with peritoneal involvement, we have then investigated whether the LIF/LIFR signaling drives the EMT transition using GC cell lines (**Figure 5A**) and found that the poorly differentiated cell line MKN45 shows the strongest expression of LIFR in comparison with KATO III and the more differentiated cell line, MKN74. In contrast, expression of LIF mRNA displayed an opposite trend, with MKN45 showing the lower expression and MKN74 showing the higher expression (**Figure 5A**), further confirming that LIF and LIFR were oppositely regulated, as observed in human samples (**Figure 4**) (36). Thus, we have used MKN45 cells in the following experiments.

To investigate the role of LIF/LIFR in modulating GC cells proliferation and function, MKN45 cells were cultured with increasing concentrations of LIF at 0.5, 5, 10, 50, and 100 ng/ml, and cell proliferation was assessed as detailed in Materials and Methods. Data shown in **Figure 5** demonstrated that exposure to LIF induced LIFR expression, as assessed by LIFR staining by ICC (**Figure 5**), and also promoted cell proliferation in a concentration-dependent manner as shown by results of MTS proliferation assay and relative mRNA expression of CMYC (**Figure 5**). Importantly, however, challenging MKN45 cells with higher concentrations of LIF at 50 and 100 ng/ml resulted in a growth-retardation effect, suggesting that, at these concentrations, LIF might be directly cytotoxic (**Figure S1**) (36).

Subsequently, we have investigated the effect of LIF on MKN45 cell cycle and apoptosis (**Figure S1**). LIF at the concentration of 10 ng/ml modulated cell proliferation and cycle, reducing the percentage of G0-G1 cells while increasing the percentage of MKN45 cells in in S-G2-M phases (**Figure S1**). Again, these effects were biphasic and higher concentrations of LIF (10 and 100 ng/ml) promoted a cell growth arrest (**Figure S1**). Thus, additional experiments were performed using LIF (10 ng/ml) as the maximal effective concentration.

Because LIFR promotes EMT in various cell systems, we have investigated the expression of E-cadherin, vimentin, and SNAIL1, the three well-recognized biomarkers of EMT, in MKN45 cells (40). The results of these experiments demonstrated that exposure of MKN45 to LIF (0.5, 5, 10, 50, and 100 ng/ml), for 48 h promoted a concentration-dependent reduction of E-cadherin mRNA expression (**Figure 5**), which was statistically significant ($p < 0.05$) at 10 ng/ml, while increasing the expression of vimentin and SNAIL1 mRNA in the same range of concentrations (**Figure 5G**). These effects were lost at higher concentrations, LIF at 50 mg/ml, due to increased apoptosis rate and cell growth arrest (**Figure S1**). Collectively, these data suggest that LIFR agonism promotes cells growth and EMT of MKN45 cells.

To further shed light in these findings to LIFR activation, we have then investigated whether LIFR inhibition effectively reversed this pattern. In these studies, we used EC359 as LIFR inhibitor. EC359 is a small molecule that selectively binds LIFR and downregulates its pro-oncogenic effects *in vitro* and *in vivo* (31). For this purposes, MKN45 cells were growth in a medium with LIF (10 ng/ml), with or without increasing concentrations of EC359 at 25, 50, 100, and 1,000 nM, for 48 h. As shown in **Figure 6**, exposure to LIF again promoted cell proliferation as measured by MTS, and this effect was reversed in a concentration-dependent manner by EC359 (**Figure 6**). The above effects were statistically significant already at a concentration of 25 nM, whereas EC359 was cytotoxic at 1,000 nM. Similarly, the mRNA expression of CMYC was statistically reduced by 25 nM EC359 (**Figure 6**). In addition, the LIFR inhibition modulated the cell cycle as shown by Ki-67/DAPI IC-FACS staining (**Figure 6**). The cell cycle analysis revealed that EC359 alone did not decreased the rate of proliferative GC cells



E

ID	Tumour with PC Avg(log2)	Tumour without PC Avg(log2)	Fold Change	P-val
OGN	3,57	0,32	9,5	0,0417
LIFR	2,99	0	7,94	0,0145
SFRP2	3,87	1,13	6,69	0,0407
SH3GL2	3,15	0,48	6,35	0,029
C7	4,84	2,3	5,81	0,0365
AGXT2L1	2,35	0	5,1	0,0334
EEF1A2	4,89	2,56	5,04	0,0103
CGNL1	6,1	3,83	4,82	0,0081
HAND2	3,31	1,26	4,16	0,028
SLC13A3	3,57	1,53	4,12	0,0177
SULT1E1	0,87	3,47	-6,04	0,0347
ANXA13	1,05	3,68	-6,22	0,0394
CDH17	2,97	5,83	-7,28	0,0369
MUC17	0	2,88	-7,36	0,0058
CYP2C18	3,46	6,37	-7,53	0,0061
GPR128	0,38	3,36	-7,87	0,0048
VIL1	3,4	6,42	-8,13	0,0031
GPA33	0,27	3,69	-10,76	0,003
ONECUT2	0,51	4,02	-11,4	0,0168
FABP1	0,33	4,56	-18,83	0,0108

FIGURE 2 | Transcriptome analysis of gastric cancer and paired normal tissues in 31 patients with advanced gastric cancer. **(A)** Heterogeneity characterization of gastric samples showed by principal component analysis (PCA) plot. **(B)** Venn diagram of differentially expressed genes showing the overlapping regions between the three comparison groups: gastric cancer with peritoneal carcinomatosis vs. healthy mucosa (red subset), gastric cancer without peritoneal carcinomatosis vs. healthy mucosa (blue subset), and gastric cancer with peritoneal carcinomatosis vs. gastric cancer without peritoneal carcinomatosis (green subset). **(C)** Scatter plots of transcripts differentially expressed between gastric cancer tissues with peritoneal carcinomatosis and gastric cancer tissues without peritoneal carcinomatosis. **(D)** For pathways analysis of green subset, identification of pathways can be grouped in four clusters: epithelial–mesenchymal transition, receptors and metabolism, inflammation, and signaling. **(E)** Table showing the fold change of expression of the top 10 upregulated and downregulated genes included in green subset (fold change of <–2 or >+2 and p-value of < 0.05).

compared with untreated cells; instead, EC359, in combination with LIF, effectively reversed the effect of LIF in a statistically significant manner ($p < 0.05$), blocking the shift from resting cell in G0-G1 cell cycle phase to S-G2-M, as also demonstrated by the calculations of ratio between percent of G0-G1 and S-G2-M cells (**Figure 6**). Moreover, EC359 increased the apoptosis cell rates, which was diminished by LIF (**Figure 6**). Consistent with these findings, LIFR inhibition by EC359 reversed EMT features in MKN45 cells challenged with LIF. As shown in **Figure 6F**, at the concentration of 25 and 100 nM, EC359 downregulated E-cadherin and reduced the expression of vimentin. Taken together, these data demonstrate that EC359 effectively reverses GC cell proliferation and EMT promoted by LIF/LIFR signaling.

Because LIF/LIFR activation promotes a downstream signaling that involves several kinases, we have then investigated whether challenging MKN45 cells with LIF promotes JAK and STAT3 phosphorylation. The results of these experiments demonstrated that LIF at the concentration of 10 ng/ml increases the expression of LIFR and promotes the phosphorylation of both JAK and STAT3 and that these effects were reversed by LIFR inhibition by EC359 at 100 nM (**Figures 7A, B**).

To evaluated whether modulation of MKN45 by LIF promotes the acquisition of a migratory phenotype, we have performed a scratch wound healing assay, a validated method to functionally assess EMT (**Figure 8A**). For these purposes, MKN45 cells were growth in a complete medium, and, on the day 0, after a scratch was produced as described in Material and Methods, cells were challenged with LIF (10 ng/ml) and EC359 (100 nM) or the combinations of the two. Cell migration was

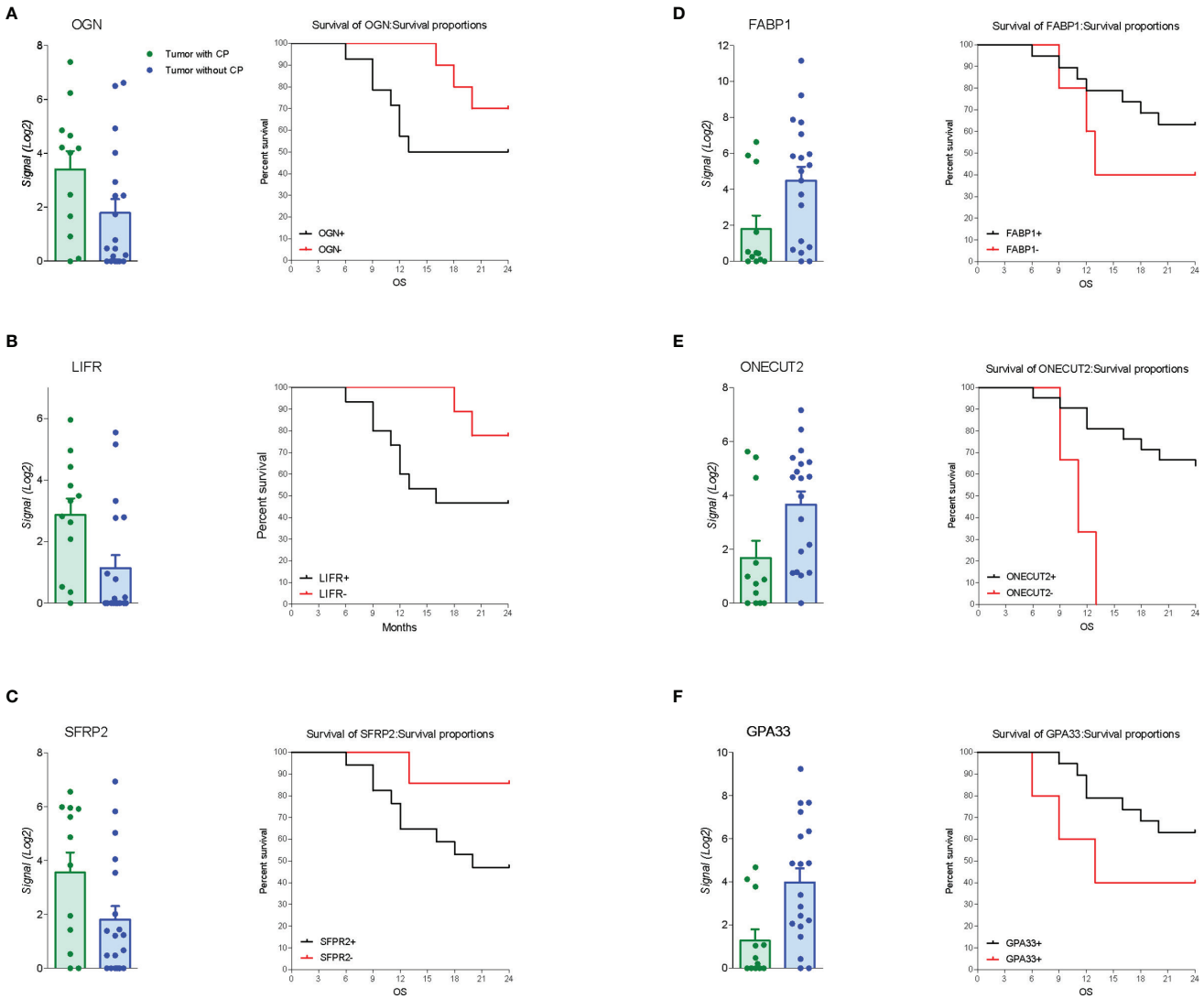


FIGURE 3 | Gene expression and survival curve. Left panel: Relative mRNA expression levels extract from RNA-seq analysis and overall survival of patients according to up regulated genes expression of **(A)** OGN, **(B)** LIFR, and **(C)** SFRP2. Right panel: Relative mRNA expression levels extract from RNA-seq analysis and overall survival of patients according to down regulated genes expression of **(D)** Fabp1, **(E)** Onecut2, and **(F)** Gpa33.

assessed by measuring the area between the two scratch margins at different time points: 0, 24, and 48 h. As illustrated in **Figure 8**, exposure to LIF promoted cell migration and wound closure with a reduction of the wound area of 45.41% at 24 h and 82.23% at 48 h. This pattern was reversed by exposure to EC359 ($p < 0.05$). In addition, EC359 alone reduced the percentage of wound closure compared with untreated cells, but these changes were not statistically significant. Similar findings were observed assessing the adhesion of MKN45 cells to the peritoneum. In this assay, although LIF promoted MKN45 adhesion to the mouse peritoneum, the effect was again significantly attenuated by co-treating the cells with EC359 by ≈30% (**Figure 8**). In summary, these results demonstrate that LIFR inhibition decreases LIF-induced ability to gain the migratory phenotype of MKN45 cells.

DISCUSSION

The LIF/LIFR signaling has been identified as a potential therapeutic target in the treatment of several cancers. In the present study, we report the transcriptome profiling (RNA-seq) of a group of patients with GC with or without PC. This investigation allowed the identification of LIFR as one of the highest expressed genes in patients with peritoneal involvement and as a strong predictor of a poor prognosis in these patients. In addition, we have shown that activation of LIF/LIFR signaling in GC cells promotes the acquisition of a mesenchymal phenotype, suggesting a potential mechanistic role of LIF/LIFR signaling in the development of peritoneal metastasis.

The PC is a relatively common localization of metastasis in GC, occurring in up to 14% of newly diagnosed patients with GC,

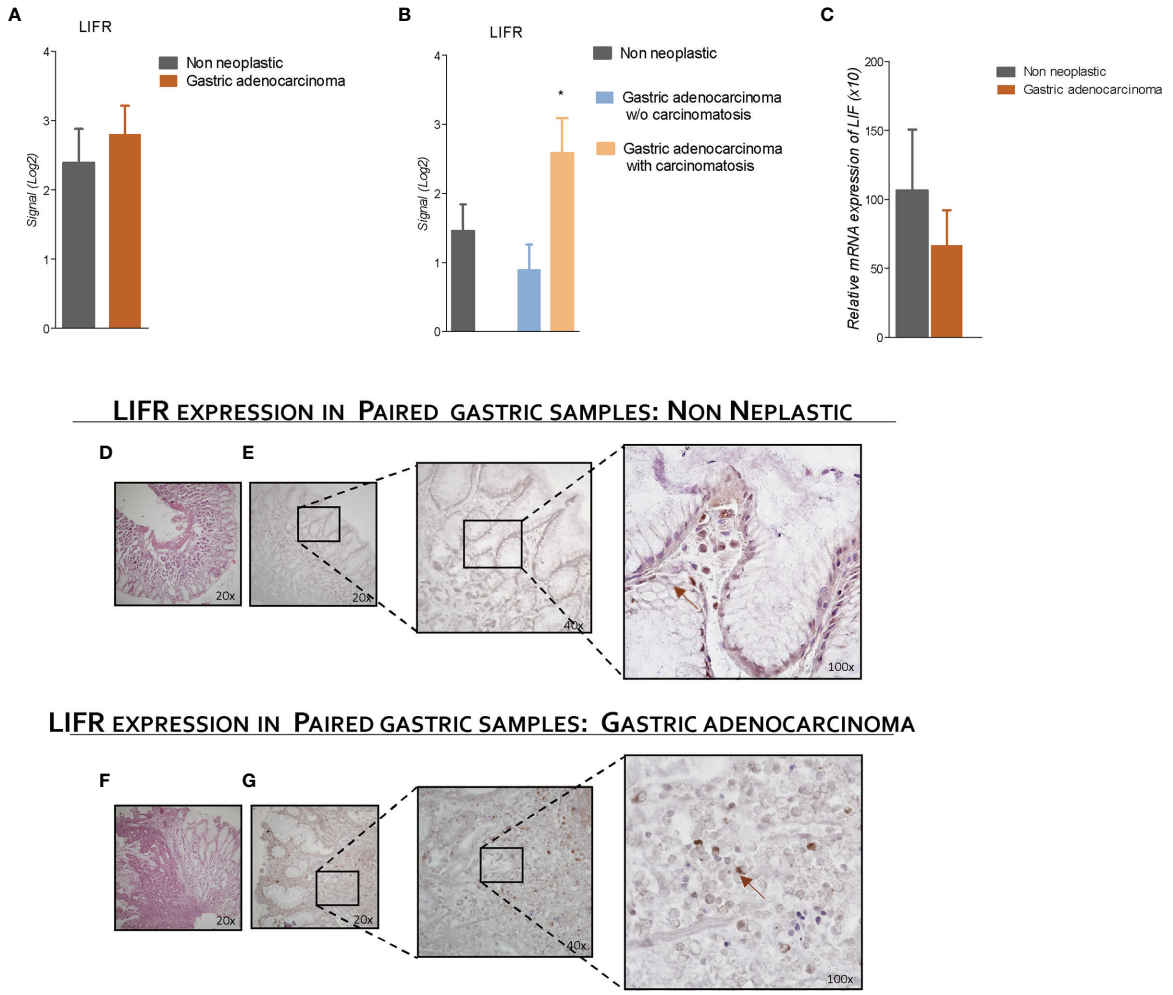


FIGURE 4 | LIFR is a negative prognostic factor for survival of a patient with GC with carcinomatosis. The expression of LIFR and LIF was examined in surgical samples from non-neoplastic and gastric adenocarcinoma mucosa obtained by patients with GC underwent surgery for GC treatment. Data shown are follows: Gene expression of LIFR (Log2) **(A)** in non-neoplastic vs. neoplastic mucosa and **(B)** in non-neoplastic and gastric adenocarcinoma w/o carcinomatosis vs. adenocarcinoma with carcinomatosis. **(C)** Relative mRNA expression of LIF. **(D)** H&E staining of non-neoplastic mucosa (magnification, ×20). **(E)** IHC staining of non-neoplastic mucosa (magnification, ×20, ×40, and ×100). **(F)** (hematoxylin and eosin) H&E staining of gastric adenocarcinoma mucosa (magnification, ×20). **(G)** IHC staining of IHC staining of gastric adenocarcinoma mucosa (magnification, ×20, ×40, and ×100). * represents statistical significance versus Non neoplastic tissue.

and is the most common site (~50%) of recurrence in patients with GC after radical surgery (41–43). In our series, as expected, presence of peritoneal metastasis was a strong predictor of poor prognosis, and mean survival time of patients with positive peritoneal cytology or macroscopic evidence of peritoneal metastasis at surgery (Cy+/P+) was approximately 12 months, significantly lower in comparison with the 60 months median survival observed in patients that were Cy0/T0. These data are in agreement with previous findings, confirming the fact that the development of a peritoneal disease is a strong predictor of a shorter-term survival in patients with GC.

In addition to the presence of peritoneal disease, by the transcriptome analysis of paired samples of neoplastic tissue and normal gastric mucosa, we have identified a group of six differentially expressed genes that predict poor prognosis: ONG, FABP1, LIF, ONECUT2, SFRP2, and GPA33. More specifically, we have shown that the top three upregulated genes—ONG, LIFR, and SFRP2, and the top three downregulated genes—FABP1, ONECUT2, and GPA33, were all associated with a poor prognosis, although statistically significant difference was detected only for the expression of ONECUT2 and LIFR ($P < 0.05$).

ONECUT2 belongs to the family of the ONECUT transcription factors, a small group of evolutionarily conserved proteins that play a role in the embryo, liver, pancreas, and neuronal system development (44). Although a role for ONECUT2 in cancer is not well defined, there is evidence that the expression of this gene is aberrantly upregulated in a variety of cancers including hepatocellular carcinoma, prostate cancer, colorectal cancer, and ovarian cancer, suggesting a role for this

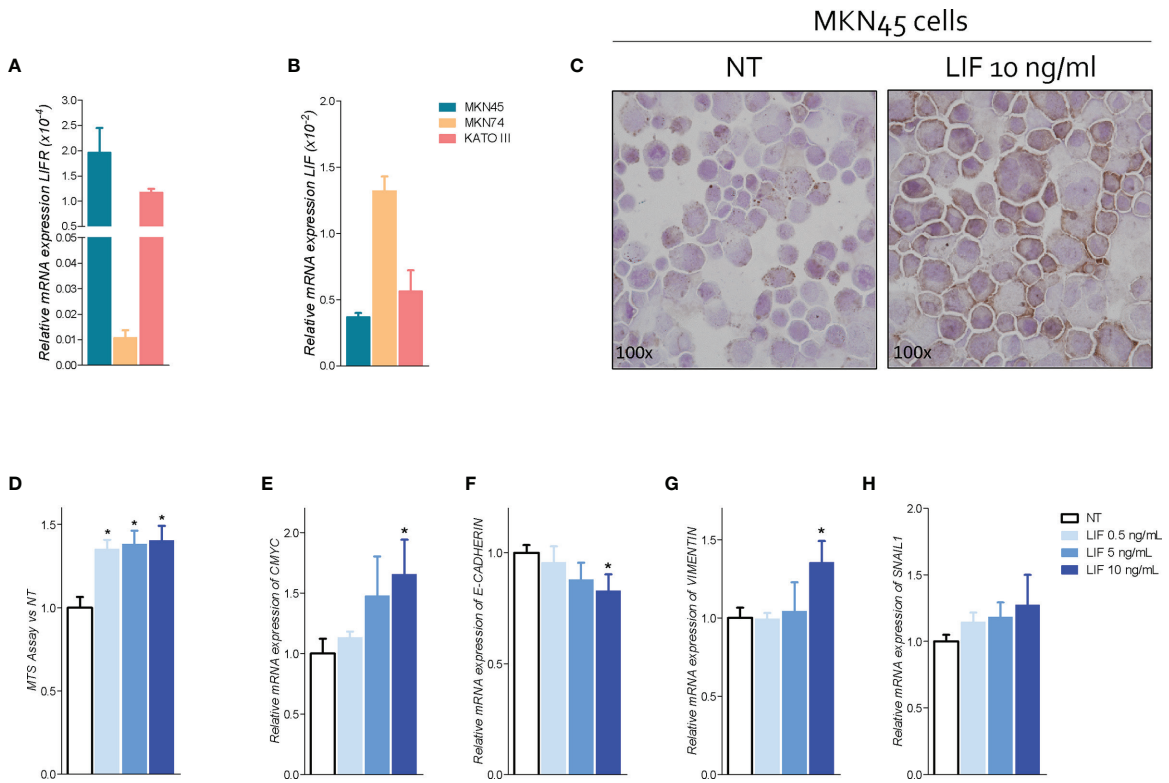


FIGURE 5 | LIFR activation promotes cell proliferation and EMT in MKN45 cells. Relative mRNA expression **(A)** LIFR and **(B)** LIF in CG cell lines. **(C)** IHC staining of LIFR in MNK45 cell lines on the left untreated and on the right triggered with LIF (10 ng/ml; magnification, ×100). MKN45 cells were serum-starved and primed with LIF (0.5, 5, and 10 ng/ml). Data shown are as follows: **(D)** dose–response curve of LIF (0.5, 5, and 10 ng/ml) determined using MTS assay on MKN45 cells. Each value is expressed relative to those of non-treated (NT), which are arbitrarily settled to 1. Results are the mean ± SEM of 10 samples per group. Relative mRNA expression of **(E)** the proliferation marker C-Myc and EMT markers **(F)** E-cadherin, **(G)** vimentin, and **(H)** Snal-1. Each value is normalized to Gapdh and is expressed relative to those of positive controls, which are arbitrarily settled to 1. Results are the mean ± SEM of five samples per group (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$).

transcription factor in the modulation of cancer progression (45). Despite the fact that, similar to a previous study, we have found that ONECUT2 gene expression was increased in the neoplastic tissues in comparison with paired samples obtained from non-neoplastic mucosa, we have found that reduced levels of ONECUT2, rather than its induction, are a poor prognosis predictor in patients with GC and peritoneal disease (39). The reason for this discrepancy is unclear, because overexpression of ONECUT2 in MKN54 and AGS, two GC cell lines, promotes cell proliferation and migration. However, others have reported that ONECUT2 regulation occurs through epigenetic regulation and hypomethylation of that CpGs in the promoter of ONECUT2, and this regulation occurs primarily in promoting intestinal differentiation of gastric mucosa. Accordingly, it has suggested that tissues levels of ONECUT2, gene and protein, might have utility in detecting intestinal metaplasia and might represent a biomarker of initial stages of gastric carcinogenesis. In contrast, the role of ONECUT2 in advanced disease is less defined (46).

The formation of peritoneal metastasis in GC is a multistep process, whereby cancer cells detach from primary tumor, migrate and attach to distant peritoneum, followed by invasion into sub-peritoneal tissues and cell proliferation to form detectable metastasis (47). Despite the clinical relevance, the specific molecular mechanisms that drive the formation of peritoneal metastasis in GC remain poorly understood, although previous studies using paired samples of primary and metastatic tumors have identified several putative mediators, mostly related to EMT remodeling, cell motility, and cytoskeleton rearrangement (48). Here, we report that the development of peritoneal disease in our series of patients with GC is associated with a robust upregulation of the LIFR in the primary tumors. This finding prompted us to further investigate whether the LIF/LIFR system was involved in promoting the EMT phenotype, a process that involves a deep reprogramming of the cancer cells genes. LIFR is an heterodimeric membrane receptor complex composed by LIFRβ and GP130 (49, 50), and although the receptor lacks an intrinsic tyrosine kinase activity, LIFR/GP130 complex constitutively associates with JAK-Tyk family of cytoplasmic tyrosine kinases, which facilitates downstream signaling and STAT3 phosphorylation activation.

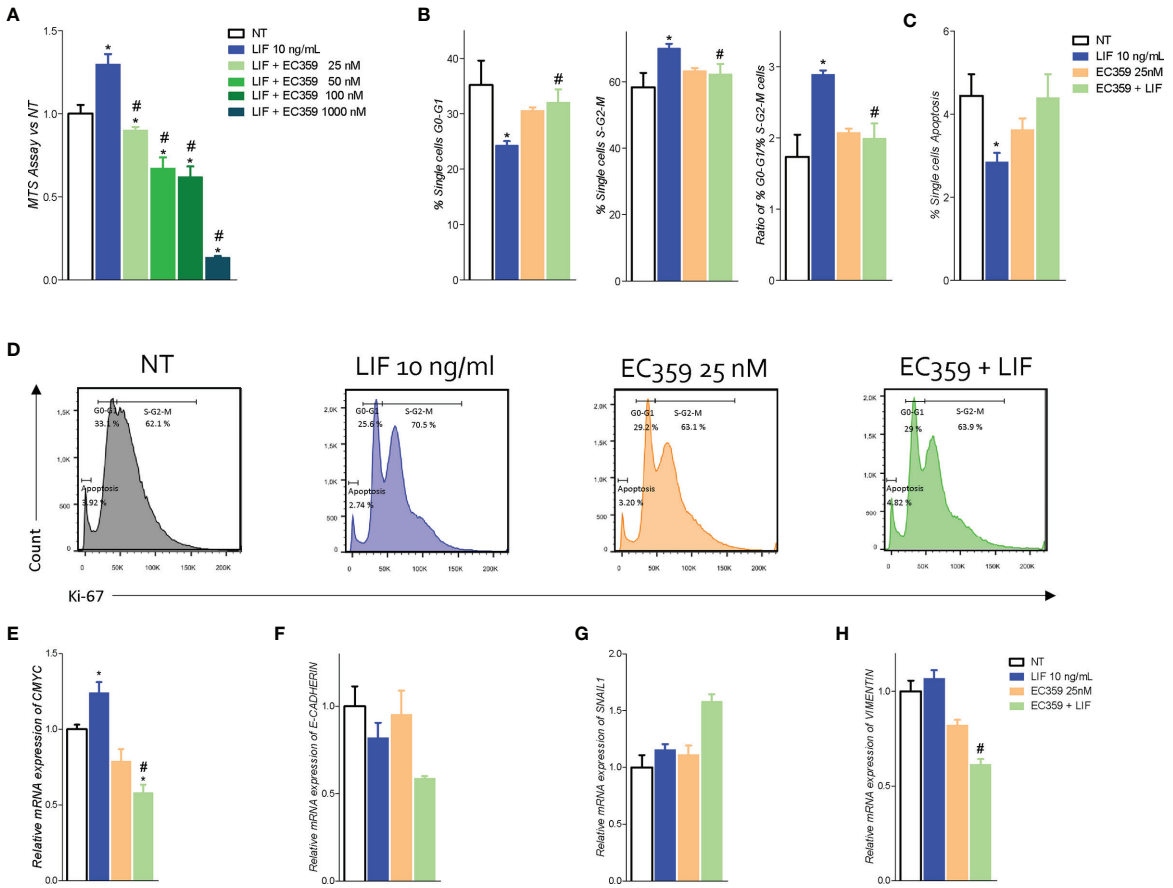


FIGURE 6 | LIFR antagonist EC359 hinders cell cycle progression, increases apoptosis rate in MKN45 cells and inhibits EMT process. **(A)** Dose–response curve of EC359 (25, 50, 100, and 1,000 nM) determined using MTS assay on MKN45 cells (n = 10). MKN45 cells were serum-starved and triggered with LIF (10 ng/ml), EC359 (100 nM), and LIF + EC359 for 48 h. Cell cycle phase analysis was performed by Ki-67/DAPI staining through IC-FACS. Data shown are follows: percentage of **(B)** from left to right cell in G0-G1 cell cycle phases, S-G2-M cell cycle phases, and ratio between % G0-G1 and % S-G2-M. **(C)** Percentage of apoptotic cells. **(D)** Representative IC-FACS showed cell cycle fraction and apoptosis rate in NT, LIF (10 ng/ml), EC359 (25 nM), and LIF + EC359. Results are the mean ± SEM of three samples for group (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$). Relative mRNA expression of **(E)** the proliferation marker C-Myc and EMT markers **(F)** E-Cadherin, **(G)** Snal-1, and **(H)** vimentin. Each value is normalized to Gapdh and is expressed relative to those of positive controls, which are arbitrarily settled to 1. Results are the mean ± SEM of five samples per group (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$).

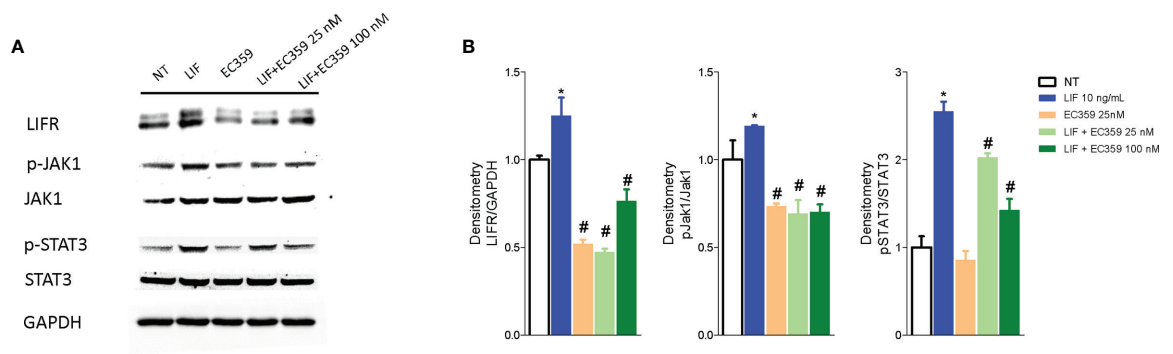


FIGURE 7 | Analysis of JAK-STAT signaling pathway. Representative Western blot analysis of **(A)** LIFR, JAK1 and phospho-JAK1, STAT3 and phospho-STAT3, and proteins in MKN45 exposed to LIF (10 nM) alone or in combination with EC359 (25 nM and 100 nM) for 20 min. GAPDH was used as loading control. **(B)** Densitometric analysis demonstrating LIFR expression, phospho-JAK1/JAK1, and phospho-STAT3/STAT3 ratio. The blot shown is representative of another one showing the same pattern. (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$).

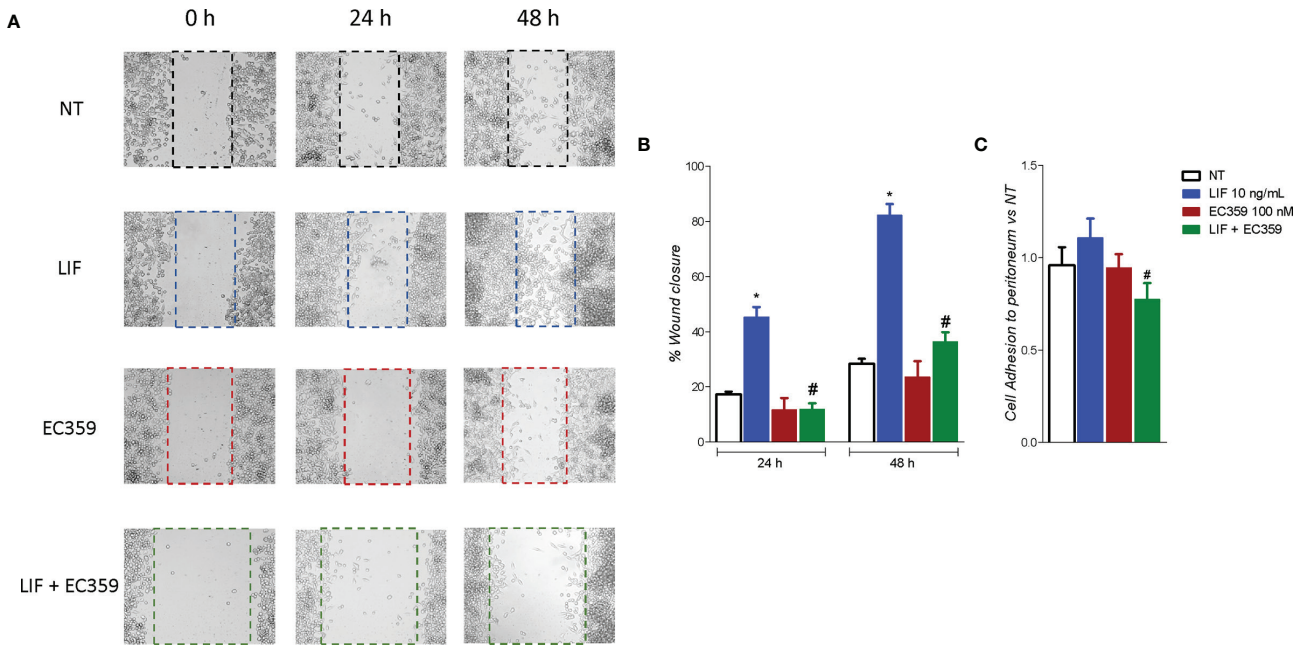


FIGURE 8 | LIFR antagonism inhibits MKN45. **(A, B)** A scratch wound healing assay is shown. MKN45 cell monolayers were scraped in a straight line using a p200 pipette tip to create a "scratch", and then, they are left untreated or primed with LIF (10 ng/ml), EC359 (100 nM), and LIF + EC359. The wound generated was imaged at 0, 24, and 48 h of incubation with the indicated compounds. **(A)** The images show cell migration at the three times point indicated. **(B)** Images obtained points were analysed measuring scraped area and its closure vs. the first time point at 0 h. Results are the mean ± SEM of three samples per group. **(C)** Cell adhesion to peritoneum. Experiment was conducted in quintuplicate (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$).

Several tumors exhibit upregulated JAK/STAT, ERK/MAPK, and PI3K/AKT signaling *via* autocrine or paracrine activation of LIF/LIFR GP130, and this pathway significantly contributes to EMT in several cancers, disease progression, and a poorer relapse-free survival in several cancers (51). LIF also participates to cross-talk between tumor cells and matrix fibroblasts to mediate the pro-invasive activation of stromal fibroblasts (52) and promotes drug resistance to HDAC inhibitors (53). By using GC cell lines, we have shown that LIFR expression varies from one line to another and that MKN45 cells were the cells with the highest expression. Challenging these cells with LIF promotes the acquisition of migratory phenotype, and this is associated with the acquisition of molecular signature of EMT and these changes are associated with LIF/LIFR signaling as assessed by measuring JAK and STAT3 phosphorylation. Of relevance, the LIFR inhibitor, EC359 (IC_{50} values of 10.2 nM), reversed these changed. Cotreating MKN45 cells with EC359 also reversed JAK and STAT3 phosphorylation induced by exposure of MKN45 cells to LIF, as well as regulation of E-cadherin and vimentin in a concentration-dependent manner, further confirming the role of LIF/LIFR in promoting EMT as well as acquisition of migratory phenotype of GC cells.

Cytoskeletal remodeling is closely related with tumor migration, invasion, and metastasis (54). LIFR plays an essential role in regulation of actin filament dynamics by modulating the expression of vimentin. Consistent with this background, we demonstrate that pharmacological inhibition of LIFR negatively regulates the expression of vimentin and that this effects are associated with a reduce cell motility and impaired migration (55, 56). Vimentin plays an important role in tumor invasion and metastasis (57), and its counter-regulation is a further evidence of the role that LIF/LIFR signaling plays in the modulation of EMT process. Of relevance, LIFR activation positively regulates vimentin expression and downregulates E-cadherin *via* JAK and STAT3 phosphorylation, and LIFR antagonism reversed this pathway (55–57).

In conclusion, by NGS RNA-seq analysis, we have identified LIF/LIFR pathway as an important mechanism in disease progression in GC. High levels of expression of LIFR mRNA in tumor tissues predict poor prognosis and reduced response to therapy. In addition, by using GC cell lines, we have shown that LIFR activation results in JAK STAT3 phosphorylation and EMT as demonstrated by vimentin induction and blunted expression of E-cadherin. These molecular changes are associated with a migratory phenotype of GC cell lines and are reversed by LIF/LIFR antagonism. Together, we suggest that targeting LIF/LIFR signaling might have utility in management of GC.

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 at https://data.mendeley.com/drafts/9t86hd78sj, 10.17632/9t86hd78sj.1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by FI00001 no. 2266/2014, granted on February 19, 2014, and by FIO0003 no. 36348 granted on May 6, 2020. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SF, LG, and AD contributed to conception and design of the study. AZ and SF provided research funding. LG, EM, and AD provided human samples. CD, SM, EM, and MB performed the data analysis. CD, SM, and EM performed the statistical analysis. SF, CD, SM, EM, and ED wrote the manuscript. CD, SM, RR, MB, RB, and GU contributed experimental settings. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by grant from the Italian MIUR/PRIN 2017 (2017FJZZRC).

ACKNOWLEDGMENTS

This manuscript has been released as a PrePrint (bioRxiv doi: https://doi.org/10.1101/2022.05.05.490785).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.939969/full#supplementary-material

Supplementary Figure 1 | MKN45 cells were serum-starved and left untreated or primed with LIF (10,50,100 ng/ml). Data shown are: **(A)** MTS assay. Each value is expressed relative to those of non-treated (NT), which are arbitrarily settled to 1. Results are the mean ± SEM of 10 samples per group. Relative mRNA expression of **(B)** the proliferation marker C-myc and EMT markers E-cadherin, Snal-1, and vimentin. Each value is normalized to Gapdh and is expressed relative to those of NT, which are arbitrarily settled to 1. Results are the mean ± SEM of five samples per group (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$). Cell cycle phase analysis were performed by Ki-67/DAPI staining through IC-FACS analysis. Data shown are: percentage of **(D)** Representative IC-FACS showed cell cycle fraction and apoptosis rate in NT, LIF (10 ng/ml), EC359 (25 nM), and LIF + EC359. **(E)** cell in G0-G1 cell cycle phases, S-G2-M cell cycle phases, and ratio between % G0-G1 and % S-G2-M. **(F)** Percentage of Apoptotic cells. Results are the mean ± SEM of three samples per group (* represents statistical significance versus NT, and # versus LIF, $p < 0.05$).

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
2. 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.
3. Ahmad SA, Xia BT, Bailey CE, Abbott DE, Helmink BA, Daly MC, et al. An Update on Gastric Cancer. *Curr Probl Surg* (2016) 53(10):449–90.
4. González CA, Sala N, Rokkas T. Gastric Cancer: Epidemiologic Aspects. *Helicobacter* (2013) 18 Suppl 1:34–8.
5. Di Giorgio C, Roselli R, Biagioli M, Marchianò S, Distrutti E, Bordoni M, et al. Organoids as *Ex Vivo* Culture System to Investigate Infection-Host Interaction in Gastric Pre-Carcinogenesis. *Recent Adv Inflammation Allergy Drug Discovery* (2022). doi: 10.2174/2772270816666220105123702
6. Lauren P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand* (1965) 64:31–49.
7. Carino A, Graziosi L, Marchianò S, Biagioli M, Marino E, Sepe V, et al. Analysis of Gastric Cancer Transcriptome Allows the Identification of Histotype Specific Molecular Signatures With Prognostic Potential. *Front Oncol* (2021) 11:663771.
8. Chia N-Y, Tan P. Molecular Classification of Gastric Cancer. *Ann Oncol Off J Eur Soc Med Oncol* (2016) 27(5):763–9.
9. Tan P, Yeoh K-GG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* (2015) 149(5):1153–62.e3.
10. Lutz MP, Zalcberg JR, Ducreux M, Adenis A, Allum W, Aust D, et al. The 4th St. Gallen EORTC Gastrointestinal Cancer Conference: Controversial Issues in the Multimodal Primary Treatment of Gastric, Junctional and Oesophageal Adenocarcinoma. *Eur J Cancer* (2019) 112:1–8.
11. Lei Z, Tan IB, Das K, Deng N, Zouridis H, Pattison S, et al. Identification of Molecular Subtypes of Gastric Cancer With Different Responses to PI3-Kinase Inhibitors and 5-Fluorouracil. *Gastroenterology* (2013) 145 (3):554–65.
12. Network CGAR. Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* (2014) 513(7517):202–9.
13. Cristescu R, Lee JH, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular Analysis of Gastric Cancer Identifies Subtypes Associated With Distinct Clinical Outcomes. *Nat Med* (2015) 21(5):449–56.
14. Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric Adenocarcinoma. *Nat Rev Dis Prim* (2017) 3:17036.
15. Schmidt B, Yoon SS. D1 Versus D2 Lymphadenectomy for Gastric Cancer. *J Surg Oncol* (2013) 107(3):259–64.
16. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-Metastatic Niches: Organ-Specific Homes for Metastases. *Nat Rev Cancer* (2017) 17(5):302–17.
17. Liu D, Li C, Trojanowicz B, Li X, Shi D, Zhan C, et al. CD97 Promotion of Gastric Carcinoma Lymphatic Metastasis is Exosome Dependent. *Gastric Cancer Off J Int Gastric Cancer Assoc Japanese Gastric Cancer Assoc* (2016) 19 (3):754–66.
18. Kim MA, Lee HS, Lee HE, Kim JH, Yang H-K, Kim WH. Prognostic Importance of Epithelial-Mesenchymal Transition-Related Protein Expression in Gastric Carcinoma. *Histopathology* (2009) 54(4):442–51.
19. Ye X, Weinberg RA. Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. *Trends Cell Biol* (2015) 25(11):675–86.
20. Marano L, Marrelli D, Sammartino P, Biacchi D, Graziosi L, Marino E, et al. Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy for Gastric Cancer With Synchronous Peritoneal Metastases: Multicenter Study of "Italian Peritoneal Surface Malignancies Oncoteam-S.I.C.O". *Ann Surg Oncol* (2021) 28(13):9060–70.
21. Thomassen I, Bernards N, van Gestel YR, Creemers G-J, Jacobs EM, Lemmens VE, et al. Chemotherapy as Palliative Treatment for Peritoneal Carcinomatosis of Gastric Origin. *Acta Oncol (Stockholm Sweden) Engl* (2014) 53:429–32.
22. Marano L, Polom K, Patriti A, Roviello G, Falco G, Stracqualursi A, et al. Surgical Management of Advanced Gastric Cancer: An Evolving Issue. *Eur J Surg Oncol* (2016) 42(1):18–27.
23. Shiozaki H, Elimova E, Slack RS, Chen HC, Staerkel GA, Sneige N, et al. Prognosis of Gastric Adenocarcinoma Patients With Various Burdens of Peritoneal Metastases. *J Surg Oncol* (2016) 113(1):29–35.

24. Seeneevassen L, Martin OCB, Lehours P, Dubus P, Varon C. Leukaemia Inhibitory Factor in Gastric Cancer: Friend or Foe? *Gastric Cancer Off J Int Gastric Cancer Assoc Japanese Gastric Cancer Assoc* (2022) 25(2):299–305.
25. Pinho V, Fernandes M, da Costa A, Machado R, Gomes AC. Leukemia Inhibitory Factor: Recent Advances and Implications in Biotechnology. *Cytokine Growth Factor Rev* (2020) 52:25–33.
26. Wang M-T, Fer N, Galeas J, Collisson EA, Kim SE, Sharib J, et al. Blockade of Leukemia Inhibitory Factor as a Therapeutic Approach to KRAS Driven Pancreatic Cancer. *Nat Commun* (2019) 10(1):3055.
27. Yu H, Yue X, Zhao Y, Li X, Wu L, Zhang C, et al. LIF Negatively Regulates Tumour-Suppressor P53 Through Stat3/ID1/MDM2 in Colorectal Cancers. *Nat Commun* (2014) 5:5218.
28. Wang J, Xie C, Pan S, Liang Y, Han J, Lan Y, et al. N-Myc Downstream-Regulated Gene 2 Inhibits Human Cholangiocarcinoma Progression and is Regulated by Leukemia Inhibitory Factor/MicroRNA-181c Negative Feedback Pathway. *Hepatology* (2016) 64(5):1606–22.
29. Li X, Yang Q, Yu H, Wu L, Zhao Y, Zhang C, et al. LIF Promotes Tumorigenesis and Metastasis of Breast Cancer Through the AKT-mTOR Pathway. *Oncotarget* (2014) 5(3):788–801.
30. Nicola NA, Babon JJ. Leukemia Inhibitory Factor (LIF). *Cytokine Growth Factor Rev* (2015) 26(5):533–44.
31. Viswanadhapalli S, Luo Y, Sareddy GR, Santhamma B, Zhou M, Li M, et al. EC359: A First-In-Class Small-Molecule Inhibitor for Targeting Oncogenic LIFR Signaling in Triple-Negative Breast Cancer. *Mol Cancer Ther* (2019) 18 (8):1341–54.
32. Zhao X, Ye F, Chen L, Lu W, Xie X. Human Epithelial Ovarian Carcinoma Cell-Derived Cytokines Cooperatively Induce Activated CD4+CD25-CD45RA+ Naïve T Cells to Express Forkhead Box Protein 3 and Exhibit Suppressive Ability *In Vitro*. *Cancer Sci* (2009) 100(11):2143–51.
33. Pascual-García M, Bonfill-Teixidor E, Planas-Rigol E, Rubio-Perez C, Iurlaro R, Arias A, et al. LIF Regulates CXCL9 in Tumor-Associated Macrophages and Prevents CD8(+) T Cell Tumor-Infiltration Impairing Anti-PD1 Therapy. *Nat Commun* (2019) 10(1):2416.
34. Lin S-R, Wen Y-C, Yeh H-L, Jiang K-C, Chen W-H, Mokgautsi N, et al. EGFR-Upregulated LIFR Promotes SUCLG2-Dependent Castration Resistance and Neuroendocrine Differentiation of Prostate Cancer. *Oncogene* (2020) 39(44):6757–75.
35. Bian S-B, Yang Y, Liang W-Q, Zhang K-C, Chen L, Zhang Z-T. Leukemia Inhibitory Factor Promotes Gastric Cancer Cell Pro LIFeration, Migration, and Invasion *via* the LIFR-Hippo-YAP Pathway. *Ann N Y Acad Sci* (2021) 1484(1):74–89.
36. Seeneevassen L, Giraud J, Molina-Castro S, Sifré E, Tiffon C, Beauvoit C, et al. Leukaemia Inhibitory Factor (LIF) Inhibits Cancer Stem Cells Tumorigenic Properties Through Hippo Kinases Activation in Gastric Cancer. *Cancers (Basel)* (2020) 12(8):2011. doi: 10.3390/cancers12082011
37. Chen Y. Scratch Wound Healing Assay. *Bio-protocol* (2012) 2(5):e100. doi: 10.21769/BioProtoc.100
38. Asao T, Yazawa S, Kudo S, Takenoshita S, Nagamachi Y. A Novel Ex Vivo Method for Assaying Adhesion of Cancer Cells to the Peritoneum. *Cancer Lett* (1994) 78(1–3):57–62. doi: 10.1016/0304-3835(94)90031-0
39. Chen J, Chen J, Sun B, Wu J, Du C. ONECUT2 Accelerates Tumor Proliferation Through Activating ROCK1 Expression in Gastric Cancer. *Cancer Manag Res* (2020) 12:6113–21.
40. Huang L, Wu R-L, Xu A-M. Epithelial-Mesenchymal Transition in Gastric Cancer. *Am J Transl Res* (2015) 7(11):2141–58.
41. Thomassen I, van Gestel YR, van Ramshorst B, Luyer MD, Bosscha K, Nienhuijs SW, et al. Peritoneal Carcinomatosis of Gastric Origin: A Population-Based Study on Incidence, Survival and Risk Factors. *Int J Cancer* (2014) 134(3):622–8.
42. Sugarbaker PH, Yu W, Yonemura Y. Gastrectomy, Peritonectomy, and Perioperative Intraperitoneal Chemotherapy: The Evolution of Treatment Strategies for Advanced Gastric Cancer. *Semin Surg Oncol* (2003) 21(4):233–48.
43. Wadhwa R, Song S, Lee J-S, Yao Y, Wei Q, Ajani JA. Gastric Cancer-Molecular and Clinical Dimensions. *Nat Rev Clin Oncol* (2013) 10(11):643–55.
44. Vanhorenbeeck V, Jenny M, Cornut J-F, Gradwohl G, Lemaigre FP, Rousseau GG, et al. Role of the Onecut Transcription Factors in Pancreas Morphogenesis and in Pancreatic and Enteric Endocrine Differentiation. *Dev Biol* (2007) 305(2):685–94.
45. Yu J, Li D, Jiang H. Emerging Role of ONECUT2 in Tumors. *Oncol Lett* (2020) 20(6):328.
46. Seo E-H, Kim H-J, Kim J-H, Lim B, Park J-L, Kim S-Y, et al. ONECUT2 Upregulation Is Associated With CpG Hypomethylation at Promoter-Proximal DNA in Gastric Cancer and Triggers ACSL5. *Int J Cancer* (2020) 146(12):3354–68.
47. Sun F, Feng M, Guan W. Mechanisms of Peritoneal Dissemination in Gastric Cancer. *Oncol Lett* (2017) 14(6):6991–8.
48. Kang X, Li W, Liu W, Liang H, Deng J, Wong CC, et al. LIMK1 Promotes Peritoneal Metastasis of Gastric Cancer and is a Therapeutic Target. *Oncogene* (2021) 40(19):3422–33.
49. Metcalf D. The Unsolved Enigmas of Leukemia Inhibitory Factor. *Stem Cells* (2003) 21(1):5–14.
50. Stahl N, Boulton TG, Farruggella T, Ip NY, Davis S, Witthuhn BA, et al. Association and Activation of Jak-Tyk Kinases by CNTF- LIF-OSM-IL-6 Beta Receptor Components. *Science* (1994) 263(5143):92–5.
51. Yue X, Zhao Y, Zhang C, Li J, Liu Z, Liu J, et al. Leukemia Inhibitory Factor Promotes EMT Through STAT3-Dependent miR-21 Induction. *Oncotarget* (2016) 7(4):3777–90.
52. Albrengues J, Bourget I, Pons C, Butet V, Hofman P, Tartare-Deckert S, et al. LIF Mediates Proinvasive Activation of Stromal Fibroblasts in Cancer. *Cell Rep* (2014) 7(5):1664–78.
53. Zeng H, Qu J, Jin N, Xu J, Lin C, Chen Y, et al. Feedback Activation of Leukemia Inhibitory Factor Receptor Limits Response to Histone Deacetylase Inhibitors in Breast Cancer. *Cancer Cell* (2016) 30(3):459–73.
54. Fife CM, McCarroll JA, Kavallaris M. Movers and Shakers: Cell Cytoskeleton in Cancer Metastasis. *Br J Pharmacol* (2014) 171(24):5507–23.
55. Mei J-W, Yang Z-Y, Xiang H-G, Bao R, Ye Y-Y, Ren T, et al. MicroRNA-1275 Inhibits Cell Migration and Invasion in Gastric Cancer by Regulating Vimentin and E-Cadherin *via* JAZF1. *BMC Cancer* (2019) 19(1):740.
56. Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, et al. Relation Between microRNA Expression and Progression and Prognosis of Gastric Cancer: A microRNA Expression Analysis. *Lancet Oncol* (2010) 11(2):136–46.
57. Du L, Li J, Lei L, He H, Chen E, Dong J, et al. High Vimentin Expression Predicts a Poor Prognosis and Progression in Colorectal Cancer: A Study With Meta-Analysis and TCGA Database. *BioMed Res Int* (2018) 2018:6387810. doi: 10.1155/2018/6387810

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

frontiers | Frontiers in Oncology

CASE REPORT
published: 07 July 2022
doi: 10.3389/fonc.2022.925582

Co-occurrence of *VHL* and *SDHA* Pathogenic Variants: A Case Report

Moon Ley Tung[1†], *Bharatendu Chandra*[1†], *Kyle Dillahunt*[1], *Matthew D. Gosse*[2], *T. Shawn Sato*[3] *and Alpa Sidhu*[1*]

[1] Division of Medical Genetics and Genomics, The Stead Family Department of Pediatrics, University of Iowa Hospitals and Clinics, Iowa City, IA, United States, [2] Department of Pathology, University of Iowa Hospitals and Clinics, Iowa City, IA, United States, [3] Division of Pediatric Radiology, The Stead Family Children's Hospital, University of Iowa Hospitals and Clinics, Iowa City, IA, United States

Von Hippel Lindau(VHL)syndrome presents with cerebellar and spinal hemangioblastomas, renal cell cancer, neuroendocrine pancreatic tumor, and pheochromocytoma and it is caused by germline mutations in the *VHL* gene. Pathogenic germline variants in the succinate dehydrogenase A (*SDHA*) gene are associated with paraganglioma and pheochromocytoma. Here we report co-occurrence of germline pathogenic variants in both *VHL* and *SDHA* genes in a patient who presented with pancreatic neuroendocrine tumor. As these genes converge on the pseudo-hypoxia signaling pathway, further studies are warranted to determine the significance of co-occurrence of these variants in relation to tumor penetrance, disease severity, treatment response and clinical outcomes in this selected group of patients.

Keywords: von Hippel-Lindau syndrome, *SDHA*-associated paraganglioma and pheochromocytoma syndrome, paraganglioma, genetics, cancer

OPEN ACCESS

Edited by:
Israel Gomy, Oncoclinicas Group, Brazil

Reviewed by:
Raymond H Kim, University of Toronto, Canada
Andrea Benedetto Galosi, Marche Polytechnic University, Italy

***Correspondence:**
Alpa Sidhu
alpa-sidhu@uiowa.edu

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

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

Received: *21 April 2022*
Accepted: *10 June 2022*
Published: *07 July 2022*

Citation:
Tung ML, Chandra B, Dillahunt K, Gosse MD, Sato TS and Sidhu A (2022) Co-occurrence of VHL and SDHA Pathogenic Variants: A Case Report. Front. Oncol. 12:925582. doi: 10.3389/fonc.2022.925582

INTRODUCTION

Von Hippel Lindau (VHL) syndrome caused by germline loss-of-function variants in the gene *VHL*, is an autosomal dominant cancer predisposition syndrome (1). Clinical features include cerebellar and spinal hemangioblastomas, renal cell cancer, neuroendocrine pancreatic tumor, pheochromocytoma, and paragangliomas. Germline pathogenic variants in the succinate dehydrogenase A (*SDHA*) gene are associated with familial paraganglioma and pheochromocytoma syndrome inherited in an autosomal dominant manner (2). At the cellular level, SDHA and VHL proteins interact and converge into a common molecular pathway *via* the hypoxia inducible factor alpha (HIF-α) (3). Here we report co-occurrence of germline pathogenic variants in both genes (*VHL* and *SDHA*) in a patient who presented with pancreatic neuroendocrine tumor. We describe the known molecular pathways involving VHL and SDHA and postulate that the disease prognosis may be dependent on the presence of co-occurring pathogenic variants in these genes through the involvement of HIF-α in the final common pathway. To our knowledge, this is the first case report of both *VHL* and *SDHA* pathogenic variants.

CASE PRESENTATION

The proband, a 23-year-old female, was reviewed at our Genetics Cancer Predisposition clinic for evaluation and recommendations for concurrent germline pathogenic variants in the *VHL* and *SDHA* genes. She presented with recurrent episodes of abdominal pain, vomiting, and anorexia at 21-years of age. Initial radiological and endoscopic evaluations were unremarkable. Her symptoms continued to persist, and a review of her prior abdominal computed tomography (**Figures 1A, B**) and magnetic resonance imaging (MRI) study showed a mass in the pancreatic tail with hepatic metastases (**Figures 1C–F**). She underwent surgical resection of pancreatic tail mass as well as right liver lobectomy, splenectomy, and cholecystectomy for metastatic disease.

Histopathologic examination of the pancreatic tail mass revealed a well-differentiated neuroendocrine tumor (**Figure 2A**). The primary tumor was centered in the distal pancreas, perineural invasion was present, and 4 mitoses per 2 mm^2 were identified. Metastatic tumor deposits were identified in two of fifteen regional lymph nodes and multiple metastatic liver foci were identified (**Figure 2B**). A Ki-67 immunostain was performed on the primary tumor, local lymph node metastasis, and liver metastasis (**Figure 2C**) and the proliferation index was determined to be 14%, 8.5%, and 11%, respectively, using digital image analysis. This met criteria for a Grade 2 (intermediate) tumor by 2019 WHO classification using both mitotic count and Ki-67 proliferation index. Immunohistochemistry on the primary tumor was positive for synaptophysin, chromogranin, pan keratin, and CK7. and negative for CK20, CD10 and progesterone receptor (PR). Somatostatin receptor subtype 2A (SSTR2A) immunostaining was positive (3+, 90%) (**Figure 2D**) and ATRX immunostain was intact. Initial testing of tumor was performed at an outside institution and additional pathology samples were not available for SDHB immunohistochemistry.

Molecular test performed on the tumor sample showed pathogenic variants in the *VHL* gene (c.500G>A, p.R167Q) (**Figure 2E**) involving 66% of tumor cells, and *SDHA* gene (c.1054C>T, p.R352*) (**Figure 2F**) involving 50% of tumor cells. Germline testing on peripheral blood sample showed identical pathogenic variants in *VHL* (c.500G>A, p.R167Q) (**Figure 2G**) and *SDHA* (c.1054C>T, p.R352*) genes (**Figure 2H**). Familial testing could not be performed due to patient's adoptive status. A Positron Emission Tomography scan following the surgery was suggestive of multiple somatostatin receptor tracer uptake in the pancreatic head, duodenum, and the liver. She was initially treated with octreotide acetate and subsequently with Lantreotide.

The patient's past medical history was significant for a right retinal hemangioma diagnosed at 18- years of age. Follow-up ophthalmological evaluation showed consistent exam without progression of disease. She was making good recovery from her initial diagnosis at the time of our review, and the pancreatic and hepatic lesions have remained stable. She was recommended studies per surveillance guidelines for her diagnosis of both *VHL* and *SDHA*-related paraganglioma and pheochromocytoma syndrome. She had normal brain, internal auditory canal, and spine magnetic resonance imaging studies without evidence of any VHL-associated lesions. There is limited prenatal, birth, and postnatal history available as she was adopted during infancy, and presented with early onset global developmental delay,undergoing interventional therapies during childhood. She was diagnosed with autism spectrum disorder and depression. To rule out underlying genetic etiology of global developmental delay, we performed chromosomal microarray and Fragile-X syndrome analysis. Her chromosomal microarray was normal whereas Fragile- X syndrome analysis incidentally revealed a premutation carrier status with 30 and 57 CGG repeats in the *FMR1* gene. The patient and her family were provided comprehensive genetic counseling about these results.

DISCUSSION

Von Hippel-Lindau (VHL) is an autosomal dominant hereditary tumor predisposition syndrome caused due to a germline pathogenic variant in the *VHL* gene (4). VHL syndrome predisposes an individual to various tumors such as cerebellar and spinal hemangioblastomas, retinal angiomas, renal cell carcinoma, and pheochromocytoma (1, 5). Other tumors such as pancreatic neoplasms, pituitary hemangioblastomas, and duodenal carcinoid tumors have also been rarely reported. The disorder has a high disease penetrance that is estimated to be about 97% by the age of 60 years (6). Patients with VHL syndrome have a shortened life expectancy secondary to complications related to cerebellar hemangioblastoma, renal cell carcinoma, and pancreatic neoplasms (1). Although VHL-associated tumors usually manifest at a younger age compared to sporadic tumors (6), they appear to be more responsive to chemotherapeutics (7), and less aggressive with respect to their local recurrence and metastatic involvement (8). Present surveillance guidelines recommend age-based screening with dilated eye examination, plasma metanephrines, MRI of brain, spine, abdomen, and internal auditory canal.

Collectively, hereditary paraganglioma-pheochromocytoma (PGL/PCC) syndromes are rare neuroendocrine tumors with an estimated incidence of approximately 2-8 cases per million per year (9). Approximately 40% of all cases of PGL/PCC are associated with germline pathogenic variants (2, 10, 11) in the pseudo-hypoxic signaling pathway (cluster I), kinase signaling (cluster II), or Wingless and Int-1 (Wnt) signaling group (cluster III) (**Figure 3**) (12). Pathogenic variants in the SDHx genes (*SDHA, SDHB, SDHC*, and *SDHD*) are classified under the cluster I genes which results in dysfunction of succinate dehydrogenase (SDH) leading to competitive inhibition of the enzyme, prolyl hydroxylase, involved in the degradation of hypoxia-inducible factor 1- α (HIF1- α) (2, 13, 14). SDHx-related PGL/PCC are relatively new tumor predisposition syndromes that include PGL and PCC, and rarely renal cell carcinoma, pituitary adenoma, gastrointestinal stromal tumors, and pancreatic neuroendocrine tumor (PNET) (15). Patients with PGL/PCC can manifest tumor at any age, although a majority would present between the third and fifth decade of

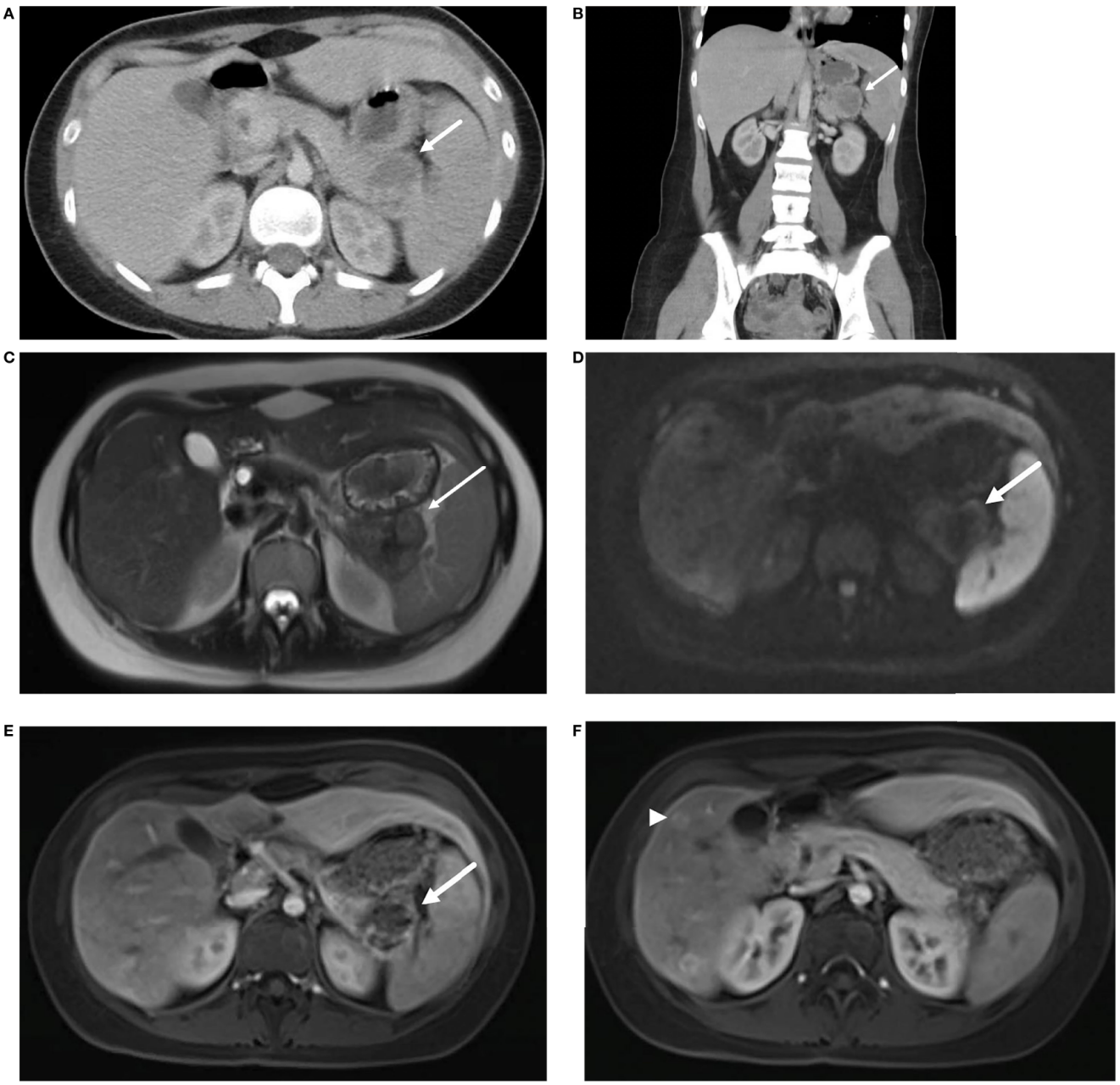


FIGURE 1 | Axial **(A)** and coronal **(B)** contrast enhanced CT demonstrate a pancreatic tail mass with central necrosis (→) positioned between the stomach, spleen, and kidney. MRI of the abdomen reveals a mass in the tail of the pancreas (→) with intermediate T2 signal **(C)**, diffusion restriction **(D)** and peripheral contrast enhancement **(E)**. Additional enhancing lesions (▶) were seen in the liver consistent with metastatic disease **(F)**.

life (16) with symptoms of excess catecholamine production, including hypertension, headache, diaphoresis, palpitations, and tremors. *SDHA* pathogenic germline variants are present in up to 30% of wild-type gastrointestinal stromal tumors and 10% of patients with PGL and PCC (Boikos et al, 2016). A recent study identified only 10 patients harboring a pathogenic germline *SDHA* variant in 4,974 pediatric and adult patients with a variety of solid tumors (17). Out of these 10 patients with underlying *SDHA* pathogenic variant, 2 had gastrointestinal stromal tumor, and the remaining had melanoma, neuroblastoma, breast, colon, renal, prostate, endometrial, and bladder cancer.

In contrast to *SDHB* which is a highly penetrant tumor predisposition gene, *SDHA* confers a much lower penetrance and severity, estimates of which are largely unknown (18, 19). Present surveillance guidelines for SDHx-associated PGL/PCC

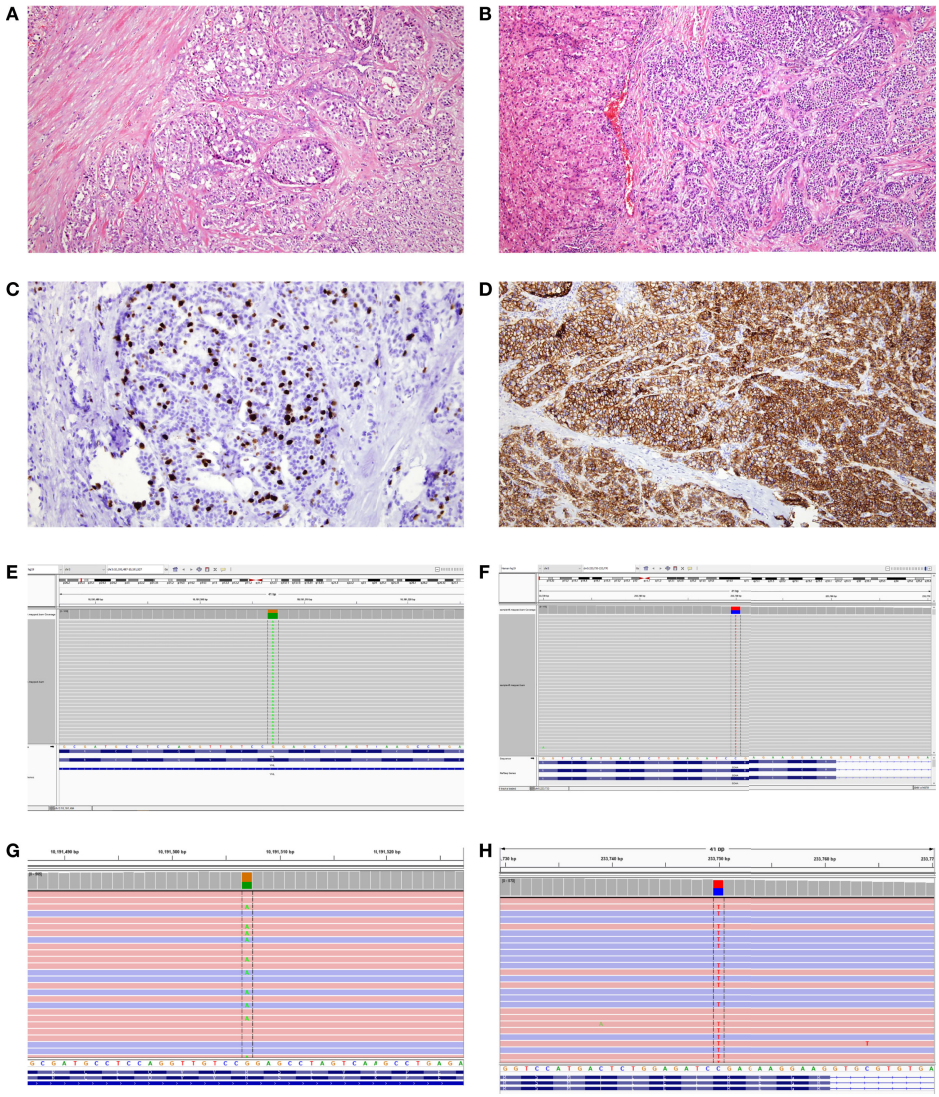


FIGURE 2 | **(A)** Hemotoxylin and eosin (H&E) staining of primary tumor. Cells are arranged in nests with finely stippled chromatin. The tumor cells expressed synaptophysin, chromogranin, and pankeratin (not shown) supporting the diagnosis of well-differentiated neuroendocrine tumor. **(B)** H&E stained section of metastatic tumor focus in the liver with similar morphology to the primary tumor. **(C)** Ki-67 immunostained section of the metastatic tumor focus in the liver. The proliferation index was found to be 11% in a similar hot spot by digital image analysis. **(D)** Somatostatin receptor subtype 2A (SSTR2A) immunostained section was positive (3+, 90%). **(E)** Tumor molecular profiling showing a variant in *VHL* (c.500G>A, p.R167Q) at a frequency of 66.3%). **(F)** Tumor molecular profiling showing a variant in *SDHA* (c.1054C>T, p.R352*) at a frequency of 50%). **(G)** Germline molecular profiling showing an identical variant in *VHL* (c.500G>A, p.R167Q). **(H)** Germline molecular profiling showing an identical variant in *SDHA* (c.1054C>T, p.R352*).

recommend plasma metanephrines, whole body and dedicated neck MRI, and complete blood count starting from age 6-8 years (19). Our patient represents the first case of a metastatic PNET in the setting of germline heterozygous pathogenic variants in the *VHL* and *SDHA* genes. PNETs are clinically heterogenous tumors originating from neuroendocrine cells of the pancreatic islets and usually follow a variable clinical course, with a low five-year survival rate in approximately 60% of patients (20). PNET occur in approximately 9 to 17% of patients with VHL disease (8) but have only been reported in one individual with a germline *SDHD* pathogenic variant (15). Since PNETs are seen in VHL and not observed so far in SDHA-related PGL/PCC syndrome, we postulate that the underlying genetic etiology would be the pathogenic variant with a second hit in the *VHL* gene. Although a second pathogenic variant was not reported by the tumor molecular testing, it is possible that the second hit could not be detected due to limitation of the tumor-based molecular analyses. These include possibility of somatic second hit being present in the promoter or deep enhancer region, or promoter methylation (21). Loss of heterozygosity (LOH) analysis could not be performedin the patient's tumor type, as the laboratory only performs LOH analysis in ovarian tumors

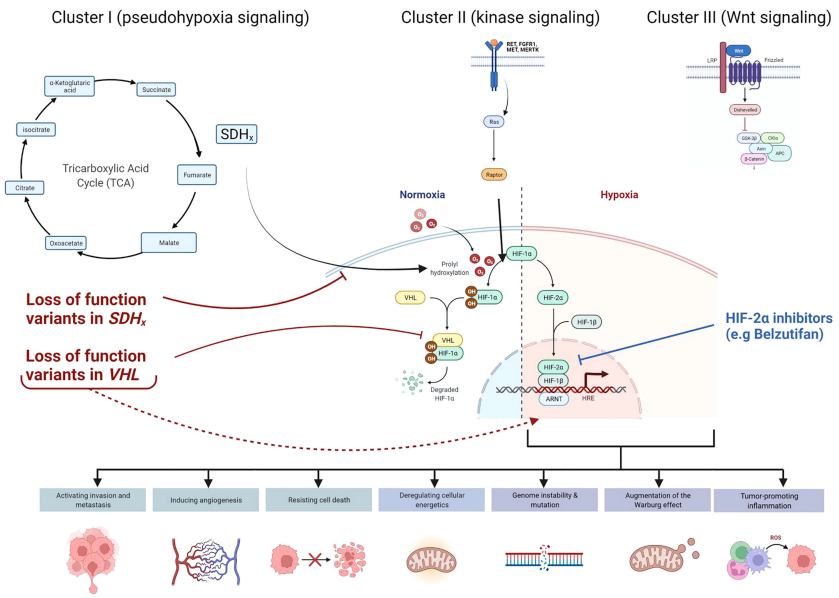


FIGURE 3 | Cellular pathways and genes involved in hereditary paraganglioma-pheochromocytoma (PGL/PCC) syndrome and Von Hippel Lindau (VHL) disease. Approximately 40% of all cases of PGL/PCC cases are associated with germline pathogenic variants in the pseudo-hypoxic signaling pathway (Cluster I), kinase signaling (Cluster II), or Wingless and Int-1 (Wnt) signaling (Cluster III). Pathogenic variants in the *SDHx* genes (*SDHA, SDHB, SDHC*, and *SDHD*) and *VHL* gene are classified under the Cluster I genes. The SDH complex plays an important role in energy metabolism through the tricarboxylic acid (TCA) cycle. Both *VHL* and *SDHx* encode for proteins that target the protein HIF1-α for degradation. Pathogenic loss of function variants in *VHL* and *SDHx* genes leads to abnormal accumulation of HIF1-α and upregulation of downstream pathways, which results in increased expression of various proteins and augmentation of the Warburg effect. These negative effects may be mitigated by novel therapies such as HIF2-α inhibitors (e.g Belzutifan).

Multi-locus inherited neoplasia allele syndrome (MINAS) is a relatively new entity that was initially reported in *BRCA1*/*BRCA2*-related cancers (21). With increasing adoption of next generation sequencing technologies for cancer susceptibility germline (CSG) testing, there has been a rise in reports of non-*BRCA1*/*BRCA2* related MINAS. A recent study by McGuigan et al. reported that atypical tumor phenotypes comprised of about 15% of non-*BRCA1*/*BRCA2* MINAS cases, which could be secondary to complex interactions between the relevant CSGs (21).

Interestingly, on a molecular level, both *VHL* and *SDHA* encode for proteins that target the protein HIF-α for degradation (22). Absence of these proteins would lead to abnormal accumulation and upregulation of HIF, resulting in increased expression of various proteins (e.g vascular endothelial growth factor, platelet-derived growth factor, matrix metalloproteinases and transforming growth factor-alpha), and augmentation of the Warburg effect by HIF-α. The Warburg effect relates to the phenomenon seen in tumor cells that effectively promotes growth and development. Even in normoxia, there is a metabolic shift in which tumor cells preferentially use glycolysis to generate adenosine triphosphate rather than the tricarboxylic acid (TCA) cycle (14). The SDH complex plays an important role in energy metabolism through the TCA cycle, and the reduction of ubiquinone to ubiquinol *via* the electron transport chain. Therefore, pathogenic variants in the genes involved in the SDH complex could result in accumulation of the 'oncometabolite' succinate, leading to disruption of the succinate to fumarate ratio that would inhibit the enzymatic degradation of HIF-α. Accumulation of succinate may also destabilize the redox state and cause mitochondrial dysfunction by increasing reactive oxygen species production and utilizing glutamine as an energy source (14). Epigenetic dysregulation has also been implicated in the malignant potential of SDHx-mutated PGL/PCC (3) and may be a contributing factor in the metastatic presentation of our patient's PNET at diagnosis. As both these genes converge on the pseudo-hypoxia signaling pathway, we hypothesize that the co-occurrence of germline pathogenic variants in both the *VHL* and *SDHA* genes could have an impact on the long-term prognosis. Although our hypothesis is limited by the lack of functional studies and a single case report, further studies looking at this genotype-phenotype correlation in similar cases will be helpful. Co-occurrence of two germline variants in *VHL* and *SDHA* in our patient also creates an opportunity for the use of novel therapeutics based on the convergence of these two genes within the pseudo-hypoxia signaling pathway. Recently approved selective HIF inhibitors could be one of the therapeutic considerations in our patient as it has been shown to be effective in patients with renal cell carcinoma due to an underlying germline *VHL* pathogenic variant (23). HIF-2α overexpression is documented in VHL disease associated renal cell carcinoma (24). Preclinical studies indicated the potential efficacy of HIF-2α subunit inhibitors, which blocks the HIF pathway proximally and limits tumor growth in clear cell renal cell carcinoma (25, 26). Subsequent clinical trials documented the benefits and safety of HIF-2α inhibitors (e.g belzutifan) in both sporadic (27) and VHL-

disease associated renal cell carcinoma (23). Belzutifan was also efficacious in reducing tumor size of non-renal cell carcinomas in VHL patients, including pancreatic neuroendocrine tumors, central nervous system hemangioblastomas and retinal hemangioblastomas (23). The U.S Food & Drug Administration (FDA) approved Belzutifan for renal cell carcinoma and non-renal cell neoplasms associated with VHL disease in August 2021 and remains a promising therapy for our patient who presented with both metastatic pancreatic neuroendocrine tumor as well as retinal hemangioma.

CONCLUSION

The co-occurrence of *VHL* and *SDHA* pathogenic variants implicated in PNET is described in this case report. It opens the way for additional exploratory studies to determine the significance of co-occurrence of these variants in terms of tumor penetrance, severity, and outcomes, as well as for the development of novel therapeutic approaches targeting the shared cellular pathway involved in VHL and SDHA-related etiopathogenesis.

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 was 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

MT and BC drafted the original manuscript and created the figure, KD obtained consent and performed counseling for the testing coordinated in the patient. MG provided the pathology images and drafted the pathology description. TS provided radiology images and drafted the imaging description. AS provided patient information and performed manuscript revisions. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank the patient and her family for consenting to participate in this publication. We would like to acknowledge Ambry Genetics for providing the IGV results of the germline sequencing data for our patient. We also acknowledge Foundation One®CDx for providing the IGV results of the tumor sequencing data for our patient. **Figure 3** was created using BioRender.com.

REFERENCES

1. Maddock IR, Moran A, Maher ER, Teare MD, Norman A, Payne SJ, et al. A Genetic Register for Von Hippel-Lindau Disease. *J Med Genet* (1996) 33 (2):120–7. doi: 10.1136/jmg.33.2.120
2. Jhawar S, Arakawa Y, Kumar S, Varghese D, Kim YS, Roper N, et al. New Insights on the Genetics of Pheochromocytoma and Paraganglioma and Its Clinical Implications. *Cancers (Basel)* (2022) 14(3):594. doi: 10.3390/cancers14030594
3. Choudhry H, Harris AL. Advances in Hypoxia-Inducible Factor Biology. *Cell Metab* (2018) 27(2):281–98. doi: 10.1016/j.cmet.2017.10.005
4. Dahia PL. Pheochromocytomas and Paragangliomas, Genetically Diverse and Minimalist, All at Once! *Cancer Cell* (2017) 31(2):159–61. doi: 10.1016/j.ccell.2017.01.009
5. Lonser RR, Glenn GM, Walther M, Chew EY, Libutti SK, Linehan WM, et al. Von Hippel-Lindau Disease. *Lancet* (2003) 361(9374):2059–67. doi: 10.1016/S0140-6736(03)13643-4
6. Maher ER, Bentley E, Yates JR, Barton D, Jennings A, Fellows IW, et al. Mapping of Von Hippel-Lindau Disease to Chromosome 3p Confirmed by Genetic Linkage Analysis. *J Neurol Sci* (1990) 100(1-2):27–30. doi: 10.1016/0022-510X(90)90008-B
7. Roma A, Maruzzo M, Basso U, Brunello A, Zamarchi R, Bezzon E, et al. First-Line Sunitinib in Patients With Renal Cell Carcinoma (RCC) in Von Hippel-Lindau (VHL) Disease: Clinical Outcome and Patterns of Radiological Response. *Fam Cancer* (2015) 14(2):309–16. doi: 10.1007/s10689-014-9771-y
8. de Mestier L, Hammel P. Pancreatic Neuroendocrine Tumors in Von Hippel-Lindau Disease. *Scand J Gastroenterol* (2015) 50(8):1054–5. doi: 10.3109/00365521.2015.1022211
9. Chen H, Sippel RS, O'Dorisio MS, Vinik AI, Lloyd RV, Pacak K. The North American Neuroendocrine Tumor Society Consensus Guideline for the Diagnosis and Management of Neuroendocrine Tumors: Pheochromocytoma, Paraganglioma, and Medullary Thyroid Cancer. *Pancreas.* (2010) 39(6):775–83. doi: 10.1097/MPA.0b013e3181ebb4f0
10. Fishbein L, Nathanson KL. Pheochromocytoma and Paraganglioma Susceptibility Genes: Estimating the Associated Risk of Disease. *JAMA Oncol* (2017) 3(9):1212–3. doi: 10.1001/jamaoncol.2017.0222
11. Jochmanova I, Pacak K. Genomic Landscape of Pheochromocytoma and Paraganglioma. *Trends Cancer.* (2018) 4(1):6–9. doi: 10.1016/j.trecan.2017.11.001
12. Lam AK. Update on Paragangliomas and Pheochromocytomas. *Turk Patoloji Derg* (2015) 31 Suppl 1:105–12. doi: 10.5146/tjpath.2015.01318
13. Ilanchezhian M, Jha A, Pacak K, Del Rivero J. Emerging Treatments for Advanced/Metastatic Pheochromocytoma and Paraganglioma. *Curr Treat Options Oncol* (2020) 21(11):85. doi: 10.1007/s11864-020-00787-z
14. Vicha A, Taieb D, Pacak K. Current Views on Cell Metabolism in SDHx-Related Pheochromocytoma and Paraganglioma. *Endocrine-related cancer.* (2014) 21(3):R261–77. doi: 10.1530/ERC-13-0398
15. Niemeijer ND, Papathomas TG, Korpershoek E, de Krijger RR, Oudijk L, Morreau H, et al. Succinate Dehydrogenase (SDH)-Deficient Pancreatic Neuroendocrine Tumor Expands the SDH-Related Tumor Spectrum. *J Clin Endocrinol Metab* (2015) 100(10):E1386–93. doi: 10.1210/jc.2015-2689
16. Guerrero MA, Schreinemakers JM, Vriens MR, Suh I, Hwang J, Shen WT, et al. Clinical Spectrum of Pheochromocytoma. *J Am Coll Surg* (2009) 209 (6):727–32. doi: 10.1016/j.jamcollsurg.2009.09.022
17. Dubard Gault M, Mandelker D, DeLair D, Stewart CR, Kemel Y, Sheehan MR, et al. Germline SDHA Mutations in Children and Adults With Cancer. *Cold Spring Harb Mol Case Stud* (2018) 4(4):a002584. doi: 10.1101/mcs.a002584
18. Maniam P, Zhou K, Lonergan M, Berg JN, Goudie DR, Newey PJ. Pathogenicity and Penetrance of Germline SDHA Variants in Pheochromocytoma and Paraganglioma (PPGL). *J Endocr Soc* (2018) 2(7):806–16. doi: 10.1210/js.2018-00120
19. Hanson H, Durkie M, Lalloo F, Izatt L, McVeigh TP, Cook JA, et al. UK Recommendations for SDHA Germline Genetic Testing and Surveillance in Clinical Practice. *J Med Genet* (2022). doi: 10.1136/jmedgenet-2021-108355

20. Lawrence B, Gustafsson BI, Chan A, Svejda B, Kidd M, Modlin IM. The Epidemiology of Gastroenteropancreatic Neuroendocrine Tumors. *Endocrinol Metab Clinics North Am* (2011) 40(1):1–18. doi: 10.1016/j.ecl.2010.12.005
21. McGuigan A, Whitworth J, Andreou A, Hearn TGenomics England Research Consortium, , Tischkowitz M, et al. Multilocus Inherited Neoplasia Allele Syndrome (MINAS): An Update. *Eur J Hum Genet* (2022) 30(3):265–70. doi: 10.1038/s41431-021-01013-6
22. Iliopoulos O. Von Hippel-Lindau Disease: Genetic and Clinical Observations. *Front Horm Res* (2001) 28. doi: 10.1159/000061052
23. Jonasch E, Donskov F, Iliopoulos O, Rathmell WK, Narayan VK, Maughan BL, et al. Belzutifan for Renal Cell Carcinoma in Von Hippel–Lindau Disease. *N Engl J Med* (2021) 385(22):2036–46. doi: 10.1056/NEJMoa2103425
24. Chen W, Hill H, Christie A, Kim MS, Holloman E, Pavia-Jimenez A, et al. Targeting Renal Cell Carcinoma With a HIF-2 Antagonist. *Nature* (2016) 539 (7627):112–7. doi: 10.1038/nature19796
25. Acaba LA, Chenworth ML, Gold AS, Wildner AC, Ehlies FJ, Berrocal AM, et al. Terson's Syndrome in a Patient With Von Hippel-Lindau Disease. *Optom Vis Sci* (2016) 93(9):1181–6. doi: 10.1097/OPX.0000000000000904
26. Cho H, Du X, Rizzi JP, Liberzon E, Chakraborty AA, Gao W, et al. On-Target Efficacy of a HIF-2α Antagonist in Preclinical Kidney Cancer Models. *Nature* (2016) 539(7627):107–11. doi: 10.1038/nature19795
27. Choueiri TK, Bauer TM, Papadopoulos KP, Plimack ER, Merchan JR, McDermott DF, et al. Inhibition of Hypoxia-Inducible Factor-2α in Renal Cell Carcinoma With Belzutifan: A Phase 1 Trial and Biomarker Analysis. *Nat Med* (2021) 27(5):802–5. doi: 10.1038/s41591-021-01324-7

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

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fonc.2022.951310

Cost-Effectiveness of BRCA 1/2 Genetic Test and Preventive Strategies: Using Real-World Data From an Upper-Middle Income Country

*Marina Lourenção[1,2], Julia Simões Correa Galendi[3], Henrique de Campos Reis Galvão[2], Augusto Perazzolo Antoniazzi[2], Rebeca Silveira Grasel[2,8], André Lopes Carvalho[4], Edmundo Carvalho Mauad[2], Jorge Henrique Caldeira de Oliveira[1], Rui Manuel Reis[2,5,6], Olena Mandrik[7] and Edenir Inêz Palmero[2,8]**

[1] School of Economics, Business Administration and Accounting at Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil, [2] Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil, [3] Institute of Health Economics and Clinical Epidemiology, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Cologne, Germany, [4] Early Detection Prevention and Infections, International Agency for Research on Cancer, Lyon, France, [5] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, [6] ICVS/3B's – PT Government Associate Laboratory, Guimarães, Portugal, [7] School of Health and Related Research, The University of Sheffield, Sheffield, United Kingdom, [8] Department of Genetics, Brazilian National Cancer Institute, Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:
Giovana Tardin Torrezan, A.C.Camargo Cancer Center, Brazil

Reviewed by:
Maria Del Pilar Estevez Diz, University of São Paulo, Brazil
Renata Sandoval, Hospital Sirio Libanes, Brazil

***Correspondence:**
Edenir Inêz Palmero
edenirip@yahoo.com.br

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

Received: 23 May 2022
Accepted: 20 June 2022
Published: 11 July 2022

Citation:
Lourenção M, Simões Correa Galendi J, Galvão HdCR, Antoniazzi AP, Grasel RS, Carvalho AL, Mauad EC, Oliveira JHCd, Reis RMV, Mandrik O and Palmero EI (2022) Cost-Effectiveness of BRCA 1/2 Genetic Test and Preventive Strategies: Using Real-World Data From an Upper-Middle Income Country. Front. Oncol. 12:951310. doi: 10.3389/fonc.2022.951310

Although BRCA1/2 genetic testing in developed countries is part of the reality for high-risk patients for hereditary breast and ovarian cancer (HBOC), the same is not true for upper-middle-income countries. For that reason, this study aimed to evaluate whether the BRCA1/2 genetic test and preventive strategies for women at high risk for HBOC are cost-effective compared to not performing these strategies in an upper-middle-income country. Adopting a payer perspective, a Markov model with a time horizon of 70 years was built to delineate the health states for a cohort of healthy women aged 30 years that fulfilled the *BRCA1/2* testing criteria according to the guidelines. Transition probabilities were calculated based on real-world data of women tested for *BRCA1/2* germline mutations in a cancer reference hospital from 2011 to 2020. We analyzed 275 *BRCA* mutated index cases and 356 *BRCA* mutation carriers that were first- or second-degree relatives of the patients. Costs were based on the Brazilian public health system reimbursement values. Health state utilities were retrieved from literature. The *BRCA1/2* genetic test and preventive strategies result in more quality-adjusted life years (QALYs) and costs with an incremental cost-effectiveness ratio of R$ 11,900.31 (U$ 5,504.31)/QALY. This result can represent a strong argument in favor of implementing genetic testing strategies for high-risk women even in countries with upper-middle income, considering not only the

cancer prevention possibilities associated with the genetic testing but also its cost-effectiveness to the health system. These strategies are cost-effective, considering a willingness-to-pay threshold of R$ 25,000 (U$ 11,563.37)/QALY, indicating that the government should consider offering them for women at high risk for HBOC. The results were robust in deterministic and probabilistic sensitivity analyses.

Keywords: breast cancer, ovarian cancer, *BRCA* genetic test, preventive strategies, cost-effectiveness

INTRODUCTION

Breast cancer is the most common cancer worldwide and the leading cause of cancer death among women, with globally more than 2.6 million new cases and almost 700,000 deaths annually (1). Despite being less frequent (about 300,000 new cases annually), ovarian cancer has a high lethality rate, with almost seven deaths for every 10 new cases diagnosed (1).

Individuals with hereditary cancer have a higher risk of developing cancer during their lifetime when compared to the general population. Although many high and moderate cancer genes have been discovered and associated with hereditary breast and ovarian cancer (HBOC) in the last years, *BRCA*1 and *BRCA*2 still account for most cases (2–4). Pathogenic germline variants in these genes confer a high risk for developing breast and/or ovarian cancer that can reach 72% and 44% for *BRCA*1 mutation carriers and 69% (for breast cancer) and 17% (for ovarian cancer) for those with *BRCA2* pathogenic alterations (5–7).

Women with personal and/or family history suggestive of HBOC should be referred for genetic counselling and genetic testing to investigate for the presence of pathogenic germline variants (8). Besides, the realization of genetic testing makes it possible to offer it to asymptomatic relatives of the index patient in a predictive context. In this context, the preventive medicine is brought into evidence, once for women who tested positive can be recommended to attend to intensified surveillance for early-detection tumors or risk reduction surgeries, such as risk-reducing bilateral mastectomy (RRBM) and/or risk-reducing salpingo-oophorectomy (RRSO) (2, 9–12). In addition, for relatives not carriers of the pathogenic variant segregating in the family, standard care can be offered (8), once they are not at increased risk of breast or ovarian cancer (13–16).

Several studies on the cost-effectiveness analysis of genetic testing have been performed worldwide. Although previous economic modeling studies indicate that it is cost-effective to provide population-based genetic tests (17–19), the Brazilian Universal Health Coverage System (SUS) does not provide BRCA genetic tests to high-risk women for HBOC. Since 71.5% of the Brazilian population relies exclusively on the SUS (20), most Brazilian women do not have access to the personalized measures for prevention and early diagnosis, as recommended by international guidelines. Thus, this study aimed to evaluate whether offering *BRCA1/2* genetic testing followed by preventive strategies for women at high risk for HBOC is cost-effective when compared to no genetic testing (i.e., and no preventive strategies) in the context of the public health system of an upper middle-income country with continental dimensions as Brazil.

METHODS

We developed a Markov model to assess whether *BRCA1/2* testing and preventive strategies for healthy women at high risk for HBOC are cost-effective compared to standard care (no testing and no preventive strategies). Using the TreeAge Pro, the model estimates the costs and benefits, and the latter is expressed as quality-adjusted life years (QALYs) and life years gained (LYG), highly used in cost-effectiveness studies. To reflect the long-term consequences of breast and ovarian cancer, the model had a 1-year cycle length and time horizon of 70 years (lifetime). The analysis was conducted from the perspective of the SUS as the payer. In line with recommendations from the Brazilian guideline for economic evaluations, costs and effects were discounted at 5% (21).

Considering that HBOC women started annual mammography at 30 years (8) and the low incidence of breast and ovarian cancer at younger ages (22), the target population was a cohort of 30-year-old Brazilian women without a history of breast or ovarian cancer but with first or second-degree relatives who have *BRCA*-related cancer and that fulfilled the National Comprehensive Cancer Network (NCCN) clinical criteria for *BRCA* testing (8).

Strategies for the Comparison

The compared strategies consisted of carrying out the genetic counselling and *BRCA*1/2 genetic testing followed by different surgical/non-surgical preventive options, compared to not performing genetic testing and carrying out these preventive/risk reduction measures. For this model, we considered the clinical criteria for offering genetic testing (and preventive options for carriers) recommended by the National Comprehensive Cancer Network (NCCN) guideline (8). Women identified as *BRCA1/2* mutation carriers were offered four alternatives based on clinical/genetic criteria and personal choice: (i) intensified surveillance with MRI and bilateral mammography annually and breast specialist consultation, CA 125 exam, and transvaginal ultrasonography biannually; (ii) salpingo-oophorectomy; (iii) bilateral mastectomy; or (iv) both salpingo-oophorectomy and bilateral mastectomy (2, 8–11). Women who tested negative and women in the control group (women who did not have genetic counselling and *BRCA* testing) were treated in consonance with the SUS standard care according to their age (e.g., bilateral mammography and medical consultation annually for HBOC women aged 30 years). **Figure 1** presents the compared strategies and the preventive options for high-risk women.

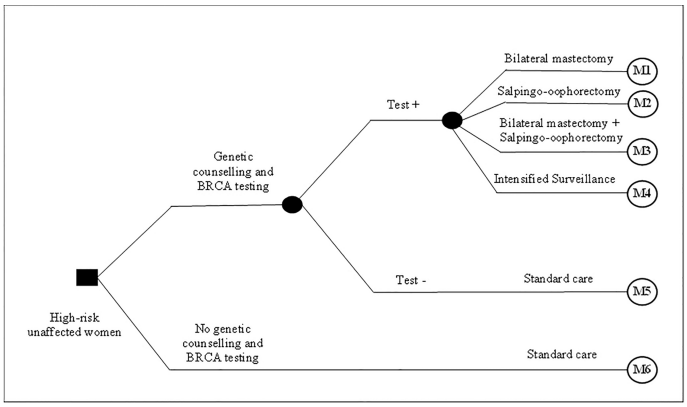


FIGURE 1 | Decision model presenting compared strategies and high-risk reduction options.

Model Overview

The Markov model structure comprises the states "well," "non-metastatic breast cancer," "metastatic breast cancer," "ovarian cancer," "post breast cancer," "post ovarian cancer," and "death" (absorbing state) (**Figure 2**).

Women in the model started in the state well and could go to the states breast or ovarian cancer or die. From breast cancer, they could either stay there, go to post breast cancer, or develop metastatic breast cancer, or ovarian cancer. Women with contralateral breast cancer returned to the initial breast cancer state. The transition from ovarian to breast cancer was not included due to the low incidence of ovarian cancer and its high mortality rates (1).

These states reflect possible clinical events for high-risk women for HBOC. The well state comprises women not diagnosed with cancer; the non-metastatic breast cancer state includes women in the first year after the diagnosis of first or contralateral breast cancer. The metastatic breast cancer state comprehends the first year of diagnosis of disseminated neoplastic cells in an organ distinct from the breast. Likewise, the ovarian cancer state includes women in the first year of diagnosis. Post-cancer states were modeled using tunnel states to reflect annual follow-up costs, utilities, and probabilities after cancer diagnosis until year 5. From the sixth year on, the patients stay in the post-cancer state unless other events occur. The ovarian cancer state was not separated between non-metastatic and metastatic due to its high risk of mortality which is caused by the difficulty to obtain an early detection of the disease (23).

Probabilities

Transition probabilities were obtained mainly from the Barretos Cancer Hospital (BCH) dataset (**Table 1**). The BCH is a Brazilian philanthropic health institution specialized in cancer care, from

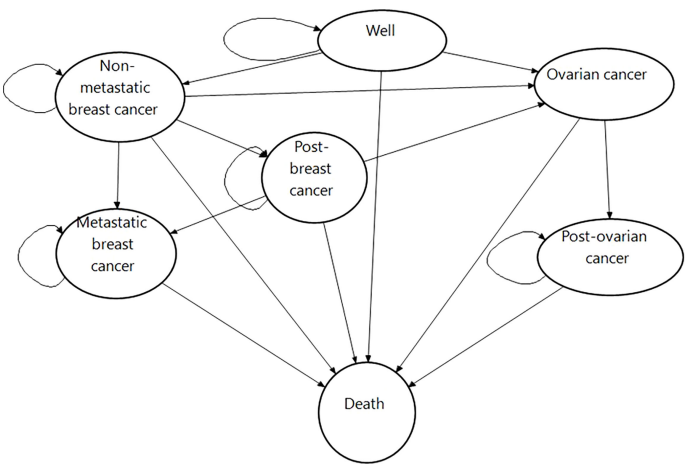


FIGURE 2 | Markov diagram.

TABLE 1 | Input data on annual probabilities and utilities and their sources.

Variable	Value (SD)	Sources
Probabilities		
To be tested positive with a genetic test	0.18	(BCH)
Choice of prophylactic option		
Mastectomy	0.03 (0.02–0.04)	(BCH)
Oophorectomy	0.12 (0.10–0.14)	(BCH)
Both	0.12 (0.09–0.17)	(BCH)
From well to BC		
Carriers	30–34 y.o.: 0.012 (0.01–0.013), 35–39: 0.016 (0.014–0.017), 40–44: 0.022 (0.020–0.024), 45–49: 0.027 (0.025–0.029), 50–54: 0.029 (0.027–0.031), ≥ 55 0.037 (0.033–0.040)	(BCH)
Non-carriers	30–34: 0.0011 (0.0001), 35–39: 0.0017 (0.0001), 40–44: 0.002 (0.0002), 45–49: 0.004 (0.0004), 50–54: 0.006 (0.0006), ≥ 55: 0.008 (0.0008)	(24, 25)
From well to OC		
Carriers	0.013 (0.052)	(BCH)
Non-carriers	0.00008728	(26, 27)
From well to death	30–34: 0.004, 35–39: 0.006, 40–44:0.009, 45–49: 0.013, 50–54: 0.019, 55–59: 0.028, 60–64: 0.043, 65–69: 0.065, 70–74: 0.1, 75–79: 0.16, 80–84: 0.25, 85> 1	(28)
From BC or post-BC to BC		
Carriers	0.069 (0.054–0.091)	BCH
Non-carriers	0.003 (0.001)	(29)
From BC or post-BC to death (BC mortality)	0.006 (0.004)	(BCH)
From BC or post-BC to Metastatic BC	0.0134 (0.0097–0.01737)	(BCH)
From BC or post-BC to OC	0.007 (0.004–0.010)	(BCH)
From metastatic BC to death	y1: 0.37 (0.31–0.48), y2: 0.61 (0.53–0.73), y3: 0.76 (0.68–0.86), y4: 0.85 (0.78–0.92), y5: 0.9 (0.85–0.96)	(BCH)
From OC to death	y1: 0.10 (0.04), y2: 0.18 (0.06); y3: 0.25 (0.07), y4: 0.32 (0.08), y5: 0.39 (0.08)	(BCH)
Development of breast cancer		
Women with bilateral mastectomy	There are no cases reported in BCH	(BCH)
Women with oophorectomy [b]	30–34: 0.014 (0.01–0.013), 35–39: 0.016 (0.014–0.017), 40–44: 0.022 (0.020–0.024), 45–49: 0.027 (0.025–0.029), 50–54: 0.029 (0.027–0.031), ≥ 55 0.037 (0.033–0.040)	(BCH)
Women with bilateral mastectomy and oophorectomy	There are no cases reported at BCH	(BCH)
Development of ovarian cancer		
Women with bilateral mastectomy	There are no cases reported at BCH	(BCH)
Women with oophorectomy	0.01 (0.0004–0.32)	(BCH)
Women with bilateral mastectomy and oophorectomy	0	(BCH)
Utility values		
Well, at age 30	0. 920 (0.0072)—baseline	(30)
Annual decrease due to age	0.00029	(30)
Healthy high-risk women	Multiplier: 0.92	(31)
Prophylactic mastectomy, oophorectomy or both	Multipliers: 0.88 (0.22), 0.95 (0.1), 0.83 (0.1)	(9)
Annual increase after prophylactic mastectomy or both oophorectomy and mastectomy in years 2–5	0.008 (0.001), 0.02 (0.011)	Assumption based on previous modeling studies (22, 32)
BC	Multiplier: 0.77 (0.18)	(33)
Post-BC	Multiplier 0.79 (0.18)	(33)
Annual increase after BC in years 2–5	0.0021 (0.0007)	Assumption based on previous modeling studies (22, 32)
Metastatic BC	Multiplier: 0.64 (0.12)	(34).
OC	Multiplier: 0.34 (0.30)	(35)
Post-OC	Multiplier: 0.83 (0.25)	(35)
Annual increase after OC in years 2–5	0.111 (0.022)	Assumption based on previous modeling studies (22, 32)

BC, breast cancer; OC, ovarian cancer; SD, standard deviation; BCH, Barretos Cancer Hospital.

[b]*It was assumed that it has the similar breast cancer risk of BRCA carrier women.*

prevention to treatment. It is a cancer center (non-profit foundation) that offers services through the Brazilian SUS. However, it differs from other public hospitals because it can receive donations from society, auctions, or organizations. Currently, the *BRCA* genetic test is not offered by hospitals operating in the Brazilian public health system, but the BCH can provide the test for its patients due to funds from donations obtained (36).

The probabilities that could not be retrieved from the BCH dataset were taken preferably from sources that reflected the Brazilian population, e.g., the National Cancer Institute of Brazil (INCA), Brazilian Geography and Statistics Institute (IBGE), and the WHO. The specific references are provided in **Table 1**.

In the BHC dataset, data were available for 2,307 women who performed BRCA1/2 genetic testing from 2011 to 2020. There were 1,544 index cases (i.e., the first member of the family to be tested)—among which there were 275 carriers and 1,269 non-carriers—and 763 first- or second-degree relatives (i.e., 356 carriers and 407 non-carriers). Using Kaplan–Meier in the SPSS software, the transition probabilities were calculated. We defined a different group eligible at baseline for each of the transition probabilities calculated to avoid selection bias regarding a previous cancer diagnosis. Briefly, we considered data from first- or second-degree relatives identified to be carriers to calculate the probability of opting for a prophylactic surgery and the respective risk of developing breast or ovarian cancer afterward. Then, we considered data from index cases to calculate transition probabilities for *BRCA* carriers from the state "well" to "cancer" only for those not submitted to prophylactic surgeries.

As BCH is a reference cancer center for women at high risk, women who tested negative were not followed up at BCH but referred back to the system for general population screening according to their ages, considering that women who tested negative have been shown to have the same risk as the general population (13–16). Thus, the incidence of cancer and mortality for women tested negative were taken from Brazilian registries (26, 27).

Because data on the probability of BC recurrence among *BRCA* non-carriers in Brazil are not available, we used the cumulative 10-year risk of secondary contralateral breast cancer for German non-carriers (29). The data choice was based on similarities in definitions of health states (37).

Utility Data

Utility data were extracted from published studies from a systematic literature search in the PubMed database (**Supplementary Table 1**). When possible, studies reporting utility values for the Brazilian population were preferred. Due to methodological heterogeneity among the studies reporting utilities (37), we used relative utility values applying decrements to the baseline ("well") state (30).

In our study, due to the distress of knowing to have a mutation and distress caused by undergoing risk-reducing surgery, utilities decreased for high-risk women (31), risk-reducing surgeries, and breast or ovarian cancer (9). All women that entered the model were considered high-risk for HBOC. Thus, if they have a negative test result, it was assumed that their utility increases to the utility of healthy women, obtained from Sullivan et al. (2005).

The decrements in utilities for high-risk women for HBOC were based on EQ-5D values of women in Croatia (31). Utilities for the prophylactic surgeries were obtained from Grann et al. (2010), in which a time trade-off instrument (TTO) was applied to *BRCA*-mutated Canadian women. Decreased utilities following prophylactic mastectomy and prophylactic salpingo-oophorectomy were assumed to increase linearly within 5 years to regain the age-specific utility of a high-risk woman, as suggested by other modeling studies (22, 32).

Utilities for breast cancer and post-breast cancer were based on the EQ-5D values from a prospective cohort of Brazilian women newly diagnosed with breast cancer and treatment naïve [32]. The utilities for metastatic breast cancer were extracted from a meta-regression of studies using a Standard Gamble approach (34).

Weighted average utilities for ovarian cancer and post-ovarian cancer were obtained from (35), in which utilities were measured in populations from different countries using the Standard Gamble approach. Following assumptions of other modeling studies (22, 32), it was assumed that women's utility declines as a result of breast or ovarian cancer and then increases linearly for 5 years to reach the age-specific utility of a post-cancer state. **Table 1** presents all input data regarding probabilities and utilities and their sources.

Cost Data

Adopting the perspective of the SUS, direct medical annual costs were calculated for each Markov model health state. Cost data were expressed in Brazilian currency (Reais). The unit cost values for 2021 were obtained from the official SUS database, namely, the Table of Procedures, Medications and Ortheses, Prostheses, and Special Materials for the National Health System (DATASUS Tabnet). Resource use (e.g., diagnostic exam and clinical procedures) was estimated based on recommendations from the NCCN guidelines and interviews with one oncologist and one gynecologist. The cost of breast cancer treatment was calculated as a weighted average that considered cancer molecular type and stage at BCH cohort, indicating a higher proportion of breast cancer diagnosis in the early stage for first- or second-degree relative women than index women (**Supplementary Table 2**).

Considering that the cost of treatment is potentially lower for breast cancer diagnosed at early stages, the annual mean cost was calculated for these two subgroups from the BCH dataset: (i) index women, that is, *BRCA*-mutated women who had breast cancer before the genetic test, and (ii) first- or second-degree relative women, that is, the *BRCA*-mutated women who had cancer after the test.

The cost of the genetic test refers to the price paid by the BCH (**Table 2**) and was obtained from the Laboratory of Molecular Diagnostics from BCH, considering reagents and personal and taking into consideration the costs of performing BRCA1/BRCA2 analysis by next-generation sequencing (NGS) complemented by rearrangement analysis by multiplex length polymorphism analysis (MLPA). Besides, the genetic test cost was calculated as the mean cost of one index and two relative women tested. The costs of intensive screening included the provision of magnetic resonance and bilateral mammography once a year, breast specialist consultation, CA 125 exam, and transvaginal ultrasound twice per year. The costs concerning standard care were related to bilateral mammography and breast specialist consultation once a year. **Table 2** summarizes the cost input data used in the sensitivity analysis. To facilitate comparisons with costs from other countries, conversion of the results presented in Brazilian real (R$) to United States dollar ($)

TABLE 2 | Costs[a] of breast and ovarian cancer (R$ and US$).

Costs of test, preventive surgeries and surveillance, value in R$ (US$)					
BRCA testing			1135[b] (524.98)		
Intensive screening and genetic counseling			428.85 (198.36)		
Standard care			55.00 (25.44)		
Prophylactic mastectomy			3484.26 (1611.59)		
Prophylactic salpingo-oophorectomy			621.00 (287.23)		
Both prophylactic mastectomy and salpingo-oophorectomy			4105.26 (1898.83)		
Breast cancer treatment					
Breast cancer—index					
Cost per procedure group (%)	1° year	2° year	3° year	4° year	5° year
Diagnostic (%)	1,085.93 (6.96)	317.62(33.12)	279.03 (38.80)	279.03 (38.80)	279.03 (38.80)
Surgical procedures (%)	2,219.87 (12.46)	-	-	-	-
Clinical procedures					
Hormonotherapy (%)	440.00 (2.47)	440.00 (45.88)	440.00 (61.19)	440.00 (61.19)	440.00 (61.19)
Neoadjuvant chemotherapy (%)	5,489.87 (30.81)	-	-	-	-
Adjuvant chemotherapy (%)	2,673.44 (15.00)	201.33 (20.99)			
Radiotherapy (%)	5,904.00 (33.14)	-	-	-	-
Sum per health state per year	17,813.11	958.95	719.03	719.03	719.03
Breast cancer—first- or second-degree relatives					
Cost per procedure group (%)	1° year	2° year	3° year	4° year	5° year
Diagnostic (%)	1021.89 (6.17)	317.38 (33.97)	279.03 (38.80)	279.03 (38.80)	279.03 (38.80)
Surgical procedures (%)	2237.39 (13.52)	-	-	-	-
Clinical procedures (%)					
Hormonotherapy (%)	440.00 (2.66)	440.00 (47.09)	440.00 (61.19)	440.00 (61.19)	440.00 (61.19)
Neoadjuvant chemotherapy (%)	4,455.75 (26.93)	-	-	-	-
Adjuvant chemotherapy (%)	2,481.06 (15.00)	176.89 (18.9)			
Radiotherapy (%)	5,904.00 (35.69)	-	-	-	-
Sum per health state per year	16,540.09	934.27	719.03	719.03	719.03
Metastatic breast cancer					
Cost per procedure group (%)	1° year	2° year	3° year	4° year	5° year
Diagnostic (%)	3,124.01 (16.98)	2,956.77 (17.10)	1,690.37 (11.30)	1699.77 (10.68)	1,699.77 (10.68)
Clinical procedures (%)					
Hormonotherapy (%)	440.00 (2.39)	1,663.45 (9.62)	138.62 (0.92)	-	-
Palliative chemotherapy (%)	12,340.32 (67.09)	12,665.36 (73.27)	13,120.92 (87.76)	14,214.25 (89.31)	14,214.25 (89.31)
Radiotherapy (%)	2,488.28 (13.52)	-	-	-	-
Sum per health state per year	18,392.61	17,285.58	14,949.91	15,914.02	15,914,02
Ovarian cancer					
Cost per procedure group (%)	1° year	2° year	3° year	4° year	5° year
Diagnostic (%)	544.18 (3.83)	270.23 (3.15)	259.77 (4.0)	240.68 (100)	240.68 (100)
Surgical procedures (%)	829.10 (5.83)	-	-	-	-
Clinical procedures (%)					
Chemotherapy (%)	6,624.18 (46.57)	-	-	-	-
Palliative chemotherapy (%)	6,227.46 (43.78)	8,303.28 (96.85)	6,227.46 (96.0)	-	
Sum per health state per year	14,224.92	1,654.11	6,487.23	240.68	240.68

[a]*Considering high uncertainty in cost values, for sensitivity analysis, an assumption of 40% standard deviation was made.*
[b]*Mean cost considering one index and two relative women tested.*

was performed by using a web-based tool (CCEMG—EPPI-Centre Cost Converter). This tool considers the Gross Domestic Product deflator index and the Purchasing Power Parities for GDP ("PPP values") to convert currencies.

Model Validation and Sensitivity Analyses

To validate the model, we consulted experts on the adequacy of input data and the conceptual appropriateness of the model. Technical accuracy was checked regarding data entry and potential programming errors (computerized model validation). For cross-model validation, we assessed the extent to which other models for breast cancer prevention came to different conclusions (38). We performed deterministic sensitivity analyses by varying probabilities and utilities considering uncertainty within the respective ranges or confidence limits to characterize overall uncertainty in the outcome measures. To obtain a comprehensive range, the costs were varied within the 40% range, as suggested by (32). Besides, a probabilistic sensitivity analysis with Monte Carlo simulation (10,000 interactions) was conducted. Gamma distributions were used for cost parameters. Probabilities and utilities were considered to be beta-distributed.

RESULTS

Base-Case Analysis

The genetic counseling and *BRCA* testing strategy cost R$ 5,298 (U$ 2,450.51) in the base-case scenario, resulting in an incremental cost of R$ 1,796 (U$ 830.71) compared with the

non-testing strategy. Accordingly, women offered the genetic test had an incremental gain of 0.2 QALYs and 0.2 LYG. The incremental cost-effectiveness ratio (ICER) for the base-case analysis was R$ 11,900.31 (U$ 5,504.31) per QALY and 10,988.67 (U$ 5082.64) per LYG.

In the Brazilian scenario, an exact value of the cost-effectiveness threshold to be applied by the National Commission for the Incorporation of Technologies (CONITEC) in the SUS was not defined (39). However, based on values of thresholds presented in CONITEC recommendation reports, the study by (40) suggested a three-level threshold: low (<R$ 25,000), medium (R$ 25,000 to R$ 70,000), and high (>R$ 70,000). Therefore, to be more conservative, the present study considered a willingness to pay of R$ 25,000. Base-case results are described in **Table 3** and **Supplementary Figure 1**.

Sensitivity Analyses

In the deterministic sensitivity analysis, the variables with the largest impact on the ICER were the discount rate, probability of moving from well to breast cancer after both risk-reducing surgeries, cost of the genetic test, probability from well to breast cancer after salpingo-oophorectomy, and breast cancer treatment costs. A discount rate of zero would reduce the ICER to R$ 3,336.10/QALY (U$ 1,543.06/QALY), and a discount rate of 10% would increase the ICER to R$ 31,617.71/QALY (U$ 14,624.29/QALY). In a scenario where the probability of moving from well to breast cancer after both risk-reducing surgeries is higher (0.08), the ICER increases to R$ 36,362.66/QALY (U$ 16,818.99/QALY). Moreover, in the scenario that only one first- or second-degree relative woman could be tested per index woman tested instead of two, the costs of a genetic test would increase to R$ 2,035.00 (U$941.26/QALY), increasing the ICER to R$ 17,862.54/QALY (U$ 8262.04/QALY), whereas, if the costs of a genetic test per woman decreased to R$ 685 (U$ 316.84) due to testing four relatives per index woman, the ICER would decrease to R$ 8,919.19/QALY (U$ 4125.43/QALY).

The cost-effectiveness ratio was also sensitive to the probability of moving from well to ovarian cancer after salpingo-oophorectomy; if this probability was higher (0.03), the ICER would increase to R$ 16,766.88, while if it was smaller (0.004), the ICER would decrease to R$ 10,179.45. Regarding the breast cancer treatment cost for non-tested women, assuming a 40% higher cost of breast cancer treatment for the non-testing group, the ICER would decrease to R$ 10,630.77/QALY (U$ 4,917.10/QALY). However, if the breast cancer treatment for this group was 40% lower, the ICER would increase to R$ 17,813.11/QALY (U$ 8,239.18/QALY).

The tornado diagram indicates that for almost all intervals considered in the analysis, the testing strategy is considered cost-effective when compared to the no testing strategy (**Figure 3**), considering a willingness to pay R$ 25,000 (U$ 11,563.37) per QALY (**Supplementary Table 3**).

Figure 4 presents the incremental cost-effectiveness plane from the probabilistic sensitivity analyses.

The cost-effectiveness acceptability curve showed a probability of genetic testing being cost-effective of 68.03% at a willingness to pay (WTP) of R$ 25,000/QALY (U$ 11,563.37/QALY) (**Figure 5**). Besides, it becomes cost-effective at a

TABLE 3 | Base case results.

Strategy	Cost (R$)	Incremental costs (R$)	QALYs	Incremental QALYs	LYG	Incremental LYG	ICER (R$) Costs/QALY	Costs/LYG
No testing	3,502		14.4		16,0			
Testing	5,298	1,726	14.6	0.2	16,01	0.2	11,900.31	10, 988.67

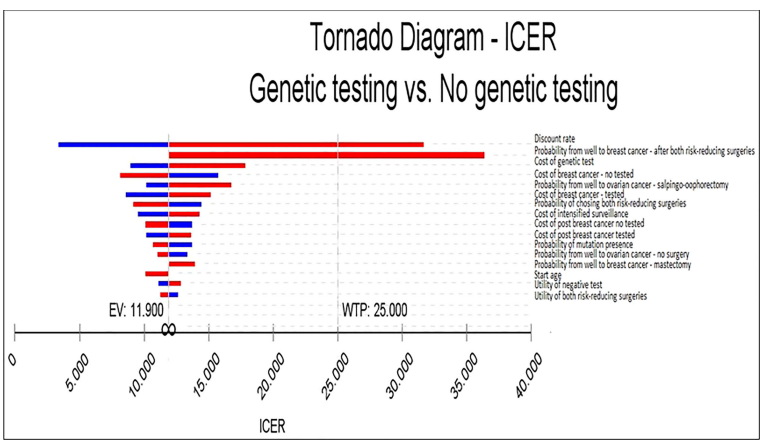


FIGURE 3 | Deterministic sensitivity analyses.

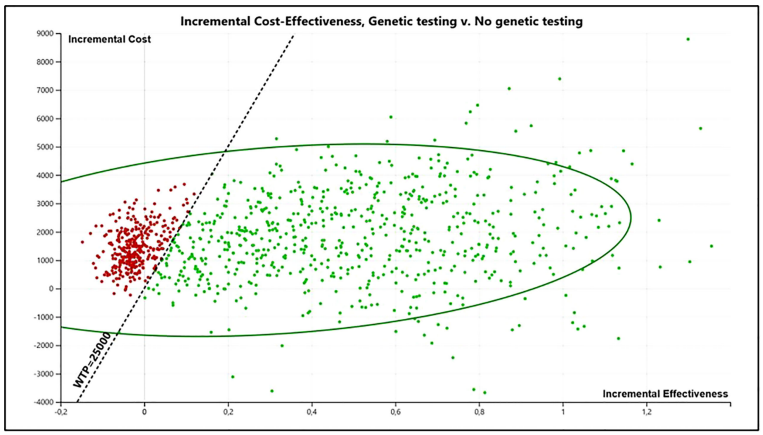


FIGURE 4 | Incremental cost-effectiveness plane from the probabilistic sensitivity analyses (PSA) (10,000 interactions).

minimum threshold of R$ 7,500/QALY (U$ 3,469.01/QALY). **Supplementary Figure 2** presents the incremental net monetary benefit (INMB) versus willingness to pay analysis.

While perceived by the upper-middle-income countries' government as a potentially highly costly intervention, as our results show, carrying out the genetic counseling, *BRCA*1/2 genetic tests, and preventive options in women at high risk for HBOC is a very cost-effective intervention compared to not carrying out these actions when considering a willingness to pay of R$ 25,000/QALY. The ICER for the base-case analysis was R$ 11,900.31 (U$ 5,504.31). The sensitivity analysis also revealed a superiority of the testing strategy. The tornado diagram points out that genetic testing is cost-effective for all scenarios. The probabilistic sensitivity analysis indicates a probability of genetic testing being cost-effective of 68.03%.

The main novelty of our results is that this is the first study for upper-middle-income countries on BRCA genetic tests whose probabilities were mainly extracted from trial-based analysis with Brazilian registries (i.e., penetrance of *BRCA*, rates of uptake prophylactic procedures, breast or ovarian cancer development, etc.). It shows that genetic testing can be cost-effective even in upper-middle-income countries. The use of patients' clinical data increases the representativeness of the results of our analysis for Brazilian women. This real-world evidence provides a more accurate representation of the target population for several reasons. First, the penetrance of BRCA is highly associated with the genetic profile of the population, and the Latin-American population is underrepresented in most international registries and databases. Second, rates of uptake prophylactic procedures vary widely worldwide since these are highly preference-sensitive decisions influenced by sociocultural factors. Therefore, our results add to the existing literature by demonstrating the cost-effectiveness of BRCA in a model that accurately reflects the epidemiology and the preferences of Latin-

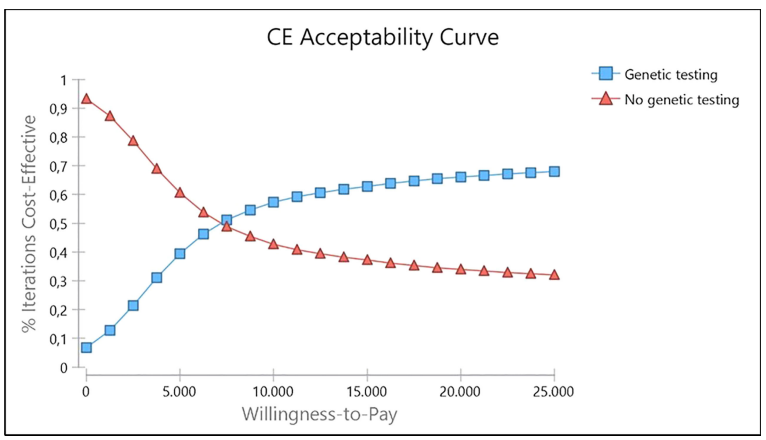


FIGURE 5 | Cost-effectiveness acceptability curve for genetic testing strategy provided to Brazilian women.

American women (from Brazil) at-risk for hereditary breast cancer.

Previous Markov model studies evaluated the cost-effectiveness of genetic *BRCA* testing for breast or ovarian cancer in high-risk women compared to no test in Brazil (17, 22, 41). All of them concluded that *BRCA* testing seems to be the cost-effective strategy with ICERs of R$ 24.264/QALY (22), R$ 908,52 per case of cancer avoided (41), and $ 20,995/QALY (17).

The Brazilian study by Simoes Correa-Galendi et al. (2021) had a similar structure. However, the data were extracted exclusively from the literature. For instance, the uptake rates taken from a UK cohort were 0.09 for prophylactic mastectomy and 0.22 for prophylactic salpingo-oophorectomy. In contrast, the uptake rates identified in our study were 0.03 and 0.12, respectively. These data demonstrate huge differences when considering real-world data from the Brazilian population. According to a recent systematic review by (42), the variability of uptake rates of the risk-reducing surgeries might be explained by several factors, such as (i) cultural differences, (ii) individual-related factors, (iii) age-dependent factors, and (iv) an improved acceptance of preventive surgeries over time. Besides these reasons, it is worth mentioning that economic factors might have also influenced the Brazilian uptake rates identified in our study, for instance, the lack of access to risk-reducing surgeries in the public setting and patients' expenses with transport, accommodation, and absence from work, once the services that provide these surgeries are usually in cancer reference hospitals or large capitals.

Ramos et al. (2019) evaluated the preventive strategies only for the relatives of patients diagnosed with ovarian cancer, while the present study analyzed preventive strategies for first- or second-degree relatives of patients diagnosed with breast or ovarian cancer. Finally, the study by (17) differs from the present research because they evaluate the population-based BRCA testing, which possibly is why the ICER is higher.

The clinical data used in the present study indicate a probability of 82% to receive a negative *BRCA* genetic test. Considering that breast or ovarian cancer risk among non-carrier women from positive families is similar to the general population (13–16), it is important to note that there is still an around 50% probability that the relative will get a negative test and no longer be considered at high risk.

In this context, the benefit that a negative test can bring to patients is not trivial and should be considered in economic modeling studies (43). Our study considers that by obtaining a negative result, the unaffected patient (from a BRCA-mutated family) stops having the utility of a patient at high risk for HBOC and starts obtaining the utility of a woman without high risk at her age due to the reduction in their level of distress (44). This increase in utility occurs as the woman may no longer be excessively concerned with intensified surveillance and breast and/or ovarian cancer (31, 45). A small utility increase due to the relief of receiving a negative test result was also assumed by a previous modeling study (46). However, while Holland et al. justified this increase in utility due to an assumption, in the present study, this analysis is supported by recent evidence from (31).

Of note, our present study provided a conservative analysis considering that only two female relatives are tested for each index. The cost of testing an index patient in Brazil is around R$ 1,800 (U$ 832—including labor, reagents, and rearrangement analysis), while the cost of testing a relative of the patient is R$ 235 (U$ 108). Thus, the average unit cost of testing a family member is calculated at R$ 1135 (U$ 525). However, if we consider that it is possible to test a higher number of family members for each index tested, the unit cost of testing a family member would be reduced, making the genetic testing strategy even more cost-effective, as its ICER would be reduced.

The main strength of the present study was to conduct a trial-based analysis to obtain transition probabilities from the data of women tested from 2011 to 2020 in a Brazilian hospital. Another advantage was the cost data analysis. The unit cost values were obtained from the official Brazilian Health System database. Concerning breast cancer treatment, the costs were calculated separately for molecular types (triple-negative, Luminal A, Luminal B, and Her2+) and stages (47, 48) to reflect a realistic scenario of the resource use. In consonance with growing evidence (17, 32, 49), the present study also highlights the genetic testing contribution to earlier cancer detection. For instance, our cohort showed a high rate of stage III/IV in the index patients (52.08%), and a tendency of earlier diagnosis (stages I/II) in patients who performed genetic testing before a cancer diagnosis (58.67%). Besides, even though the uptake of risk-reducing surgeries was low in our cohort, these patients had access to intensified surveillance with breast magnetic resonance imaging (MRI) which might also contribute to earlier diagnosis.

Limitations might have affected the results. Although data from BCH were preferred, probabilities of events that happen with *BRCA* non-carriers had to be taken from the literature because those patients were not followed up at the BCH. In addition, because of the sample size of the BCH database, the transition probabilities calculated in the present study might not generalize to the entire Brazilian population. Another limitation of the model is the unknown risk of *BRCA*-negative women with a family history of cancer compared with cancer risks observed in the general population. It was considered that non-carriers of genetic mutations that came from positive families did not show an increased risk for HBOC (13–16) and, according to the NCCN guideline, should have a standard care (8). Lastly, because most data on utility specific to the Brazilian population are not available, we used studies from other countries that reported the most similar health-related quality of life; moreover, the available data on utility are not homogeneous. Nevertheless, the sensitivity analysis reveals that for all intervals considered, the testing strategy is considered cost-effective compared to the no testing strategy, considering a willingness to pay of R$ 25,000 per QALY.

Importantly, the present results can support policy development on the topic. Currently, genetic testing is not covered by the Universal Health Coverage in Brazil. The present study uses Brazilian women's clinical data to support the argumentation that the Brazilian public health system should offer the *BRCA* genetic test for women with a family history that

leads to increased risk for HBOC. Our results indicate that a comprehensive genetic test-and-screen strategy for high-risk Brazilian women results in a substantial gain of QALY at moderate additional costs. Although genetic testing followed by preventive surgeries appears to be the most economically advantageous option, women's preferences should always be considered and drive the final treatment decision.

CONCLUSION

In this study, we showed that a screen-and-treat strategy for healthy women at risk for HBOC results in more QALYs and moderately more costs, with an ICER of R$ 11,900.31 (U$ 5,504.31) per QALY gained. The cost-effectiveness of the screen-and-treat intervention depends on a still undecided cost-effectiveness threshold for Brazil, but it would be cost-effective considering a willingness to pay of R$ 25,000 (U$ 11,563.37) per QALY. These results might be reproducible in other upper middle-income countries.

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 Barretos Cancer Hospital's research ethical committee (approval number: 56164716.9.0000.5437). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conceptualization, ML and EP. Methodology, ML, JC-G, AA, OM and EP. Software, ML and JC-G. Validation, ML, JC-G and OM. Formal analysis, ML. Investigation, ML and HG. Resources, HG and EP. Data curation, ML. Writing—original draft preparation, ML and RG. Writing—review and editing, EP, JO, JC-G, OM, RR, AC and EM. Visualization, ML. Supervision, EP. Project administration, EP. Funding acquisition, EP. All authors have read and agreed to the published version of the manuscript.

FUNDING

This project was funded through grants from the National Oncology Care Support Program (PRONON, Grant number 25000.056766/2015-64) from the Brazilian Ministry of Health. EIP and RMR are recipients of the National Council for Scientific and Technological Development (CNPq) productivity fellowships. The study sponsors had no involvement in the study design, collection, analyses or interpretation of data.

ACKNOWLEDGMENTS

We wish to thank members of the Center of Molecular Diagnosis, Onco-genetics Department, and Molecular Oncology Research Center of Barretos Cancer Hospital for their contributions to the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.951310/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: A Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Howell A, Gandhi A, Howell S, Wilson M, Maxwell A, Astley S, et al. Long-Term Evaluation of Women Referred to a Breast Cancer Family History Clinic (Manchester UK 1987–2020). *Cancers* (2020) 12:3697. doi: 10.3390/cancers12123697
3. Azribi F, Abdou E, Dawoud E, Ashour M, Kamal A, Al Sayed M, et al. Prevalence of BRCA1 and BRCA2 Pathogenic Sequence Variants in Ovarian Cancer Patients in the Gulf Region: The PREDICT Study. *BMC Cancer* (2021) 21:1350. doi: 10.1186/s12885-021-09094-8
4. Gorodetska I, Kozeretska I, Dubrovska A. *BRCA* Genes: The Role in Genome Stability, Cancer Stemness and Therapy Resistance. *J Cancer* (2019) 10:2109–27. doi: 10.7150/jca.30410
5. Felicio PS, Grasel RS, Campacci N, de Paula AE, Galvão HCR, Torrezan GT, et al. Whole-Exome Sequencing of non-BRCA1/BRCA2 Mutation Carrier Cases at High-Risk for Hereditary Breast/Ovarian Cancer. *Hum Mutat* (2021) 42:290–9. doi: 10.1002/humu.24158
6. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. *JAMA* (2017) 317:2402. doi: 10.1001/jama.2017.7112
7. Wu Y, Pan X, Dou J, Zhang Q, Li Y, Sheng Y, et al. A Novel Germline BRCA1 Mutation Identified in a Family With Hereditary Breast and Ovarian Cancer Syndrome. *Clin Med Insights Oncol* (2021) 15:11795549211028568. doi: 10.1177/11795549211028569
8. Daly MB, Karlan BY, Pal T, Pilarski R, Pederson HJ, Reiser G, et al. NCCN Guidelines Index Table of Contents Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Discussion. *Risk Assess* (2020) 8:119. doi: 10.6004/jnccn.2020.0017
9. Grann VR, Patel P, Bharthuar A, Jacobson JS, Warner E, Anderson K, et al. Breast Cancer-Related Preferences Among Women With and Without BRCA Mutations. *Breast Cancer Res Treat* (2010) 119:177–84. doi: 10.1007/s10549-009-0373-6
10. Gupta S, Kadayaprath G, Gupta N, Barthwal V. Bilateral Risk-Reducing Prophylactic Mastectomies in an Unaffected BRCA1 Carrier Using Dermal

Sling and Implant. *Indian J Surg Oncol* (2021) 12. doi: 10.1007/s13193-021-01370-0
11. Sekine M, Nishino K, Enomoto T. BRCA Genetic Test and Risk-Reducing Salpingo-Oophorectomy for Hereditary Breast and Ovarian Cancer: State-Of-the-Art. *Cancers* (2021) 13:2562. doi: 10.3390/cancers13112562
12. Cadiz F, Kuerer HM, Puga J, Camacho J, Cunill E, Arun B. Establishing a Program for Individuals at High Risk for Breast Cancer. *J Cancer* (2013) 4:433–46. doi: 10.7150/jca.6481
13. Girardi F, Barnes DR, Barrowdale D, Frost D, Brady AF, Miller C, et al. Risks of Breast or Ovarian Cancer in BRCA1 or BRCA2 Predictive Test Negatives: Findings From the EMBRACE Study. *Genet Med* (2018) 20:1575–82. doi: 10.1038/gim.2018.44
14. Guedaoura S, Pelletier S, Foulkes WD, Hamet P, Simard J, Wong N, et al. No Evidence of Excessive Cancer Screening in Female Noncarriers From BRCA1/2 Mutation–Positive Families. *Curr Oncol* (2017) 24:352–9. doi: 10.3747/co.24.3759
15. Korde LA, Mueller CM, Loud JT, Struewing JP, Nichols K, Greene MH, et al. No Evidence of Excess Breast Cancer Risk Among Mutation-Negative Women From BRCA Mutation-Positive Families. *Breast Cancer Res Treat* (2011) 125:169–73. doi: 10.1007/s10549-010-0923-y
16. Nelson HD, Fu R, Goddard K, Mitchell JP, Okinaka-Hu L, Pappas M, et al. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer: Systematic Review to Update the U.S. Preventive Services Task Force Recommendation, in: *Agency for Healthcare Research and Quality (Us)* (2013). Available at: http://www.ncbi.nlm.nih.gov/books/NBK179201/ (Accessed July 1, 2021).
17. Manchanda R, Sun L, Patel S, Evans O, Wilschut J, De Freitas Lopes AC, et al. Economic Evaluation of Population-Based BRCA1/BRCA2 Mutation Testing Across Multiple Countries and Health Systems. *Cancers (Basel)* (2020) 12:1–38. doi: 10.3390/cancers12071929
18. Manchanda R, Patel S, Antoniou AC, Levy-Lahad E, Turnbull C, Evans DG, et al. Cost-Effectiveness of Population Based BRCA Testing With Varying Ashkenazi Jewish Ancestry. *Am J Obstetrics Gynecol* (2017) 217:578.e1–578.e12. doi: 10.1016/j.ajog.2017.06.038
19. Michaan N, Leshno M, Safra T, Sonnenblick A, Laskov I, Grisaru D. Cost Effectiveness of Whole Population BRCA Genetic Screening for Cancer Prevention in Israel. *Cancer Prev Res* (2021) 14:455–62. doi: 10.1158/1940-6207.CAPR-20-0411
20. IBGE. *IBGE | Biblioteca | Detalhes | Pesquisa Nacional De Saúde : 2019 : Informações Sobre Domicílios, Acesso E Utilização Dos Serviços De Saúde : Brasil, Grandes Regiões E Unidades Da Federação / IBGE, Coordenação De Trabalho E Rendimento* (2020). Available at: https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101748 (Accessed July 6, 2021).
21. Antonini Ribeiro R, Lavanholi Neyeloff J, Itria A, Cristina Canuto Santos V, Manso de Mello Vianna C, Nunes da Silva E, et al. Diretriz Metodológica Para Estudos De Avaliação Econômica De Tecnologias Em Saúde No Brasil. *JBES* (2016) 8:174–84. doi: 10.21115/JBES.v8.n3.p174-184
22. Simoes Correa-Galendi J, Del Pilar Estevez Diz M, Stock S, Müller D. Economic Modelling of Screen-And-Treat Strategies for Brazilian Women at Risk of Hereditary Breast and Ovarian Cancer. *Appl Health Econ Health Policy* (2021) 19:97–109. doi: 10.1007/s40258-020-00599-0
23. Stewart C, Ralyea C, Lockwood S. Ovarian Cancer: An Integrated Review. *Semin Oncol Nurs* (2019) 35:151–6. doi: 10.1016/j.soncn.2019.02.001
24. *IBGE B Geography and Statistics Institute*. Available at: https://cidades.ibge.gov.br/brasil/sp/barretos/pesquisa/23/25207?indicador=25207 (Accessed June 30, 2021).
25. INCA Brazil. *Informações De Registro De Câncer De Base Populacional* (2010). Available at: https://www.inca.gov.br/BasePopIncidencias/PreparарConsultarRelatorioValorAbsoluto.action?tipoFaixaEtaria=1 (Accessed June 30, 2021).
26. IBGE. *Projeções Da População | IBGE* (2018). Available at: https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?=&t=resultados (Accessed June 30, 2021).
27. INCA. INCA, in: *INCA - Instituto Nacional De Câncer* (2018). Available at: https://www.inca.gov.br/numeros-de-cancer (Accessed June 30, 2021).
28. *Mortality Country Profile*. Available at: https://www.who.int/data/mortality/country-profile (Accessed December 21, 2021).
29. Engel C, Fischer C, Zachariae S, Bucksch K, Rhiem K, Giesecke J, et al. Breast Cancer Risk in BRCA1/2 Mutation Carriers and Noncarriers Under Prospective Intensified Surveillance. *Int J Cancer* (2020) 146:999–1009. doi: 10.1002/ijc.32396
30. Sullivan PW, Lawrence WF, Ghushchyan V. A National Catalog of Preference-Based Scores for Chronic Conditions in the United States. *Med Care* (2005) 43:736–49. doi: 10.1097/01.mlr.0000172050.67085.4f
31. Žigman T, Lukša I, Mihaljević G, Žarković M, Kirac I, Vrdoljak DV, et al. Defining Health-Related Quality of Life in Localized and Advanced Stages of Breast Cancer – the First Step Towards Hereditary Cancer Genetic Counseling. *Acta Clinica Croatica* (2020) 59:209–15. doi: 10.20471/acc.2020.59.02.02
32. Müller D, Danner M, Schmutzler R, Engel C, Wassermann K, Stollenwerk B, et al. Economic Modeling of Risk-Adapted Screen-and-Treat Strategies in Women at High Risk for Breast or Ovarian Cancer. *Eur J Health Econ* (2019) 20:739–50. doi: 10.1007/s10198-019-01038-1
33. Guerra RL, Dos Reis NB, Corrêa FDM, Fernandes MM, Ribeiro Alves Fernandes R, Cancela MDC, et al. Breast Cancer Quality of Life and Health-State Utility at a Brazilian Reference Public Cancer Center. *Expert Rev Pharmacoecon Outcomes Res* (2020) 20:185–91. doi: 10.1080/14737167.2019.1621752
34. Peasgood T, Ward SE, Brazier J. Health-State Utility Values in Breast Cancer. *Expert Rev Pharmacoeconomics Outcomes Res* (2010) 10:553–66. doi: 10.1586/erp.10.65
35. Havrilesky LJ, Broadwater G, Davis DM, Nolte KC, Barnett JC, Myers ER, et al. Determination of Quality of Life-Related Utilities for Health States Relevant to Ovarian Cancer Diagnosis and Treatment. *Gynecol Oncol* (2009) 113:216–20. doi: 10.1016/j.ygyno.2008.12.026
36. Hospital BC. Barretos Cancer Hospital, in: *Barretos Cancer Hospital*. Available at: http://www.hcancerbarretos.com.br/en/ (Accessed July 23, 2021).
37. Brazier J, Ara R, Azzabi I, Busschbach J, Chevrou-Séverac H, Crawford B, et al. Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health* (2019) 22:267–75. doi: 10.1016/j.jval.2019.01.004
38. Vemer P, Corro Ramos I, van Voorn GAK, Al MJ, Feenstra TL. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. *Pharmacoeconomics* (2016) 34:349–61. doi: 10.1007/s40273-015-0327-2
39. Soarez PCD, Novaes HMD. Limiares De Custo-Efetividade E O Sistema Único De Saúde. *Cad Saúde Pública* (2017) 33:1–5. doi: 10.1590/0102-311x00040717
40. Zimmermann IR, de Oliveira EF, Vidal ÁT, Santos VCC, Petramale CA. A Qualidade Das Evidências E as Recomendações Sobre a Incorporação De Medicamentos No Sistema Único De Saúde:: Uma Análise Retrospectiva. *Rev Eletrônica Gestão Saúde* (2015) 6:3043–65. doi: 10.18673/gs.v6i4.22099
41. Ramos MC de A, Folgueira MAAK, Maistro S, Campolina AG, de Soárez PC, de Bock GH, et al. Cost Effectiveness of the Cancer Prevention Program for Carriers of the BRCA1/2 Mutation. *Rev Saúde Pública* (2018) 52:1–10. doi: 10.11606/S1518-8787.2018052000643
42. Simões Corrêa Galendi J, Kautz-Freimuth S, Stock S, Müller D. Uptake Rates of Risk-Reducing Surgeries for Women at Increased Risk of Hereditary Breast and Ovarian Cancer Applied to Cost-Effectiveness Analyses: A Scoping Systematic Review. *Cancers* (2022) 14:1786. doi: 10.3390/cancers14071786
43. Simões Corrêa Galendi J, Vennedey V, Kentenich H, Stock S, Müller D. Data on Utility in Cost–Utility Analyses of Genetic Screen-And-Treat Strategies for Breast and Ovarian Cancer. *Cancers* (2021) 13:4879. doi: 10.3390/cancers13194879
44. Lombardi L, Bramanti SM, Babore A, Stuppia L, Trumello C, Antonucci I, et al. Psychological Aspects, Risk and Protective Factors Related to BRCA Genetic Testing: A Review of the Literature. *Support Care Cancer* (2019) 27:3647–56. doi: 10.1007/s00520-019-04918-7
45. Wenzel L, Osann K, Lester J, Kurz R, Hsieh S, Nelson EL, et al. Biopsychological Stress Factors in BRCA Mutation Carriers. *Psychosomatics* (2012) 53:582–90. doi: 10.1016/j.psym.2012.06.007

46. Holland ML, Huston A, Noyes K. Cost-Effectiveness of Testing for Breast Cancer Susceptibility Genes. *Value Health* (2009) 12:207–16. doi: 10.1111/j.1524-4733.2008.00418.x
47. Brandão M, Morais S, Lopes-Conceição L, Fontes F, Araújo N, Dias T, et al. Healthcare Use and Costs in Early Breast Cancer: A Patient-Level Data Analysis According to Stage and Breast Cancer Subtype. *ESMO Open* (2020) 5:e000984. doi: 10.1136/esmoopen-2020-000984
48. Yarden Y, Baselga J, Miles D. Molecular Approach to Breast Cancer Treatment. *Semin Oncol* (2004) 31:6–13. doi: 10.1053/j.seminoncol.2004.07.016
49. Sun L, Brentnall A, Patel S, Buist D, Bowles E, Evans D, et al. A Cost-Effectiveness Analysis of Multigene Testing for All Patients With Breast Cancer. *JAMA Oncol* (2019) 5:1718–30. doi: 10.1001/jamaoncol.2019.3323

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

The reviewer MD declared a shared affiliation, with no collaboration, with the authors ML and JO to the handling editor at the time of review.

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 24 August 2022
DOI 10.3389/fonc.2022.957472

OPEN ACCESS

EDITED BY
Giovana Tardin Torrezan,
A.C.Camargo Cancer Center, Brazil

REVIEWED BY
Katia Candido Carvalho,
Clinical Hospital, Faculty of Medicine, University of São Paulo, Brazil
Qian Chen,
Guangxi Medical University Cancer Hospital, China

*CORRESPONDENCE
Li Hong
drhongli7777@gmail.com

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

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

RECEIVED 31 May 2022
ACCEPTED 26 July 2022
PUBLISHED 24 August 2022

CITATION
Tang J, Tian X, Min J, Hu M and Hong L (2022) RPP40 is a prognostic biomarker and correlated with tumor microenvironment in uterine corpus endometrial carcinoma.
Front. Oncol. 12:957472.
doi: 10.3389/fonc.2022.957472

RPP40 is a prognostic biomarker and correlated with tumor microenvironment in uterine corpus endometrial carcinoma

Jianming Tang[1†], Xiaoli Tian[2†], Jie Min[1], Ming Hu[1] and Li Hong[1*]

[1]Department of Obstetrics and Gynecology, Renmin Hospital of Wuhan University, Wuhan, China, [2]Department of Pathology, Renmin Hospital of Wuhan University, Wuhan, China

Ribonuclease P/MRP Subunit P40 (RPP40), a component of ribonuclease P and multimeric ribonuclease P complex, was reported as one of the promoting factors for the chemoresistance of acute myeloid leukemia and a recurrence predictor of early-stage triple-negative breast cancer. However, the functional role of RPP40 in uterine corpus endometrial carcinoma (UCEC) is unclear. In this study, comprehensive bioinformatic analyses were conducted to explore the predictive role of RPP40 on UCEC diagnosis and prognosis, as well as the underlying mechanism. Differential analyses of multiple databases showed that both messenger RNA (mRNA) and the protein expression of RPP40 were significantly upregulated in UCEC tumor tissues. Furthermore, the RPP40 mRNA expression level was significantly correlated with the clinicopathological characteristics of UCEC patients, including the clinical stage, primary therapy outcome, histological type, histologic grade, overall survival event, disease-specific survival event, and progression-free interval event. Receiver operating characteristic (ROC) analysis showed that RPP40 was a reliable predictor for UCEC diagnosis with an area under the curve (AUC) of 0.775, a sensitivity of 0.829, and a specificity of 0.719. Kaplan–Meier, Cox regression, and nomogram analyses showed that high RPP40 expression was an independent prognostic factor for the 1-year, 3-year, and 5-year survival of UCEC patients. In addition, the enrichment analysis of RPP40-associated differentially expressed genes and correlation analyses showed that the expression of RPP40 was correlated with the regulation of extracellular matrix and immune cell infiltration. In conclusion, the upregulation of RPP40 is significantly correlated with the poor survival and tumor microenvironment of UCEC, suggesting that RPP40 is a promising biomarker of poor prognosis and a potential target of chemotherapy or immunotherapy in UCEC.

KEYWORDS

RPP40, uterine corpus endometrial carcinoma, immune infiltration, extracellular matrix, tumor microenvironment, prognosis

Introduction

Uterine corpus endometrial carcinoma (UCEC) is the third most commonly diagnosed gynecological cancer and the seventh most common malignant tumor in women worldwide (1, 2). Over 60,000 new cases are expected next year in American women (3). Generally, early screening and therapies can significantly reduce the incidence, recurrence, and mortality of UCEC. Nevertheless, the patients in advanced stages usually respond poorly to conventional treatments, with a 5-year survival rate as low as 17% (4). In recent years, evolving medical drugs and technologies have slowed the decline in the long-term survival rate in UCEC patients. However, novel prognostic biomarkers and therapeutic targets for improving the survival rate of UCEC patients still need continuous exploration.

The tumor microenvironment (TME), composed of multiple cellular and molecular components, has been implicated in cancer cell survival, proliferation, invasion, and therapeutic response (5–7). Various members of TME, such as cancer-associated fibroblasts (CAFs), immune cells, extracellular matrix (ECM), cytokines, and chemokines, act together to regulate phenotypes, antitumor immunity, and the therapeutic response of malignant tumors (5–8). The metabolic and biologic changes of malignant cells driven by oncogenes can influence the TME to suppress antineoplastic immune responses and induce therapeutic resistance (7). Meanwhile, this also reveals a novel strategy for cancer therapy to remodel the TME by targeting hub oncogenes and related signaling pathways.

Ribonuclease P/MRP Subunit P40 (RPP40), a 40-KDa protein subunit of ribonuclease P (RNase P), was reported to enable RNase P RNA binding activity and then contribute to the generation of mature tRNA molecules (9–11). Moreover, RPP40 is also a component of the multimeric ribonuclease P (MRP) complex, which cleaves pre-rRNA sequences (12). At present, the molecular function of RPP40 remains unclear since there are few studies concerning it. As other components of RNase P or MRP, both RPP25 and RPP30 were reported as reliable prognostic risk factors for glioblastoma multiforme (11, 13) and also have been reported to promote the proliferation, migration, invasion, and cell cycle program of cervical cancer cells (14). Similarly, RPP40 was also regarded as one of the promoting factors for the chemoresistance of acute myeloid leukemia (15) and recurrence predictor of early-stage triple-negative breast cancer (16). Furthermore, the result of bioinformatics analysis in this study showed that RPP40 was one of the potential prognostic genes for UCEC (Supplementary File 1). Therefore, we speculated that RPP40 might be a potential prognostic biomarker or therapeutic target of UCEC and might be involved in its tumorigenesis or progression.

Based on the above speculation, we first analyzed the expression difference, survival prognosis, and possible molecular function of RPP40 in UCEC in this study. We found that both mRNA and protein expression were significantly upregulated in UCEC tumor tissues. Moreover, RPP40 was an effective diagnostic and prognostic predictor of UCEC. In addition, gene enrichment analysis revealed that RPP40 was involved in regulating the TME, especially ECM dysregulation and immune cell infiltration.

Materials and methods

TCGA database and data processing

Transcriptional expression data of 21 types of cancer and paired clinical data of UCEC were downloaded from the The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). RNA sequencing data were transformed from the format of fragments per kilobase per million (FPKM) to transcripts per million reads (TPM) for further analyses. This study did not require ethical approval since the research data we used was acquired from public online databases. Then, the mRNA expression differences between tumor tissues and normal tissues were determined in 21 types of cancer, including bladder urothelial carcinoma (BLCA), breast-invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and UCEC.

The University of Alabama at Birmingham cancer data analysis portal and clinical proteomic tumor analysis consortium

The University of Alabama at Birmingham Cancer data analysis Portal (UALCAN; http://ualcan.path.uab.edu/analysis-prot.html) is a public online database that provides protein expression analysis option using data from Clinical Proteomic Tumor Analysis Consortium (CPTAC) and the International

Cancer Proteogenome Consortium (ICPC) datasets (17, 18). In this study, we compared the protein expression difference between primary UCEC tumor samples (n=100) and normal endometrial samples (n=31) using the data from CPTAC on ULCAN. The z-value represents standard deviations from the median across samples for UCEC. Log2 spectral count ratio values from CPTAC were first normalized within each sample profile and then normalized across samples.

The Human Protein Atlas

The Human Protein Atlas (HPA; https://www.proteinatlas.org/) is a public database that contains the protein expression data of human protein-coding genes (19, 20). The immunohistochemical staining pictures of normal and tumor tissues were publicly available in this database. In this study, we compared the protein expression of RPP40 between UCEC tumor tissue and normal endometrial tissue on HPA.

Study design, grouping, and sample size

The study flowchart is illustrated in Figure 1. In this study, 552 UCEC patients were divided into two groups, high- and low-RPP40-expression groups, according to the median value of RPP40 expression in UCEC tumor samples. There were 276 patients in each group. Then the patients were divided into subgroups for further analyses based on each clinicopathological characteristics. The sample size of each subgroup is shown in the Supplementary Table 1.

Correlation analysis for RPP40 expression and clinicopathological characteristics of uterine corpus endometrial carcinoma patients

The clinicopathological characteristics of UCEC patients between high- and low-RPP40-expression groups were compared using the Wilcoxon rank sum test (continuous variables) or Pearson's chi-square test (rank variables). Secondly, the correlation research of RPP40 expression with clinicopathological characteristics was investigated *via* logistic analysis. Next, the expression differences of RPP40 among different subgroups of clinicopathological characteristics were compared by an independent t-test. A p-value <0.05 was regarded statistically significant.

Clinical significance evaluation of RPP40 expression in uterine corpus endometrial carcinoma

To evaluate the predictive value of RPP40 in UCEC diagnosis, receiver operating characteristic (ROC) analysis was conducted using an R package of "pROC" (21). Next, Kaplan–Meier (K-M), univariate, and multivariate Cox regression analyses were employed for prognosis analysis, including overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). The R packages "rms" and "survival" were applied to construct nomograms and calibration plots. The R packages "forestplot" and "survival" were used for the clinicopathological subgroup study. All survival data in this study were acquired from

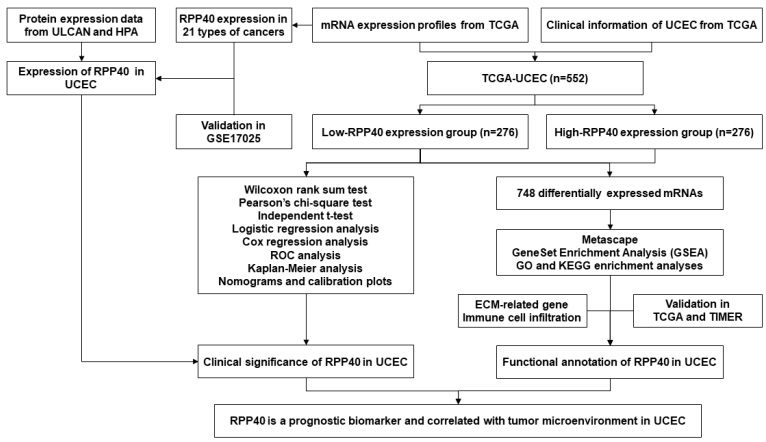


FIGURE 1
Flow diagram of this study.

the published research (22). All the above analyses were all accomplished with R (v3.6.3), and a p-value <0.05 was considered as the statistical threshold.

RPP40-related differentially expressed genes in uterine corpus endometrial carcinoma tumors

Differentially expressed genes (DEGs) between high- and low-RPP40-expression groups were screened out using R package "DESeq2" (23). Furthermore, The R package "ggplot2" was used to illustrate results as volcano plots and heatmaps. P<0.05 and |log2 Fold change|>1.0 were set as thresholds for DEGs with statistical significance.

Enrichment analysis of RPP40-associated DEGs in uterine corpus endometrial carcinoma tumors

The DEGs with significance were then processed for enrichment analysis on the Metascape database (http://metascape.org) (24), with the analysis thresholds of counts≥3, enrichment factors>1.5, and P-value<0.01. Furthermore, the R package "clusterProfiler" (25) was utilized for the gene set enrichment analysis (GSEA) (26) of the DEGs between two groups, as well as the Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. The data set of "c2.cp.v7.2.symbols.gmt" from MSigDB collections were selected as reference gene sets in GSEA analysis. The number of analysis permutations was set to 1000, and False discovery rate (FDR)<0.25 and adjusted P-value<0.05 were set as analysis thresholds in GSEA.

Association of RPP40 and immune cell infiltration in uterine corpus endometrial carcinoma tumors

Firstly, a single-sample GSEA method from R package "GSVA" (27) was used to analyze the correlation between the RPP40 expression and infiltration of 24 common immune cell types (28), including dendritic cells (DCs), activated DCs (aDCs), B cells, CD8 T cells, cytotoxic cells, eosinophils, immature DCs (iDCs), macrophages, mast cells, neutrophils, natural killer (NK) cells, NK CD56bright cells, NK CD56dim cells, plasmacytoid DCs (pDCs), T cells, T helper cells, T central memory (Tcm), T effector memory (Tem), T follicular helper (TFH), T gamma delta (Tgd), Th1 cells, Th17 cells, Th2 cells, and Treg. Secondly, the immune cell infiltration levels between high- and low-RPP40-expression groups were compared by an independent-samples t-test. Furthermore, the association between the RPP40 expression and gene marker levels of immune cells in UCEC tumor tissues was determined *via* the Tumor IMmune Estimation Resource (TIMER) database (https://cistrome.shinyapps.io/timer/). A p-value <0.05 was regarded statistically significant in all above analyses.

Results

Expression profiles of RPP40 in pan-cancer perspective

To determine the mRNA expression pattern of RPP40 in different cancers, the mRNA expression data of RPP40 in the tumors and corresponding normal tissues of different cancer types based on the TCGA database were analyzed. As shown in Figure 2A, when compared with normal samples, the RPP40 mRNA expression of tumor samples were significantly upregulated in the tumor samples of 17 cancer types, including BLCA, BRCA, CESC, CHOL, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, STAD, and UCEC according to the TCGA database. These results indicate that the mRNA expression of RPP40 is significantly upregulated in a variety of cancer types.

Next, the mRNA expression characteristic of RPP40 in UCEC was further determined. As shown in Figure 2B, paired data analysis showed that the mRNA expression levels of RPP40 in UCEC tumor tissues were significantly upregulated than those in normal endometrial tissues (n=23) according to the TCGA database. Similarly, as shown in Figure 2C, the RPP40 mRNA expression levels of UCEC tumor tissues (n=552) were significantly increased than those of normal tissues (n=35) in unpaired data analysis. We also validated the mRNA expression of RPP40 in the GSE17025 dataset. As shown in Supplementary Figure 1, RPP40 mRNA expression levels in UCEC tumor tissues (n=91) were significantly upregulated than those in normal tissues (n=12). Furthermore, the protein expression of RPP40 in UCEC was analyzed on both UALCAN and HPA databases. As shown in Figure 2D, the protein expression of RPP40 in primary UCEC (CPTAC samples, n=100) was significantly higher than those in normal endometrial tissues (CPTAC samples, n=31). As same as the research result from CPTAC samples, immunohistochemical staining results from the HPA database also confirmed that the protein level of RPP40 was markedly upregulated in UCEC tumor tissues (Figures 2E, F). These results indicate that both the mRNA and protein expression of RPP40 are significantly upregulated in UCEC tumor tissues.

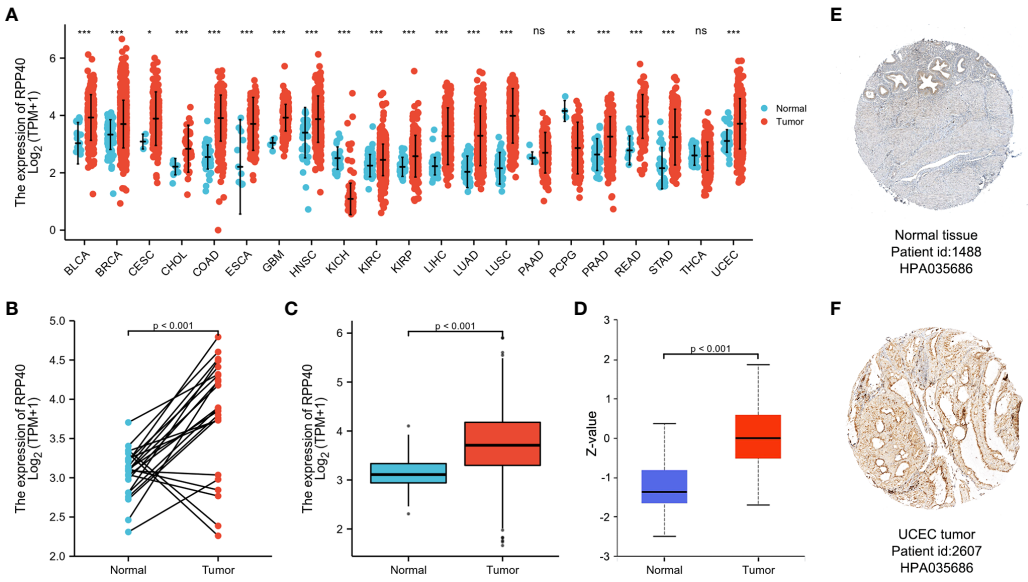


FIGURE 2
Expression of Ribonuclease P/MRP Subunit P40 (RPP40) in pan-cancer perspective. **(A)** The comparison of RPP40 mRNA expression between tumor and normal tissues in different cancer types based on the TCGA database. **(B)** Paired analysis of the mRNA expression levels of RPP40 in 23 UCEC samples and matched normal samples from the TCGA database. **(C)** Unpaired analysis of the mRNA expression levels of RPP40 in 552 UCEC samples and 35 normal samples from the TCGA database. **(D)** The protein expression difference of RPP40 between UCEC tumor tissues and normal endometrial tissues based on CPTAC. **(E, F)** The immunohistochemical staining of RPP40 protein in normal endometrial sample **(E)** and UCEC tumor sample **(F)** based on Human Protein Atlas. ns, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Association between RPP40 expression and clinicopathological characteristics in uterine corpus endometrial carcinoma patients

To evaluate the potential clinical significance of RPP40 in UCEC, 552 UCEC patients were divided into two groups, high- (n=276) and low- (n=276) RPP40-expression groups, based on the RPP40 mRNA expression levels in UCEC tumor tissues. Then, the clinicopathological characteristics of UCEC patients between different RPP40 expression levels were compared (Table 1). The results showed that the RPP40 mRNA expression level was significantly correlated with the clinical stage, primary therapy outcome, histological type, histologic grade, OS event, DSS event, and PFI event of UCEC patients. Moreover, logistics analysis was applied to further confirm the correlation between RPP40 expression and clinicopathological characteristics. As shown in Table 2, RPP40 expression was positively correlated with the clinical stage (OR=1.617, *P*=0.011), histological grade (OR=3.280, *P*<0.001), histological type (OR=3.166, *P*<0.001), and primary therapy outcome (OR=2.864, *P*=0.004). Moreover, we also investigated the expression differences of RPP40 among different subgroups of clinicopathological characteristics. The result showed that RPP40 expression was significantly increased in patients with clinical stages III–IV (Figure 3A), histological grade G3 (Figure 3B), the histological type of serous and mixed (Figure 3C), age over 60 years old (Figure 3D), the primary therapy outcome of PD&SD&PR (Figure 3E), and dead patients in the survival event of OS, DSS, and PFI (Figures 3F–H). At the same time, there were no significant differences in RPP40 expression between the two subgroups of BMI, residual tumor, tumor invasion, menopause status, hormone therapy, diabetes, radiation therapy, and surgical approach (Supplementary Figure 2).

Predictive values of RPP40 for the diagnosis and prognosis of UCEC patients

ROC curve analysis was conducted to further explore the clinical significance of RPP40 in UCEC patients. The result showed that RPP40 was a reliable predictive biomarker for the diagnosis of UCEC, with an area under the curve (AUC) of 0.775, a sensitivity of 0.829, and a specificity of 0.719 (Figure 4A). Furthermore, K-M analyses were conducted to evaluate the prognostic value of RPP40 in UCEC patients. As shown in Figures 4B–D, the OS (HR=2.42, *P*<0.001), DSS (HR=2.50, *P*=0.001), and PFI (HR=1.80, *P*=0.001) of the patients in high-RPP40-expression patients were all significantly shorter than those in low-RPP40-expression patients. Moreover, to further evaluate the prognostic value of RPP40 in UCEC patients, univariate and multivariate Cox

TABLE 1 Clinicopathological characteristics of uterine corpus endometrial carcinoma (UCEC) patients with differential RPP40 expression.

Characteristic	Low-RPP40 expression (N = 276)	High-RPP40 expression (N = 276)	*P*-value
Clinical stage, n (%)			0.038
Stage I	185 (33.5%)	157 (28.4%)	
Stage II	25 (4.5%)	26 (4.7%)	
Stage III	57 (10.3%)	73 (13.2%)	
Stage IV	9 (1.6%)	20 (3.6%)	
Primary therapy outcome, n (%)			< 0.001
PD	8 (1.7%)	12 (2.5%)	
SD	3 (0.6%)	3 (0.6%)	
PR	0 (0%)	12 (2.5%)	
CR	238 (49.6%)	204 (42.5%)	
Race, n (%)			0.176
Asian	11 (2.2%)	9 (1.8%)	
Black or African American	45 (8.9%)	63 (12.4%)	
White	195 (38.5%)	184 (36.3%)	
Age, n (%)			0.088
<=60	113 (20.6%)	93 (16.9%)	
>60	161 (29.3%)	182 (33.2%)	
BMI, n (%)			0.179
<=30	99 (19.1%)	113 (21.8%)	
>30	163 (31.4%)	144 (27.7%)	
Histological type, n (%)			< 0.001
Endometrioid	234 (42.4%)	176 (31.9%)	
Mixed	11 (2%)	13 (2.4%)	
Serous	31 (5.6%)	87 (15.8%)	
Residual tumor, n (%)			0.857
R0	193 (46.7%)	182 (44.1%)	
R1	10 (2.4%)	12 (2.9%)	
R2	8 (1.9%)	8 (1.9%)	
Histologic grade, n (%)			< 0.001
G1	74 (13.7%)	24 (4.4%)	
G2	73 (13.5%)	47 (8.7%)	
G3	125 (23.1%)	198 (36.6%)	
Tumor invasion (%), n (%)			0.697
<50	134 (28.3%)	125 (26.4%)	
>=50	116 (24.5%)	99 (20.9%)	
Menopause status, n (%)			0.500
Pre	20 (4%)	15 (3%)	
Peri	10 (2%)	7 (1.4%)	
Post	223 (44.1%)	231 (45.7%)	
Hormones therapy, n (%)			0.416
No	148 (43%)	149 (43.3%)	
Yes	27 (7.8%)	20 (5.8%)	
Diabetes, n (%)			0.403
No	171 (37.9%)	157 (34.8%)	
Yes	58 (12.9%)	65 (14.4%)	
Radiation therapy, n (%)			0.980
No	142 (26.9%)	137 (26%)	
Yes	125 (23.7%)	123 (23.3%)	
Surgical approach, n (%)			0.874
Minimally invasive	103 (19.4%)	105 (19.8%)	
Open	163 (30.8%)	159 (30%)	
OS event, n (%)			< 0.001
Alive	246 (44.6%)	212 (38.4%)	
Dead	30 (5.4%)	64 (11.6%)	
DSS event, n (%)			0.003
Alive	256 (46.5%)	231 (42%)	
Dead	20 (3.6%)	43 (7.8%)	
PFI event, n (%)			0.005
Alive	226 (40.9%)	197 (35.7%)	
Dead	50 (9.1%)	79 (14.3%)	

regression analyses were accomplished in this study. As shown in Table 3, RPP40 expression was an independent risk factor for OS (HR=2.491, *P*=0.007), DSS (HR= 3.060, *P*=0.011) and PFI (HR=1.811, *P*=0.045) in multivariate Cox regression. Moreover, the clinical stage and primary therapy outcome also showed prognostic values for OS, DSS, and PFI, the residual tumor also showed a prognostic value for DSS, and the histological type also showed a prognostic value for PFI in multivariate Cox regression analyses.

Next, all the significant prognostic factors in multivariate Cox regression analyses were used for prognostic nomogram construction. Then, the corresponding calibration curves were drawn for further testing the efficiency of each nomogram. As shown in Figure 5, the clinical stage, primary therapy outcome, and RPP40 expression were used to predict 1-, 3-, and 5-year OS with a C-index of 0.779 (Figures 5A, B). The clinical stage, primary therapy outcome, residual tumor, and RPP40 expression were used to predict 1-, 3-, and 5-year DSS with a C-index of 0.871 (Figures 5C, D). The clinical stage, primary therapy outcome, histological type, and RPP40 expression were used to predict 1-, 3-, and 5-year PFI with a C-index of 0.728 (Figures 5E, F). The calibration curves showed a desirable prediction of OS and DSS nomograms for the 1-, 3-, and 5-year clinical outcomes, with a slightly overestimated mortality in patients with predicted mortality higher than 50% in the 3- and 5-year prediction of OS. These results indicated that RPP40 was a reliable prognostic biomarker for UCEC, especially in predicting OS and DSS.

TABLE 2 Logistic regression analysis of association between clinicopathological characteristics and RPP40 expression in UCEC patients.

Characteristics	Odds Ratio (95%CI)	*P*-value
Clinical stage (Stage III–IV vs. Stage I–II)	1.617 (1.116-2.353)	0.011
Histologic grade (G3 vs. G1–2)	3.280 (2.293-4.724)	<0.001
Histological type (Mixed and Serous vs. Endometrioid)	3.166 (2.114-4.808)	<0.001
Age (>60 vs. ≤60)	1.374 (0.972-1.945)	0.073
BMI (>30 vs. ≤30)	0.774 (0.544-1.099)	0.152
Menopause status (Post vs. Pre and Peri)	1.413 (0.794-2.549)	0.243
Primary therapy outcome (PD and SD and PR vs. CR)	2.864 (1.422-6.156)	0.004
Residual tumor (R1 and R2 vs. R0)	1.178 (0.603-2.317)	0.630
Tumor invasion (%) (≥50 vs. <50)	0.915 (0.636-1.314)	0.630
Hormones therapy (Yes vs. No)	0.736 (0.391-1.364)	0.333
Diabetes (Yes vs. No)	1.221 (0.806-1.851)	0.346
Radiation therapy (Yes vs. No)	1.020 (0.724-1.436)	0.910
Surgical approach (Open vs. Minimally Invasive)	0.957 (0.675-1.356)	0.804

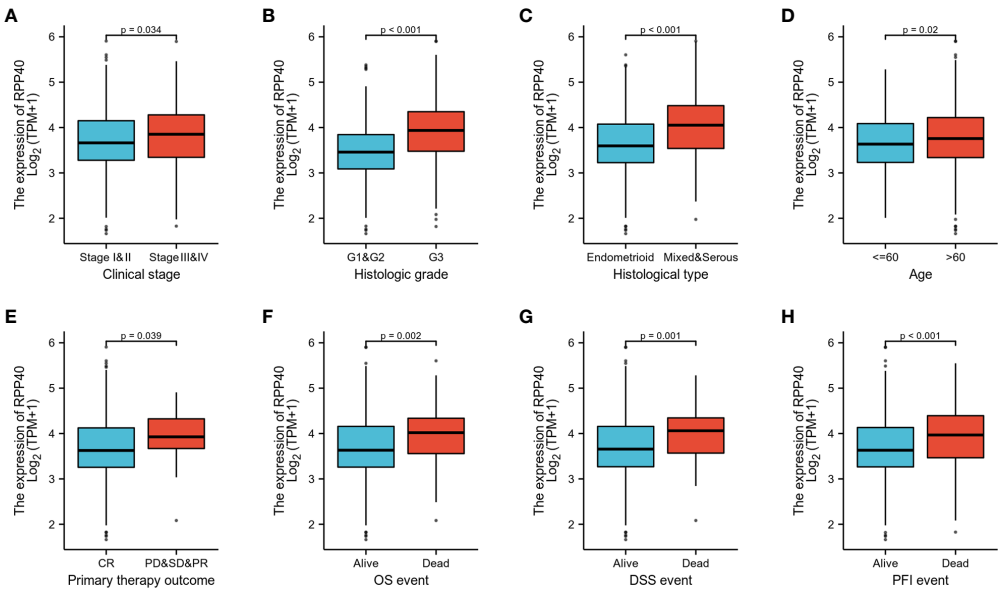


FIGURE 3
RPP40 expression is associated with clinicopathological characteristics in UCEC patients. The expression differences of RPP40 between distinct subgroups of UCEC patients are based on different clinicopathological characteristics, including clinical stage **(A)**, histological grade **(B)**, histological type **(C)**, age **(D)**, primary therapy outcome **(E)**, OS event **(F)**, DSS event **(G)**, and PFI event **(H)**.

Prognostic values of RPP40 in uterine corpus endometrial carcinoma clinicopathological subgroups

To further confirm the prognostic value of RPP40 in UCEC, a subgroup study of each clinicopathological factor was conducted by Cox regression analysis, and the results of subgroup analyses were presented as forest plots. As shown in Figure 6A, the upregulation of RPP40 was a risk factor for OS in UCEC patients with clinical stage I–II (HR=1.920, *P*=0.044), clinical stage III–IV (HR=3.170, *P*<0.001), histological grade G3 (HR=1.720, *P*=0.025), histological type of endometrioid (HR=2.500, *P*=0.002), age below 60 years old (HR=7.760, *P*=0.001), age over 60 years old (HR=1.700, *P*=0.031), a BMI less than 30 kg/m^2 (HR=2.270, *P*=0.015), a BMI over 30 kg/m^2 (HR=2.190, *P*=0.011), postmenopause status (HR=2.180, *P*=0.001), the primary therapy outcome of CR (HR=2.790, *P*=0.001), residual tumor R0 (HR=2.060, *P*=0.012), tumor invasion less than 50% of the muscular layer (HR=3.720, *P*=0.005), or tumor invasion over 50% of the muscular layer

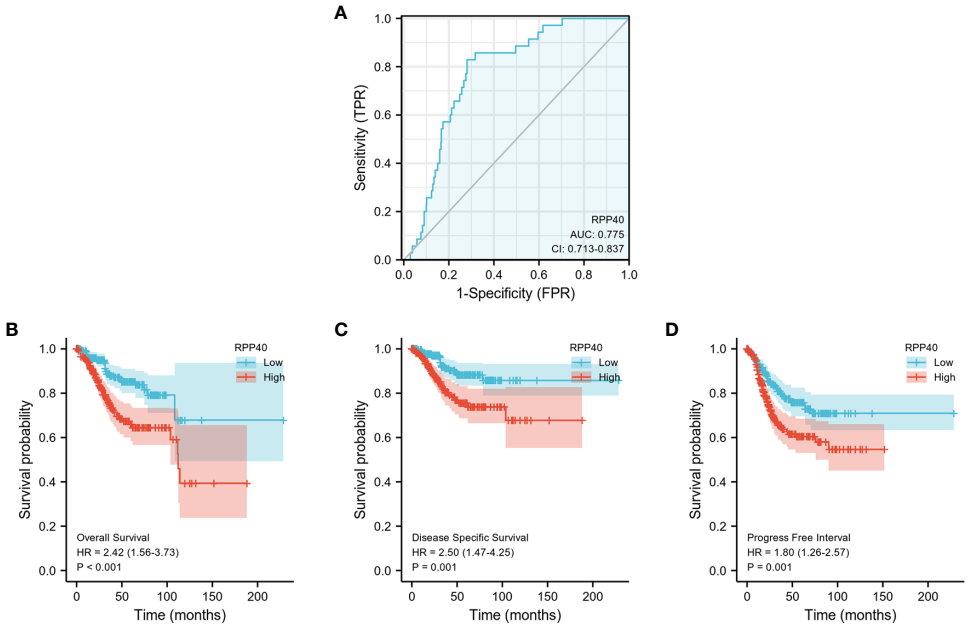


FIGURE 4
Predictive value of RPP40 expression for diagnosis and survival in UCEC patients. **(A)** ROC curve analysis was used to evaluate the performance of RPP40 for UCEC diagnosis. **(B-D)** K-M analyses were used to compare OS **(B)**, DSS **(C)**, and PFI **(D)** between high- and low-RPP40-expression groups of UCEC patients.

TABLE 3 Cox regression analysis for clinical outcomes in UCEC patients.

Characteristics	HR (95% CI) for OS		HR (95% CI) for DSS		HR (95% CI) for PFI	
	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
Clinical stage (III–IV vs. I–II)	3.543***	3.849***	7.030***	5.641**	3.169***	2.692**
Histologic grade (G3 vs. G1–2)	3.281***	1.062^{NS}	7.851***	1.615^{NS}	2.088***	0.673^{NS}
Histological type (Mixed and Serous vs. Endo)	2.628***	1.286^{NS}	3.572***	1.450^{NS}	2.109***	2.079*
Age (>60 vs. ≤60)	1.847*	1.549^{NS}	0.215^{NS}		1.353^{NS}	
BMI (>30 vs. ≤30)	0.967^{NS}		0.948^{NS}		1.046^{NS}	
Menopause status (Post vs. Pre and Peri)	1.050^{NS}		1.214^{NS}		1.637^{NS}	
Residual tumor (R1 and R2 vs. R0)	3.101***	2.201^{NS}	5.310***	3.309*	2.724***	1.963^{NS}
Diabetes (Yes vs. No)	1.172^{NS}		1.195^{NS}		1.169^{NS}	
Surgical approach (Open vs. Minimally Invasive)	0.709^{NS}		0.661^{NS}		0.629*	0.587^{NS}
Primary therapy outcome (PD and SD and PR vs. CR)	7.729***	3.409**	13.602***	5.412***	8.331***	6.283***
Tumor invasion (%) (≥50 vs. <50)	2.813***	0.259^{NS}	3.281***	1.026^{NS}	1.885**	1.439NS
RPP40 (High vs. Low)	2.417***	2.491**	2.497***	3.060*	1.799**	1.811*

^{NS}P>0.05, *P<0.05, **P<0.01, ***P<0.001.

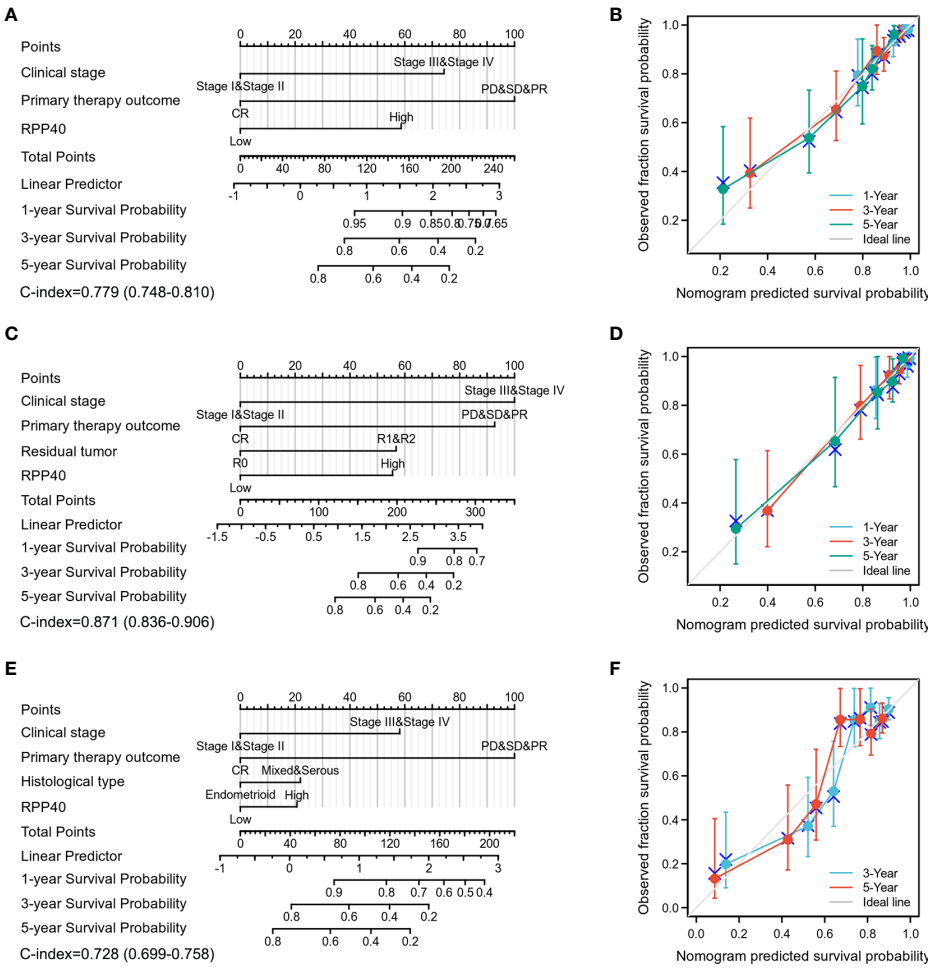


FIGURE 5
Construction and validation of nomograms in UCEC prognosis based on RPP40 expression. The nomograms were constructed to establish RPP40 expression-based risk scoring models for 1-, 3-, and 5-year OS **(A)**, DSS **(C)**, and PFI **(E)**. Calibration plots were drawn to validate the efficiency of nomograms for OS **(B)**, DSS **(D)**, and PFI **(F)**.

(HR=2.080, *P*=0.011). Furthermore, as shown in Figure 6B, the upregulation of RPP40 was a risk factor for DSS in patients with clinical stage III–IV (HR=3.660, *P*<0.001), histological grade G3 (HR=1.960, *P*=0.018), a histological type of endometrioid (HR=2.270, *P*=0.034), age below 60 years old (HR=11.090, *P*=0.001), BMI less than 30 kg/m^2 (HR=2.670, *P*=0.020), BMI over 30 kg/m^2 (HR=2.160, *P*=0.038), postmenopause status (HR=2.240, *P*=0.004), the primary therapy outcome of CR (HR=4.350, *P*=0.001), residual tumor R0 (HR=2.400, *P*=0.025), or tumor invasion over 50% of the muscular layer (HR=2.800, *P*=0.004). Moreover, as shown in Figure 6C, the upregulation of RPP40 was also a risk factor for PFI in patients with clinical stage III–IV (HR=2.320, *P*=0.001), histological grade G3 (HR=1.650, *P*=0.017), a histological type of endometrioid (HR=1.640, *P*=0.033), age below 60 years old (HR=1.900, *P*=0.027), a BMI over 30 kg/m^2 (HR=2.130, *P*=0.003), postmenopause status (HR=1.760, *P*=0.003), the primary therapy outcome of CR (HR=1.940, *P*=0.005), residual tumor R0 (HR=1.710, *P*=0.027), tumor invasion less than 50% of the muscular layer (HR=2.560, *P*=0.004), or tumor invasion over 50% of the muscular layer (HR=1.920, *P*=0.010).

Next, K-M analyses for the OS, DSS, and PFI of clinicopathological subgroups were performed to compare clinical outcomes between high- and low-RPP40 groups. As shown in Figure 7 and Supplementary Figure 3, except DSS for tumor invasion less than 50% of the muscular layer subgroup, the RPP40 expression level exhibited a significantly prognostic value in different clinicopathological subgroups, including clinical stage III–IV, histological grade G3, the histological type of endometrioid, the primary therapy outcome of CR,

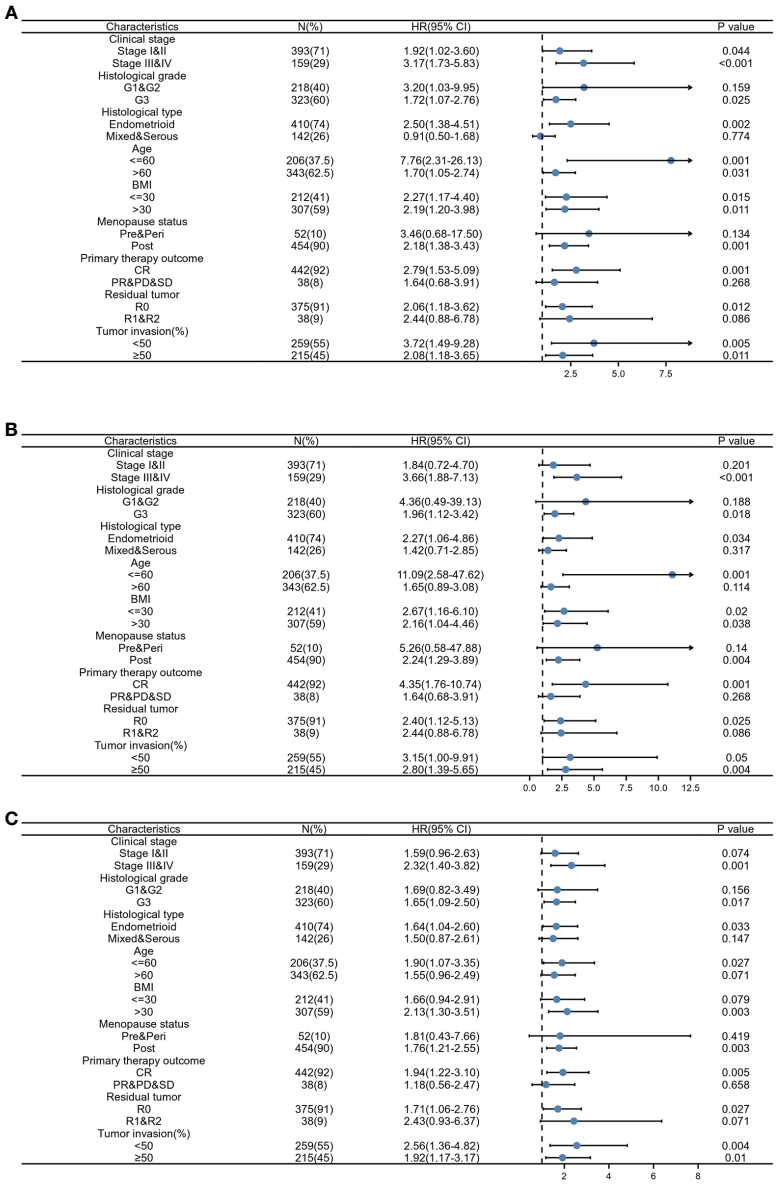


FIGURE 6
Prognostic performance of RPP40 on clinical outcomes in different subgroups of UCEC patients. Patients were divided into different subgroups according to clinical stage, histological grade, histological type, age, BMI, menopause status, primary therapy outcome, residual tumor, and tumor invasion. For each subgroup, the prognostic performance of RPP40 on OS **(A)**, DSS **(B)**, and PFI **(C)** were evaluated by Cox regression, and the results are presented as a hazard ratio. The bar represents the 95% confidence interval of the hazard ratio.

residual tumor R0, tumor invasion over 50% of the muscular layer, tumor invasion less than 50% of the muscular layer, age below 60 years old, a BMI less than 30 kg/m^2, and postmenopause status. These results indicated the prognostic value of RPP40 in UCEC was independent of the above clinicopathological factors, and the patients with low RPP40 expression possess significantly better clinical outcomes than those with high RPP40 expression.

Identification and functional annotation of RPP40-associated DEGs in uterine corpus endometrial carcinoma

In order to explore the function of RPP40 in UCEC, the DEGs between high- and low- RPP40 expression groups were identified. As shown in Figure 8 and Supplementary Figure 4, 748 mRNAs (including 200 upregulated and 548 downregulated

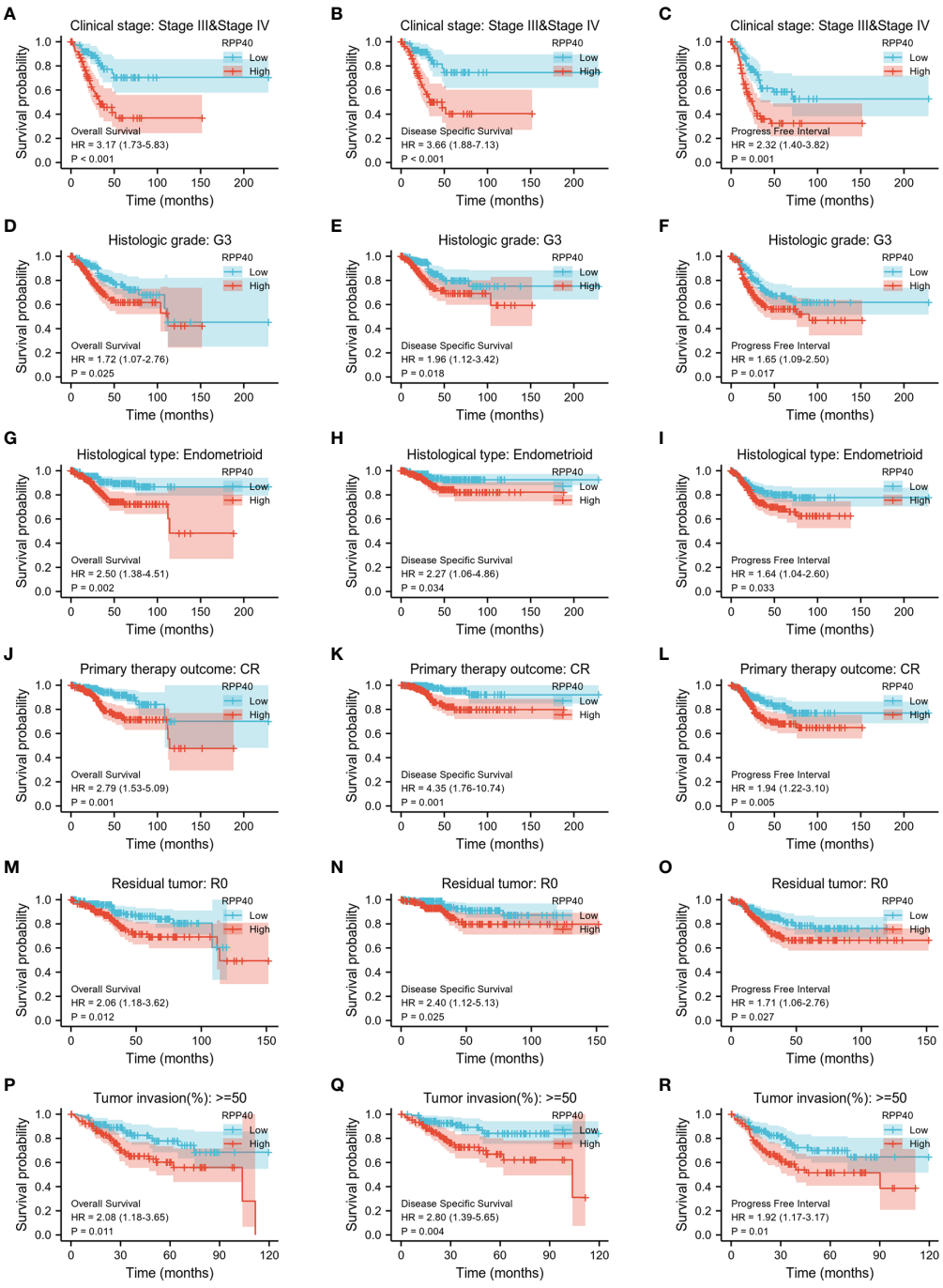


FIGURE 7
The association of clinical outcomes with RPP40 expression in UCEC patient form different subgroups based on clinicopathological factors. The result of K-M analysis showing distinct clinical outcomes of OS **(A, D, G, J, M, P)**, DSS **(B, E, H, K, N, Q)**, and PFI **(C, F, I, L, O, R)** between high- and low-RPP40-expression groups of UCEC patients in several subgroups, including clinical stage III–IV **(A-C)**, histological grade G3 **(D-F)**, the histological type of endometrioid **(G-I)**, the primary therapy outcome of CR **(J-L)**, residual tumor R0 **(M-O)**, and tumor invasion more than 50% of the muscular layer **(P-R)**.

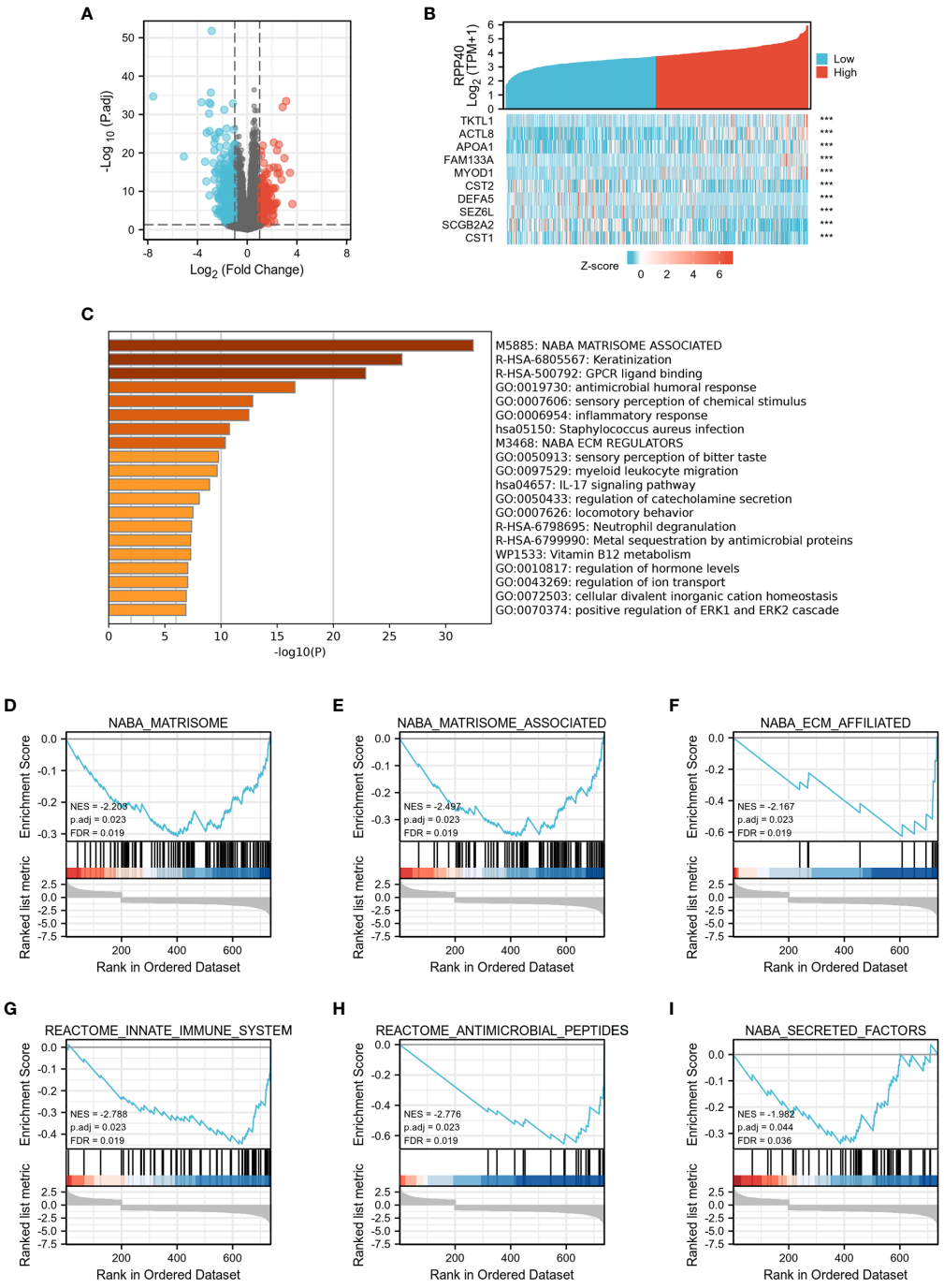


FIGURE 8
Identification and functional enrichment analysis of DEGs in UCEC patients with distinct RPP40 mRNA levels. The mRNAs of DEGs between two groups are presented by volcano plots **(A)**, and represented DEGs are shown as heatmaps **(B)**. Statistically enriched terms identified by the Metascape database are shown; the threshold of kappa score is set as 0.3; ***$P < 0.001$ **(C)**. Gene set enrichment analysis (GSEA) of differentially expressed mRNAs between high- and low-RPP40-expression groups in UCEC tumors has been conducted, and representative clusters are shown **(D-I)**.

mRNAs, Figure 8A and Supplementary Table 2), 90 microRNAs (miRNAs) (including 1 upregulated and 89 downregulated miRNAs, Supplementary Figure 4A and Supplementary Table 3), and 1,408 lncRNAs (including 131 upregulated and 1,277 downregulated lncRNAs, Supplementary Figure 4B, Supplementary Table 4) were screened out as DEGs in the high-RPP40 group. Representative DEGs were presented in heatmaps (Figure 8B).

To uncover the function of RPP40 in UCEC, functional enrichment analyses of DEGs were conducted. Firstly, as shown in Figure 8C and Supplementary Table 5, an online analysis *via* "Metascape" showed that several pathways associated with RPP40 were enriched, including "NABA_MATRISOME_ASSOCIATED", "Keratinization", "NABA_ECM_REGULATORS", "antimicrobial humoral response", "inflammatory response", "myeloid leukocyte migration", "IL-17 signaling pathway", "Neutrophil degranulation", "regulation of hormone levels", and "positive regulation of ERK1 and ERK2 cascade". This indicated that the function of RPP40 may be mainly related to the regulation of ECM, immune or inflammatory responses, and the ERK signaling pathway. Furthermore, as shown in Figures 8D–I and Supplementary Table 6, the result of GSEA analysis showed that RPP40-associated DEGs were mainly significantly enriched in ECM-related clusters (such as NABA_MATRISOME, NABA_MATRISOME_ASSOCIATED, NABA_ECM_AFFILIATED, and NABA_SECRETED_FACTORS), and immune system-related clusters (such as REACTOME_INNATE_IMMUNE_SYSTEM, and REACTOME_ANTIMICROBIAL_PEPTIDES). Moreover, GO and KEGG enrichment analyses (Supplementary File 2) also showed that the enriched biological processes, molecular functions, and pathways of RPP40 were closely related to ECM regulation and an immune or inflammatory response. These results indicated that the function of RPP40 in UCEC may be associated with the regulation of ECM and immune function.

Association of RPP40 and immune cell infiltration in uterine corpus endometrial carcinoma tumors

The possible association between RPP40 and the immune system was uncovered by the functional annotation of RPP40-associated DEGs. To further confirm the possible effect of RPP40 on tumor immunity, the relationship between RPP40 expression and immune cell infiltration in UCEC was firstly determined. As shown in Figure 9A, the infiltration of Th2 cells (R=0.310, $P<0.001$), Tcm (R=0.145, $P<0.001$), and T helper cells (R=0.183, $P<0.001$) were significantly positively correlated with RPP40 expression. In contrast, the tumor infiltration levels of NK CD56bright cells (R=−0.365, $P<0.001$), pDC (R=−0.347, $P<0.001$), iDC (R=−0.347, $P<0.001$), neutrophils (R=−0.322, $P<0.001$), NK cells (R=−0.175, $P<0.001$), TFH (R=−0.211, $P<0.001$), mast cells (R=−0.211, $P<0.001$), Treg (R=−0.242, $P<0.001$), cytotoxic cells (R=−0.193, $P<0.001$), Tem (R=−0.213, $P<0.001$), NK CD56dim cells (R=−0.179, $P<0.001$), eosinophils (R=−0.173, $P<0.001$), T cells (R=−0.165, $P<0.001$), Th17 cells (R=−0.118, $P<0.001$), CD8 T cells (R=−0.065, $P=0.004$), DC (R=−0.139, $P=0.009$), and B cells (R=−0.125, $P=0.013$) were all significantly negatively correlated with RPP40 expression levels. Moreover, the infiltration levels of 24 immune cell types in UCEC tumor tissues between high- and low-RPP40-expression groups were compared. As shown in Figures 9B–U, the infiltration levels of Th2 cells, Tcm, and T helper cells were significantly increased in the high-RPP40 group. At the same time, there were 17 immune cell types (including NK CD56bright cells, pDC, iDC, neutrophils, NK cells, TFH, mast cells, Treg, cytotoxic cells, Tem, NK CD56dim cells, eosinophils, T cells, Th17 cells, CD8 T cells, DC, and B cells) significantly decreased in the high-RPP40 group. In addition, the association between the RPP40 expression and gene marker levels of immune cells in UCEC tumor tissues was evaluated *via* TIMER, as shown in Table 4, the RPP40 expression level in UCEC tumor tissues was closely related to the immune marker expressions of $CD8^+$ T cells, T cells (general), monocytes, macrophages, neutrophils, NK cells, DC, and Th1 cells. These data indicate that RPP40 may play a specific role in the infiltration of immune cells in UCEC tumor tissues.

Discussion

Although patients with early clinical stages of UCEC have a relatively good prognosis, the patients with advanced or relapsed UCEC still respond poorly to conventional therapies (1, 2, 4). Therefore, the mining of novel prognostic biomarkers and therapeutic targets to improve the survival rate of UCEC patients is of great scientific interest and clinical importance.

At present, the molecular function of RPP40 remains unclear since there are few studies on it. As a component of RNase P or MRP, RPP25 has been reported to promote the proliferation, migration, invasion, and cell cycle programs of cervical cancer cells (14). Furthermore, both RPP25 and RPP30, another component of RNase P and MRP, were reported as reliable prognostic risk factors for glioblastoma multiforme (11, 13). Similarly, RPP40 was also regarded as one of the promoting factors for the chemoresistance of acute myeloid leukemia (15), and the member of a prognostic signature includes seven mRNAs and could accurately predict the recurrence risks of early-stage triple-negative breast cancer (16). In addition, the result of bioinformatics analysis in this study showed that RPP40 was one of the potential prognostic genes for UCEC (Supplementary File 1). Therefore, we speculated that RPP40 might be involved in the tumorigenesis or progression of UCEC.

In the present study, we found that the mRNA expression of RPP40 was significantly upregulated in the tumor tissues of various cancer types, especially in UCEC. Furthermore, the

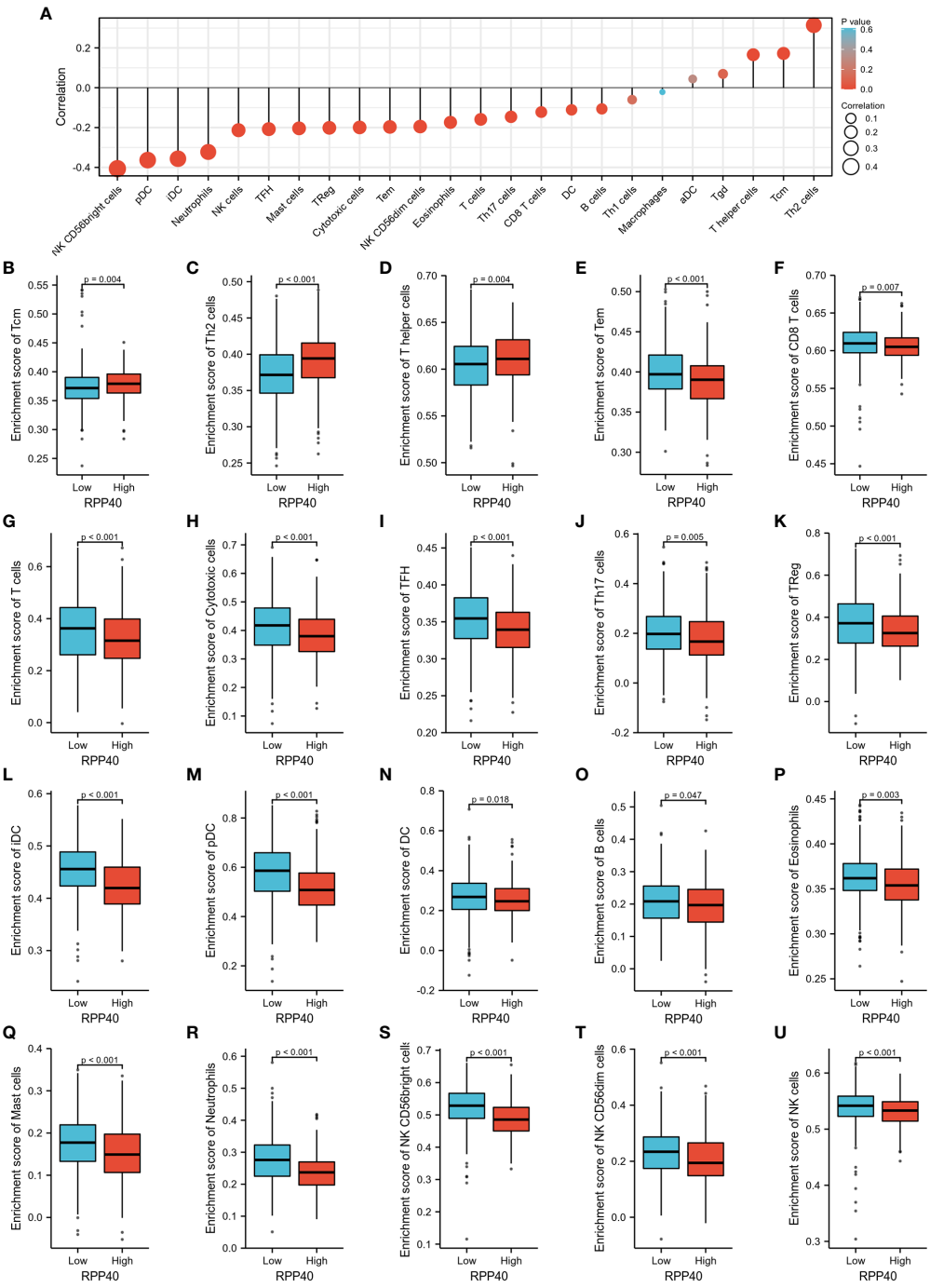


FIGURE 9
Relationships between RPP40 expression and immune cell infiltration in UCEC tumors. **(A)** The correlation of immune cell infiltration levels (24 cell types) and RPP40 mRNA expression was evaluated by Spearman's analysis. **(B–U)** The comparison of the infiltration levels of significantly correlated immune cells, including Tcm **(B)**, Th2 cells **(C)**, T helper cells **(D)**, Tem **(E)**, CD8 T cells **(F)**, T cells **(G)**, cytotoxic cells **(H)**, TFH **(I)**, Th17 cells **(J)**, Treg **(K)**, iDC **(L)**, pDC **(M)**, DC **(N)**, B cells **(O)**, eosinophils **(P)**, mast cells **(Q)**, neutrophils **(R)**, NK CD56bright cells **(S)**, NK CD56dim cells **(T)**, and NK cells **(U)** between high- and low-RPP40-expression groups of UCEC patients.

TABLE 4 Correlation analysis between RPP40 expression and immune cell markers in UCEC tumors.

Description	Gene markers	None		Purity	
		Cor	*P*-value	Cor	*P*-value
$CD8^+$ T cells	CD8A	-0.037	0.386	-0.058	0.325
	CD8B	-0.227	***	-0.27	***
T cells (general)	CD3D	-0.158	***	-0.157	**
	CD3E	-0.167	***	-0.192	***
	CD2	-0.054	0.205	-0.056	0.338
B cells	CD19	-0.057	0.182	0.042	0.474
	CD79A	-0.112	**	-0.081	0.165
Monocyte	CD86	0.019	0.665	0.02	0.733
	CD115(CSF1R)	-0.21	***	-0.144	*
TAM	CCL2	0.04	0.357	0.09	0.126
	CD68	-0.003	0.478	0.004	0.945
	IL10	-0.03	0.272	0.064	0.272
M1 macrophage	INOS(NOS2)	-0.096	*	-0.12	*
	IRF5	0.114	**	0.109	0.062
	COX2(PTGS2)	-0.123	**	-0.058	0.325
M2 macrophage	CD163	0.132	**	0.193	***
	VSIG4	0.005	0.899	0.092	0.114
	MS4A4A	0.065	0.129	0.111	0.058
Neutrophils	CD66b(CEACAM8)	-0.188	***	-0.154	**
	CD11b(ITGAM)	-0.201	***	-0.146	**
	CCR7	-0.169	***	-0.16	**
NK cells	KIR2DL1	-0.061	0.157	-0.073	0.21
	KIR2DL3	-0.085	*	-0.131	*
	KIR2DL4	-0.03	0.491	-0.043	0.465
	KIR3DL1	-0.112	**	-0.213	***
	KIR3DL2	-0.018	0.679	-0.095	0.105
	KIR3DL3	-0.055	0.202	-0.111	0.057
	KIR2DS4	-0.108	*	-0.183	**
DC	HLA-DPB1	-0.221	***	-0.206	***
	HLA-DQB1	-0.201	***	-0.209	***
	HLA-DRA	-0.137	**	-0.129	*
	HLA-DPA1	-0.126	**	-0.111	0.057
	BDCA-1(CD1C)	-0.265	***	-0.229	***
	BDCA-4(NRP1)	-0.041	0.337	0.018	0.757
	CD11c(ITGAX)	-0.184	***	-0.164	**
Th1 cells	T-bet (TBX21)	-0.054	0.209	-0.051	0.384
	STAT4	-0.092	*	-0.064	0.271
	STAT1	0.374	***	0.367	***
	IFN-γ(IFNG)	0.025	0.554	0.014	0.814
	TNF-α(TNF)	0.085	*	0.138	*
Th2 cells	GATA3	-0.093	*	0.006	0.918
	STAT6	-0.081	0.06	0.056	0.342
	STAT5A	-0.05	0.247	0.004	0.948
	IL13	-0.039	0.369	-0.009	0.881
TFH	BCL6	-0.047	0.269	-0.049	0.404
	IL21	-0.04	0.349	-0.045	0.444
Th17 cells	STAT3	0.043	0.317	0.107	0.067
	IL17A	0.015	0.718	-0.018	0.754

*P < 0.05, **P < 0.01, ***P < 0.001. Cor: Spearman's rho value; None: no adjusted correlation; Purity: correlation adjusted by tumor purity. The results were based on TIMER database analysis.

protein expression of RPP40 is also significantly upregulated in UCEC tumor tissues. In addition, RPP40 expression was positively correlated with the clinical stage, histological grade, histological type, and primary therapy outcome. Based on these observations, we speculate that RPP40 might be a potential biomarker and therapeutic target of UCEC. To verify this hypothesis, we evaluated the predictive values of RPP40 in the diagnosis and prognosis of UCEC patients; the results showed that RPP40 was an effective predictor for the diagnosis of UCEC with an AUC of 0.775, a sensitivity of 0.829, and a specificity of 0.719. Furthermore, RPP40 also possessed a significant prognostic value independent of clinicopathological factors in UCEC patients, and the patients with low RPP40 expression possess significantly better clinical outcomes than those with high RPP40 expression. Therefore, we considered RPP40 as a promising prognostic biomarker for UCEC. However, studies targeting the function of RPP40 in malignant tumors are rarely reported.

The TME, composed of multiple cellular and molecular factors, has been widely implicated in tumorigenesis, progression, metastasis, and therapeutic resistance (5–7). Various components of TME, such as cancer-associated fibroblasts (CAFs), immune cells, extracellular matrix (ECM), cytokines, chemokines, and other soluble factors, act together to influence antitumor immunity, therapeutic response, and clinical outcomes (5–8). As an essential component of TME, ECM regulates cell proliferation and differentiation, and its remodeling contributes to tumor growth and metastasis (29, 30). CAFs, the main productor of ECM, interact with almost all cells within the TME that could enable them to promote the tumorigenic alterations of ECM components (29–31). Studies have confirmed that ECM stiffness and degradation always result in the proliferation, migration, and invasion of cancer cells (29). ECM stiffness was mainly regulated by integrin and transforming growth factor-β (TGF-β)–related pathways. ECM degradation was regulated primarily by matrix metalloproteinases (MMPs)/tissue inhibitors of MMP (TIMPs)-related pathways; both of these pathways have been reported to contribute to

cancer cell invasion and metastasis (29, 32–34). Until now, there have been no relevant studies about the regulatory role of RPPs on ECM remodeling. In this study, "Metascape" analysis showed that several ECM-related pathways associated with RPP40 were enriched, including "NABA_MATRISOME_ASSOCIATED" and "NABA_ECM_REGULATORS". Furthermore, the result of GSEA analysis showed that RPP40-associated DEGs were mainly significantly enriched in ECM-related clusters, such as NABA_MATRISOME, NABA_MATRISOME_ASSOCIATED, NABA_ECM_AFFILIATED, and NABA_SECRETED_FACTORS. In addition, the RPP40 expression level was significantly associated with the expression levels of ECM-related genes. In particular, RPP40 expression was positively correlated to the expressions of TGFB2, SMAD2, ITGA1, ITGB1, ITGB5, MMP1, and MMP12, and negatively correlated to the expressions of COL1A1, COL3A1, COL6A2, TGFB1, and TIMP1 (Supplementary Figure 5), suggesting that ECM stiffness and degradation might occur in the UCEC tumors of high-RPP40 patients. These results indicated that the function of RPP40 in UCEC might be closely related to ECM dysregulation in the TME.

Tumor-infiltrating immune cells and the cytokines, chemokines, and other soluble factors secreted by them are also crucial components of the TME (5, 7, 8). Most tumor cells express antigens that can mediate recognition by immune cells and then promote immune cell infiltration and activate the tumor immunity (35). Existing studies confirmed that tumor-infiltrating immune cells are closely associated with the clinical outcome of cancer patients (36, 37). Meanwhile, tumor cells can alter the TME and then induce immune escape and adaptive immune tolerance, which are currently considered essential for the metastases, recurrence, and therapeutic resistance of malignant tumors (5, 38). In patients with systemic autoimmune rheumatic disease, almost all RNase P and MRP complexes' components have been reported as autoantibody targets (39–41). In addition, the expression of RPP25 was strongly correlated with immune cell infiltration levels in glioblastoma multiforme (11). Similarly, in present research, an online analysis of "Metascape" showed that several pathways associated with RPP40 were enriched, including "antimicrobial humoral response", "inflammatory response", "myeloid leukocyte migration", "IL-17 signaling pathway", "Neutrophil degranulation", and "regulation of hormone levels". Furthermore, the result of GSEA analysis showed that RPP40-associated DEGs were mainly significantly enriched in immune system–related clusters, such as "REACTOME_INNATE_IMMUNE_SYSTEM" and "REACTOME_ANTIMICROBIAL_PEPTIDES". Moreover, the upregulation of RPP40 was significantly negatively correlated with the tumor infiltration levels of most of immune cell types, such as NK cells, DCs, cytotoxic cells, and CD 8 T cells. DCs are a group of specialized antigen-presenting cells; CD 8 T cells are essential cancer antigen recognition cells that act together and have critical roles in initiating and regulating anti-tumor immune responses (42, 43). NK cells and cytotoxic cells are important effectors of antitumor immunity and can directly kill cancer cells (44, 45). These results indicated that the function of RPP40 in UCEC might also be closely related to the TME.

Based on the above results, we believe that RPP40 is a promising prognostic biomarker correlated with the TME in UCEC. Meanwhile, the mechanism underlying the regulatory function of RPP40 on the TME is still not clear. We notice that RPP40-related DEGs were also significantly enriched in the "positive regulation of ERK1 and ERK2 cascade" in the functional annotation analysis based on "Metascape". As protein-serine/threonine kinases, both ERK1 and ERK2 are essential components of the Ras-Raf-MEK-ERK signaling cascade, which has been reported to regulate cell proliferation, survival, differentiation, metabolism, adhesion, and migration (46). In malignant tumors, the ERK signaling pathway has been confirmed to promote the transformation of fibroblasts to CAFs in colorectal cancer (47). Furthermore, the ERK1/2 signaling pathway has also been reported as a promoting factor of tumor ECM degradation and angiogenesis, contributing to the proliferation, invasion, and metastasis of malignant tumors (48). Similarly, ERK1/2 signaling cascade has been proven to regulate the tumor immune microenvironment by recruiting immune cells in glioblastoma (49). Therefore, we speculate that the regulatory mechanism of RPP40 in the TME of UCEC may be closely related to the regulation of ERK signaling pathways, whereas further verification studies are needed.

Although we revealed a potential role and the possible mechanism of RPP40 in UCEC tumorigenesis and prognosis, there are still several limitations in this research. Firstly, we just evaluated the association of RPP40 expression and the expression of ECM-related genes in UCEC tumors based on the TCGA database, while CAFs are the main productor of ECM. Therefore, the association analysis between RPP40 expression in tumor cells and the ECM-related gene expressions in CAFs is more convincing. Secondly, further *in vivo* and *in vitro* experiments, and the confirming studies in protein levels are all needed to verify the effect and direct mechanism of RPP40 in UCEC.

Conclusions

The upregulation of RPP40 might play an important role in the tumorigenesis and progression of UCEC by regulating the TME and exhibiting a reliable diagnostic and prognostic value for clinical outcomes. The results of this study indicate the possibility of RPP40 as a promising biomarker and therapeutic target for UCEC.

Data availability statement

All data used in this research was acquired from the TCGA database. This data can be found online: https://portal.gdc.cancer.gov/. 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

JT, XT, JM and MH: project investigation. JT and XT: methodology. JT and XT: writing–original draft. JT and LH: writing–review and editing. LH: project administration and supervision. All authors contributed to this article and approved the submitted version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82001527), and Open Project of Hubei Key Laboratory from Renmin Hospital of Wuhan University (No. 2021KFY003).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.957472/full#supplementary-material

References

1. Liu W, Sun L, Zhang J, Song W, Li M, Wang H. The landscape and prognostic value of immune characteristics in uterine corpus endometrial cancer. *Biosci Rep* (2021) 41(4):BSR20202321. doi: 10.1042/BSR20202321

2. Romero I, Rubio MJ, Medina M, Matias-Guiu X, Santacana M, Schoenenberger JA, et al. An olaparib window-of-Opportunity trial in patients with early-stage endometrial carcinoma: Polen study. *Gynecol Oncol* (2020) 159 (3):721–31. doi: 10.1016/j.ygyno.2020.09.013

3. Mahdy H, Casey MJ, Crotzer D. *Endometrial cancer*. Treasure Island (FL: Statpearls (2022).

4. Zhang M, Liu Y, Hou S, Wang Y, Wang C, Yin Y, et al. Kdm4b, a potential prognostic biomarker revealed by Large-scale public databases and clinical samples in uterine corpus endometrial carcinoma. *Mol Omics* (2022) 18(6):506–19. doi: 10.1039/d1mo00287b

5. Arneth B. Tumor microenvironment. *Medicina (Kaunas)* (2019) 56(1):15. doi: 10.3390/medicina56010015

6. Ino Y, Yamazaki-Itoh R, Shimada K, Iwasaki M, Kosuge T, Kanai Y, et al. Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. *Br J Cancer* (2013) 108(4):914–23. doi: 10.1038/bjc.2013.32

7. Xiao Y, Yu D. Tumor microenvironment as a therapeutic target in cancer. *Pharmacol Ther* (2021) 221:107753. doi: 10.1016/j.pharmthera.2020.107753

8. Soysal SD, Tzankov A, Muenst SE. Role of the tumor microenvironment in breast cancer. *Pathobiology* (2015) 82(3-4):142–52. doi: 10.1159/000430499

9. Jarrous N, Eder PS, Guerrier-Takada C, Hoog C, Altman S. Autoantigenic properties of some protein subunits of catalytically active complexes of human ribonuclease p. *RNA* (1998) 4(4):407–17.

10. Wu J, Niu S, Tan M, Huang C, Li M, Song Y, et al. Cryo-em structure of the human ribonuclease p holoenzyme. *Cell* (2018) 175(5):1393–404.e11. doi: 10.1016/j.cell.2018.10.003

11. Xiao D, Wu J, Zhao H, Jiang X, Nie C. Rpp25 as a prognostic-related biomarker that correlates with tumor metabolism in glioblastoma. *Front Oncol* (2021) 11:714904. doi: 10.3389/fonc.2021.714904

12. Goldfarb KC, Cech TR. Targeted crispr disruption reveals a role for rnase mrp rna in human preribosomal rna processing. *Genes Dev* (2017) 31(1):59–71. doi: 10.1101/gad.286963.116

13. Li G, Zhai Y, Liu H, Wang Z, Huang R, Jiang H, et al. Rpp30, a transcriptional regulator, is a potential pathogenic factor in glioblastoma. *Aging (Albany NY)* (2020) 12(16):16155–71. doi: 10.18632/aging.103596

14. Feng L, Zhao M, Wu A. Circasap1 promotes the development of cervical cancer through sponging mir-338-3p to upregulate Rpp25. *Anticancer Drugs* (2022) 33(1):e155–e65. doi: 10.1097/CAD.0000000000001167

15. Aasebo E, Berven FS, Hovland R, Doskeland SO, Bruserud O, Selheim F, et al. The progression of acute myeloid leukemia from first diagnosis to chemoresistant relapse: A comparison of proteomic and phosphoproteomic profiles. *Cancers (Basel)* (2020) 12(6):1466. doi: 10.3390/cancers12061466

16. Yang YS, Ren YX, Liu CL, Hao S, Xu XE, Jin X, et al. The early-stage triple-negative breast cancer landscape derives a novel prognostic signature and therapeutic target. *Breast Cancer Res Treat* (2022) 193(2):319–30. doi: 10.1007/s10549-022-06537-z

17. Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-Spectrometry-Based proteomic characterization of more than 500 human cancers. *Nat Commun* (2019) 10 (1):5679. doi: 10.1038/s41467-019-13528-0

18. Zhang Y, Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways. *Nat Commun* (2022) 13(1):2669. doi: 10.1038/s41467-022-30342-3

19. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science* (2015) 347 (6220):1260419. doi: 10.1126/science.1260419

20. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science* (2017) 357(6352): eaan2507. doi: 10.1126/science.aan2507

21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. Proc: An open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinf* (2011) 12:77. doi: 10.1186/1471-2105-12-77

22. 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

23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with Deseq2. *Genome Biol* (2014) 15(12):550. doi: 10.1186/s13059-014-0550-8

24. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6

25. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A.* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

27. Hanzelmann S, Castelo R, Guinney J. Gsva: Gene set variation analysis for microarray and rna-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

28. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003

29. Najafi M, Farhood B, Mortezaee K. Extracellular matrix (Ecm) stiffness and degradation as cancer drivers. *J Cell Biochem* (2019) 120(3):2782–90. doi: 10.1002/jcb.27681

30. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep* (2014) 15(12):1243–53. doi: 10.15252/embr.201439246

31. Walker C, Mojares E, Del Rio Hernandez A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci* (2018) 19(10):3028. doi: 10.3390/ijms19103028

32. Nolte M, Margadant C. Controlling immunity and inflammation through integrin-dependent regulation of tgf-beta. *Trends Cell Biol* (2020) 30(1):49–59. doi: 10.1016/j.tcb.2019.10.002

33. Jablonska-Trypuc A, Matejczyk M, Rosochacki S. Matrix metalloproteinases (Mmps), the main extracellular matrix (Ecm) enzymes in collagen degradation, as a target for anticancer drugs. *J Enzyme Inhib Med Chem* (2016) 31(sup1):177–83. doi: 10.3109/14756366.2016.1161620

34. Salimi Sartakhti J, Manshaei MH, Sadeghi M. Mmp-timp interactions in cancer invasion: An evolutionary game-theoretical framework. *J Theor Biol* (2017) 412:17–26. doi: 10.1016/j.jtbi.2016.09.019

35. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* (2013) 14(10):1014–22. doi: 10.1038/ni.2703

36. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. *BioMed Pharmacother* (2017) 95:55–61. doi: 10.1016/j.biopha.2017.08.003

37. Zhang SC, Hu ZQ, Long JH, Zhu GM, Wang Y, Jia Y, et al. Clinical implications of tumor-infiltrating immune cells in breast cancer. *J Cancer* (2019) 10(24):6175–84. doi: 10.7150/jca.35901

38. Ferrone S, Whiteside TL. Tumor microenvironment and immune escape. *Surg Oncol Clin N Am* (2007) 16(4):755–74. doi: 10.1016/j.soc.2007.08.004

39. Kuwana M, Kimura K, Hirakata M, Kawakami Y, Ikeda Y. Differences in autoantibody response to th/to between systemic sclerosis and other autoimmune diseases. *Ann Rheum Dis* (2002) 61(9):842–6. doi: 10.1136/ard.61.9.842

40. Van Eenennaam H, Vogelzangs JH, Bisschops L, Te Boome LC, Seelig HP, Renz M, et al. Autoantibodies against small nucleolar ribonucleoprotein complexes and their clinical associations. *Clin Exp Immunol* (2002) 130(3):532–40. doi: 10.1046/j.1365-2249.2002.01991.x

41. Van Eenennaam H, Vogelzangs JH, Lugtenberg D, Van Den Hoogen FH, Van Venrooij WJ, Pruijn GJ. Identity of the rnase mrp- and rnase p-associated th/to autoantigen. *Arthritis Rheum* (2002) 46(12):3266–72. doi: 10.1002/art.10673

42. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent advances in targeting Cd8 T-cell immunity for more effective cancer immunotherapy. *Front Immunol* (2018) 9:14. doi: 10.3389/fimmu.2018.00014

43. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. *Nat Rev Immunol* (2020) 20(1):7–24. doi: 10.1038/s41577-019-0210-z

44. Oh DY, Fong L. Cytotoxic Cd4(+) T cells in cancer: Expanding the immune effector toolbox. *Immunity* (2021) 54(12):2701–11. doi: 10.1016/j.immuni.2021.11.015

45. Wu SY, Fu T, Jiang YZ, Shao ZM. Natural killer cells in cancer biology and therapy. *Mol Cancer* (2020) 19(1):120. doi: 10.1186/s12943-020-01238-x

46. Roskoski RJr. Erk1/2 map kinases: Structure, function, and regulation. *Pharmacol Res* (2012) 66(2):105–43. doi: 10.1016/j.phrs.2012.04.005

47. Wang L, Yang D, Tian J, Gao A, Shen Y, Ren X, et al. Tumor necrosis factor receptor 2/Akt and erk signaling pathways contribute to the switch from fibroblasts to cafs by progranulin in microenvironment of colorectal cancer. *Oncotarget* (2017) 8(16):26323–33. doi: 10.18632/oncotarget.15461

48. Guo YJ, Pan WW, Liu SB, Shen ZF, Xu Y, Hu LL. Erk/Mapk signalling pathway and tumorigenesis. *Exp Ther Med* (2020) 19(3):1997–2007. doi: 10.3892/etm.2020.8454

49. Lailler C, Louandre C, Morisse MC, Lhossein T, Godin C, Lottin M, et al. Erk1/2 signaling regulates the immune microenvironment and macrophage recruitment in glioblastoma. *Biosci Rep* (2019) 39(9):BSR20191433. doi: 10.1042/BSR20191433

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 31 August 2022
DOI 10.3389/fonc.2022.976959

OPEN ACCESS

EDITED BY
Jorge Melendez-Zajgla,
Instituto Nacional de Medicina Genómica (INMEGEN), Mexico

REVIEWED BY
Rachel Ellsworth,
Walter Reed National Military Medical Center, United States
Yasmine Kanaan,
Howard University, United States

*CORRESPONDENCE
Dirce Maria Carraro
dirce.carraro@accamargo.org.br

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

RECEIVED 23 June 2022
ACCEPTED 10 August 2022
PUBLISHED 31 August 2022

CITATION
Paixão D, Torrezan GT, Santiago KM, Formiga MN, Ahuno ST, Dias-Neto E, Tojal da Silva I, Foulkes WD, Polak P and Carraro DM (2022) Characterization of genetic predisposition to molecular subtypes of breast cancer in Brazilian patients. *Front. Oncol.* 12:976959. doi: 10.3389/fonc.2022.976959

Characterization of genetic predisposition to molecular subtypes of breast cancer in Brazilian patients

Daniele Paixão[1], Giovana Tardin Torrezan[2,3], Karina Miranda Santiago[2], Maria Nirvana Formiga[1], Samuel Terkper Ahuno[4], Emmanuel Dias-Neto[3,5], Israel Tojal da Silva[3,6], William D. Foulkes[7], Paz Polak[8] and Dirce Maria Carraro[2,3]*

[1]Oncogenetics Department, A.C.Camargo Cancer Center, São Paulo, SP, Brazil, [2]Clinical and Functional Genomics Group, International Research Center/CIPE, A.C.Camargo Cancer Center, São Paulo, SP, Brazil, [3]National Institute of Science and Technology in Oncogenomics and Therapeutic Innovation (INCITO), São Paulo, SP, Brazil, [4]Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY, United States, [5]Genomic Medicine Group, - International Research Center/CIPE, A.C.Camargo Cancer Center, São Paulo, SP, Brazil, [6]Bioinformatics and Computational Biology Group, - International Research Center/CIPE, A.C.Camargo Cancer Center, São Paulo, SP, Brazil, [7]Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montreal, QC, Canada, [8]Computational Biology, C2i Genomics, New York, NY, United States

Introduction: *BRCA1* and *BRCA2* germline pathogenic variants (GPVs) account for most of the 5-10% of breast cancer (BC) that is attributable to inherited genetic variants. *BRCA1* GPVs are associated with the triple negative subtype, whereas *BRCA2* GPVs are likely to result in higher grade, estrogen-receptor positive BCs. The contribution of other genes of high and moderate risk for BC has not been well defined and risk estimates to specific BC subtypes is lacking, especially for an admixed population like Brazilian.

Objective: The aim of this study is to evaluate the value of a multigene panel in detecting germline mutations in cancer-predisposing genes for Brazilian BC patients and its relation with molecular subtypes and the predominant molecular ancestry.

Patients and Methods: A total of 321 unrelated BC patients who fulfilled NCCN criteria for *BRCA1/2* testing between 2016-2018 were investigated with a 94-genes panel. Molecular subtypes were retrieved from medical records and ancestry-specific variants were obtained from off-target reads obtained from the sequencing data.

Results: We detected 83 GPVs in 81 patients (positivity rate of 25.2%). Among GPVs, 47% (39/83) were identified in high-risk BC genes (*BRCA1/2, PALB2* and *TP53*) and 18% (15/83) in moderate-penetrance genes (*ATM, CHEK2* and *RAD51C*). The remainder of the GPVs (35% - 29/83), were identified in lower-

risk genes. As for the molecular subtypes, triple negative BC had a mutation frequency of 31.6% (25/79), with predominance in *BRCA1* (12.6%; 10/79). Among the luminal subtypes, except Luminal B HER2-positive, 18.7% (29/155) had GPV with *BRCA1/2* genes contributing 7.1% (11/155) and non-*BRCA1/2* genes, 12.9% (20/155). For Luminal B HER2-positive subtype, 40% (16/40) had GPVs, with a predominance of *ATM* gene (15% - 6/40) and *BRCA2* with only 2.5% (1/40). Finally, HER2-enriched subtype presented a mutation rate of 30.8% (4/13) with contribution of *BRCA2* of 7.5% (1/13) and non-*BRCA1/2* of 23% (3/13). Variants of uncertain significance (VUS) were identified in 77.6% (249/321) of the patients and the number of VUS was increased in patients with Asian and Native American ancestry.

Conclusion: The multigene panel contributed to identify GPVs in genes other than *BRCA1/2*, increasing the positivity of the genetic test from 9.6% (*BRCA1/2*) to 25.2% and, considering only the most clinically relevant BC predisposing genes, to 16.2%. These results indicate that women with clinical criteria for hereditary BC may benefit from a multigene panel testing, as it allows identifying GPVs in genes that directly impact the clinical management of these patients and family members.

KEYWORDS

breast cancer, hereditary cancer, multigene panel, cancer genetics, molecular subtype of breast cancer

Introduction

Breast cancer is the most common non-cutaneous cancer, and according to the World Health Organization, it is the second leading cause of cancer death among women worldwide (1). Between 5-10% of BCs are attributed to inherited genetic variations mainly in two high-risk genes, *BRCA1* and *BRCA2*, associated to the hereditary breast and ovarian cancer syndrome (HBOC), which confer a high risk of breast, ovarian, pancreatic and prostate cancer (2–4). However, a significant proportion of the suspected genetic risk patients remains unexplained when only the two genes, *BRCA1/2*, are investigated. Apart from *BRCA1/2*, GPVs located in seven other genes - *ATM, BARD1, CHEK2, PALB2, RAD51C, RAD51D* and *TP53* – have shown to be clinically relevant, increasing the risk to develop BC (5–9). For *ATM, CHEK2* and *PALB2* appropriate evidences have been gathered in the clinical setting to warrant the screening for GPVs in these genes, even in the absence of familial BC history (10–12). Additionally, the role of widespread clinical testing for GPVs in other BC-risk genes, such as *CDH1, STK11* and *PTEN* that increase the risk for BC in the context of Hereditary Diffuse Gastric Cancer, Peutz-Jeghers and Cowden's syndrome, respectively, continues to be debated (9).

The implementation of genetic testing for multiple genes for hereditary cancer syndromes offers many benefits, including lower cost and time per gene when compared to single-gene testing (13–16). Currently available commercial multigene panels range widely from phenotype specific, for familial cancer such as BC, to panels covering multiple phenotypes. These panels may include high-risk genes, with established clinical utility, as well as moderate and low-risk genes, with limited data about clinical significance and cancer risk and even genes with no management guidelines (16, 17).

Important issues have been widely discussed about the clinical indications, benefits and genetic counseling impact of multigene panels (18). In general, these panels are indicated when more than one gene may be associated to the phenotype, due to its increased efficacy and reasonable cost as compared to single genes (16). Indications can also be considered for patients with a negative test for particular syndromes, whose personal and familial history may suggest hereditary cancer (13, 19). Nevertheless, nowadays, National Comprehensive Cancer Network (NCCN) guidelines recommend high-penetrance susceptibility gene analysis, beyond *BRCA1/2*, for BC patients with testing criteria (20).

The use of multigene panels in the clinical practice still faces some challenges. These include the proper interpretation of the sometimes complex results, such as the identification of variants of uncertain significance (VUS), specially concerning in populations that have been less characterized by genomic

studies, as well as the find of unexpected results like variants without genotype-phenotype correlation and of potentially GPVs in moderate and low-risk genes, which makes genetic counseling and clinical management more challenging (11, 16, 17). As an example, after the implementation of multigene panels in suspected BC risk patients, some studies reported that the identification of at least one VUS in different cohorts varied from 20% to 40% (13, 19). Still, studies with multigene panels have shown that patients with suspected hereditary breast and ovarian cancer and negative for *BRCA1/2*, presented a prevalence of mutations in other genes ranging from 4% to 16%, substantially increasing the ability to discovery the genetic cause for the increased cancer risk in these patients (18, 19).

The aim of the study presented here is to evaluate the impact of the use of a multigene panel in clinical practice of patients suspected of BC risk in both, overall and subtype-specific BC scenarios, and to evaluate the VUS repertoire in groups distinct predominant ancestries. To this end, we used a 94-genes panel in a series of BC patients and compared the identified clinically relevant variants to clinical, pathological and ancestry data. Our results contribute to the understanding of the genetic architecture of BC risk in a very admixed and scarcely genetic characterized population.

Methods

Patients selection

We selected a total of 321 unrelated patients diagnosed with BC, all under investigation at the Department of Oncogenetics, A.C.Camargo Cancer Center (ACCCC), between September 2016 and May 2018. Inclusion criteria: patients with a current or previous BC diagnosed at any age, of any histological type (including bilateral BC), who fulfilled NCCN criteria (2016 to 2018) of HBOC syndrome and performed genetic test at the Genomic Diagnostic Laboratory/Pathological Anatomy of the ACCCC. All patients received pre- and post-testing genetic counseling.

All patients have signed a written informed consent after genetic counseling. This study was approved by the local Ethics Committee of the ACCCC (protocol number 2483/18).

Genetic testing

Genomic DNA from peripheral blood sample or saliva was extracted and was used to perform capture by hybridization of the exons and exon-intron boundaries of the 94 genes using the commercial kits TruSight Cancer Sequencing Panel and TruSight Rapid Capture (Illumina). Next generation sequencing (NGS) was performed on the NextSeq 500 System (Illumina) platform. Sequences corresponding to the requested genes of each patient were compared with the respective reference sequences for calling variants using bioinformatics tools (Isaac Enrichment v3.0 and Illumina Variant Studio 2.2).

All identified variants were imported into the VarSeq software (Golden Helix) for function, classification and frequency annotations in public databases. Variants were filtered according to the criteria: quality >30; variant base present in at least 25% of the reads; absent in population databases (gnomAD, dbSNP, 1000genomes and Abraom - database of variants of exomes of the Brazilian population: http://abraom.ib.usp.br/) or, when present, presenting minor allele frequencies (MAF) ≤ 0.01.

Multiplex Ligation-dependent Probe Amplification (MLPA – P087, MRC-Holland, Amsterdam, NL) was used for *BRCA1* and *BRCA2* copy number variation analysis, according to the manufacturer's recommendations. Coffalyzer software (MRC-Holland, Amsterdam, NL) was used at default settings for data analyses.

Variant classification and analysis

The variants were noted for their increased changes of impacting protein function: loss of function (LoF) changes, indels and mutations at canonical splice sites; indels disrupting reading frames; amino-acid substitution variants (missenses) and synonymous alterations, and evaluated in the ClinVar database. For the classification and final interpretation of the identified variants we have followed the recommendations of the American College of Medical Genetics and Genomics (ACMG) (21).

Ancestry analysis

Analysis of African, European, Native/Latin American, and South/East Asian ancestries were performed using sequencing data obtained from the TruSight Cancer, containing data from the 94 cancer predisposing genes as well as all off-target reads. The data was processed using the software PLINK and a set of quality control criteria was applied (22). First, SNPs with call rates across samples < 95% and minor allele frequency (MAF) < 1% were filtered out. Then, SNPs were excluded for Hardy-Weinberg Equilibrium test < 0.000001 and pruned for linkage disequilibrium for window size = 50, step size = 5 and r2threshold = 0.2. For quality control, the set of selected SNPs was applied to the reference dataset and the unsupervised mode of ADMIXTURE was used. The values obtained were then compared with the ancestry previously calculated. This comparison was made using the graphical method Bland-Altman (23). As a reference, populations from 1000 Genomes

Project (1000G) and Human Genome Diversity Project (HGDP) were extracted (24–26). Altogether eight populations were selected from 1000G and these were grouped into four superpopulations: European, African, Native American and Asian. Only the most homogenous populations were selected. This estimate was obtained using the unsupervised mode of ADMIXTURE with K=4 (27). The chosen populations were: YRI (Yoruba in Ibadan, Nigeria), LWK (Luhya in Webuye, Kenya), GBR (British in England and Scotland), TSI (Toscani in Italia), JPT (Japanese from Tokyo, Japan), CHB (Han Chinese from Beijing, China), ITU (Indian Telugu from UK) and STU (Sri Lankan Tamil from the UK). In HGDP, five populations were selected, and classified as Native Americans. From all our subjects, a total of 534,734 SNPs were found.

The classification in categorical variables was performed similarly to two previous studies (28, 29). Individuals that had the secondary ancestry with less than 20% were noted as the first ancestry (e.g., an individual with 78% European ancestry, 15% African and 7% native American was annotated as predominantly EUR). Individuals that had greater than 20% of the secondary ancestry were classified as admixed samples, and both the primary and secondary most relevant ancestries were noted (e.g., an individual with 65% European ancestry and 35% African ancestry was annotated as EUR_admixAFR). Individuals without any ancestry above 50% were noted as highly admixed.

Statistical analysis

Clinical, anatomopathological and familial characteristics were described with descriptive statistics, including medians, means and standard deviations for continuous data. For categorical data, proportions with a 95% confidence interval (CI) were calculated using the Clopper-Pearson method. The demographic, clinical and pathological data were compared by the T test (continuous variables) and the T test/analysis of variance for continuous variables. Statistical significance was set at a $p \leq 0.05$.

Breast cancer (BC) molecular subtypes were classified according to Immunohistochemistry (IHC) status of Progesterone/Estrogen receptor and HER2 protein overexpression by IHC/amplified by FISH. Ki67 status was not available in our records. Thus we classified BC in three molecular subtypes: Triple-negative breast cancer (TNBC), Luminal (HR-positive/HER2-negative), Luminal B HER2 (HR-positive/HER2-positive) and HER2-enriched (HR-negative/HER2-positive).

To investigate the predictors of number of VUS, we used a multivariate poison regression. First set number of VUS as response variable and used all predictors (“European”, “African”, “Asian”, “America”, “Ashkenazi Jewish ethnicity”, “Age at diagnosis years”, “ Family history of breast & ovarian cancer”, “ Family history of non-breast & ovarian cancer”) in building a model.

In order to be able to select the features, which could well serve as predictors, we used backward stepwise approach where the least significant variable is gradually eliminated until we get a final model with relatively lower AIC values. To take a more robust approach, we bootstrapped the backward stepwise elimination approach running it for (n) number of times each time with subset of (n) samples from original dataset with replacement. We repeated the process with a) all patients are selected, b) selected patients with VUS in genes that form top 10% in the cohort, c) patients with VUS of 10 breast and ovarian cancer genes previously reported (30). For GPVs, we used a multivariate binomial logistic regression model of all predictors followed by bootstrap backward stepwise feature selection process. The response variable in this exercise was the presence or absence of GPVs.

Results

Germline characterization in breast cancer patients with overall or specific subtype

Among 321 women included, the mean age at diagnosis was 45.2 years (ranging from 26 to 85 years), and the median was 44 years, just over half of them developed BC at age 45 years or younger (57%) and 29 (9%) had bilateral disease. Only 1.2% of the studied population declared to have Ashkenazi Jewish ancestry. The predominant histological type was ductal carcinoma, found in 82.6% of the cases. Triple-negative breast cancer (TNBC) was found in 24.6% of the patients, Luminal subtype in 48.3%, Luminal B HER2 in 12.5%, HER2-enriched in 4.0%. Tumor subtype data was not available for 10.6% (Table 1). We observed that 12.1% of women had diagnosis of another cancer, and of these, 2.8% had a diagnosis of ovarian cancer (5 of them had diagnosis of ovarian after BC, one was synchronized and two had ovary before BC). Most women (65.1%) had a first, second or third degree relative with breast or ovary cancer. Clinical and pathological findings for all patients are given in Table 1.

Using a 94-genes panel a total of 83 GPVs in 24 genes were found in 81 women (25.2%). Most GPVs were LoF (78.3%): 32 frameshift variants (38.6%), 18 nonsense variants (21.7%) and 10 splice site variants (12%), and 21.7% were missense variants. The frequency of GPVs among 321 women was 25.2% (81/321) (Table 2). Two patients had more than one GPV and were diagnosed with the Multilocus Inherited Neoplasia Alleles Syndrome (MINAS), involving a combination of mutations in *BRCA2*/*ATM* and *BRCA2*/*CHEK2* genes.

GPVs in genes of high-penetrance for BC were found in 39 patients (12.15% - 39/321) including: 17 (5.3%) in *BRCA1*; 14 in

TABLE 1 Clinical and anatomopathological characteristics found in the 321 studied individuals.

Study Characteristic (N = 321)	No.	%
Age at diagnosis, years		
Mean ± SD	45,21 ± 11,22	
Median	44	
Range	26-85	
≤45	183	57.0
46-60	106	33.0
>60	32	10.0
Ashkenazi Jewish ethnicity		
Yes	4	1.2
No	317	98.8
Breast cancer subtypes, receptor status		
TNBC	79	24.6
HR positive/HER2 negative (Luminal)	155	48.3
HR negative/HER2 positive (HER2- enriched)	13	4.0
HR positive/HER2 positive (Luminal B HER2)	40	12.5
HR positive/HER2 not available	22	6.8
Unknown	12	3.7
Histology		
Ductal	265	82.6
Lobular	27	8.4
Ductal and lobular	2	0.6
Other	20	6.2
Unknown	7	2.2
Bilateral disease		
Yes	29	9.0
No	292	91.0
Patient history of second breast cancer		
Yes	38	11.2
No	283	88.2
Patient history of prior cancer (excluded breast cancer)		
Yes	39	12.1
No	282	87.9
Ovarian cancer	9	2.8
First-/second-/third- degree relative with breast or ovarian cancer		
Yes	209	65.1
No	106	33.0
Unknown	6	1.9
First-/second-/third- degree relative with cancer (excluded breast and ovarian)		
Yes	244	76.0
No	71	22.1
Unknown	6	1.9

SD, standard deviation; TNBC, triple-negative breast cancer; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

BRCA2 (4.3%); 4 in *TP53* (1.2%) and 4 in *PALB2* (1.2%), corresponding to 47% (39/83) of the positive results. Mutations in genes of moderate penetrance for BC (*ATM, CHEK2, RAD51C*) were found in 4.6% (15/321) (Figure 1; Table 2).

BRCA1/2 GPVs were found in 31 patients, corresponding to 9.6% (31/321) and to 37.3% (31/83) of the GPVs. In other 50 patients, corresponding to 15.6% (50/321), GPVs were detected in additional high, moderate and lower-risk BC genes: 23 GPVs (23/321 – 7.1%) were detected in high- and moderate-risk BC genes: 8 in *ATM* (9.6%), 6 with *CHEK2* (7.2%), 4 with *TP53* (4.8%), 4 with *PALB2* (4.8%), 1 with *RAD51C* (1.2%) and 29 GPVs (29/321 – 9%) in others of unknown clinical relevance for BC. (Figure 1; Table 2). No GPVs were identified in the other 70 genes. All GPVs detected are described in Table 3.

10.3389/fonc.2022.976959

TABLE 2 Frequency of germline pathogenic variants found in the study population.

Gene	No. of Patients	%	95% CI
Negative for GPV	240	74.8	
Positive for GPV	81	25.2	
Total of GPV	83		
High risk breast cancer genes	39	47.0	
BRCA1	17	20.5	12.4 – 30.8
BRCA2	14	16.9	9.5 – 26.7
TP53	4	4.8	1.3 – 11.9
PALB2	4	4.8	1.3 – 11.9
Moderate risk breast cancer genes	15	18.0	
ATM	8	9.6	4.3 – 18.1
CHEK2	6	7.2	2.7 – 15.1
RAD51C	1	1.2	0.0 – 6.5
Low risk breast cancer genes	29	35.0	
MUTYH (monoallelic)	7	8.4	3.5 – 16.6
SBDS	3	3.6	0.8 – 10.2
FANCI	2	2.4	0.3 – 8.4
HNF1A	2	2.4	0.3 – 8.4
PFR1	2	2.4	0.3 – 8.4
RECQL4	2	2.4	0.3 – 8.4
BLM	1	1.2	0.0 – 6.5
BRIP1	1	1.2	0.0 – 6.5
FANCA	1	1.2	0.0 – 6.5
FANCD2	1	1.2	0.0 – 6.5
FANCE	1	1.2	0.0 – 6.5
FANCL	1	1.2	0.0 – 6.5
FANCM	1	1.2	0.0 – 6.5
FH	1	1.2	0.0 – 6.5
PHOX2B	1	1.2	0.0 – 6.5
PMS2	1	1.2	0.0 – 6.5
SLX4	1	1.2	0.0 – 6.5

GPV, germline pathogenic variants.
Genes included in the multigene panel: AIP, ALK, APC, ATM, BAP1, BLM, BMPR1A, BRCA1, BRCA2, BRIP1, BUB1B, CDC73, CDH1, CDK4, CDKN1C, CDKN2A, CEBPA, CEP57, CHEK2, CYLD, DDB2, DICER1, DIS3L2, EGFR, EPCAM, ERCC2, ERCC3, ERCC4, ERCC5, EXT1, EXT2, EZH2, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FH, FLCN, GATA2, GPC3, HNF1A, HRAS, KIT, MAX, MEN1, MET, MLH1,, MSH2, MSH6, MUTYH, NBN, NF1, NF2, NSD1, PALB2, PHOX2B, PMS1, PMS2, PRF1, PRKAR1A, PTCH1, PTEN, RAD51C, RAD51D, RB1, RECQL4, RET, RHBDF2, RUNX1, SBDS, SDHAF2, SDHB, SDHC, SDHD, SLX4, SMAD4, SMARCB1, STK11, SUFU, TMEM127, TP53, TSC1, TSC2, VHL, WRN, WT1, XPA, XPC.

The presence of GPV was analyzed according to BC molecular subtypes (Figure 2). Among 79 TNBC patients, 31.6% (26/79) had a GPV and the gene with the highest number of GPVs was *BRCA1* (12.6%) followed by *BRCA2*, *MUTYH* and *PALB2* (2.5% each), and *TP53* and *RAD51C* had 1.3% each. Luminal BC was diagnosed in 155 patients, and of these, 18.7% (29/155) had a P/LP variant, with *BRCA2* and *BRCA1* showing 5.8% of mutations (3.2% and 2.6%, respectively). The subtype Luminal B HER2 was found in 40 patients, and of these, 40% (16/40) had a P/LP variant, with *BRCA2* mutation being found in 10% (4/40) and *ATM* in 12.5% (5/40), the most frequently mutated gene in this subgroup. Finally, for the 13 patients diagnosed with HER2-enriched BC, 30.8% (4/13) had a P/LP variant in *BRCA2, TP53, FH* and *RECQL4* genes (7.7% each), these latter two genes with an still to be determined relevance in BC (Figure 2).

At least one VUS was identified in 249 patients (249/321 - 77.5%). A total of 470 variants were found in 81 of the 94 genes of the panel. *FANCM* gene harbored the larger number of VUS, found in 22 patients (22/249 - 8.8%), followed by *ATM* and *RECQL4*, with VUS identified in 20 patients each (20/249 - 8%), *MSH6* with VUS identified in 18 patients (18/249 - 7.2%), *SLX4* in 17 patients (17/249 - 6.8%) and *BRCA2* in 16 (16/249 – 6.4%). Some patients had more than one VUS in the same gene (Figure 3). Most VUS were identified in genes without strong evidence of association with breast cancer.

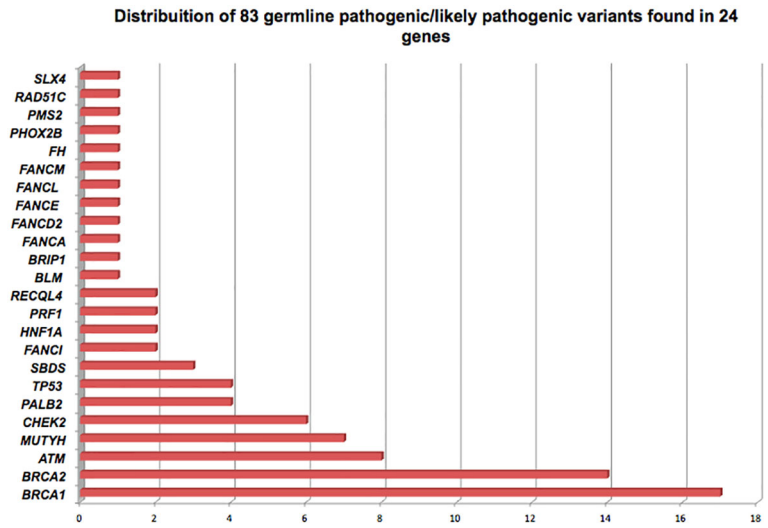


FIGURE 1
Distribution of 83 germline pathogenic/likely pathogenic variants of breast cancer-related genes detected in 81 Brazilian patients, found in 24 cancer susceptibility genes.

Two patients were diagnosed with MINAS, the first (P013) was diagnosed with a ductal carcinoma at the age of 49 years, molecular subtype luminal B HER2 with family history of breast, prostate, colorectal and gastric cancer. For this subject we detected the GPV c.5617_5621delGTAAT; p.(Val1873*) in *BRCA2* gene and the variant c.470T>C; p.(Ile157Thr) in *CHEK2* gene. The variant c.470T>C is a founder variant with low penetrance in Finnish and Polish individuals (31, 32). Mutations in these two genes increase the risk for BC and *BRCA2* is also related to increase risk for prostate cancer even as *CHEK2* increases risk for colorectal cancer, probably explaining the family history of cancer at multiple distinct sites.

The second patient (P211) was diagnosed with a ductal carcinoma at 31 years old, molecular subtype luminal B HER2 with a family history of BC in a second-degree relative. The GPV c.156_157insAlu was detected in *BRCA2* gene and the variant c.6529C>T; p.(Gln2177Ter) in *ATM* gene, both genes related to BC.

We evaluated potential associations among clinical variants and the BC carriers with GPVs in both *BRCA1* and *BRCA2* genes, and 22 additional genes (*ATM, BLM, BRIP1, CHEK2, FANCA, FANCD2, FANCE, FANCI, FANCL, FANCM, FH, HNF1A, MUTYH, PALB2, PHOX2B, PMS2, PRF1, RAD51C, RECQL4, SBDS, SLX4* and *TP53*). Significant associations were found only with *BRCA1/2* GPVs and self-reported Ashkenazi Jewish ethnicity. No other significant associations were found with the other clinical variables evaluated (age of onset, BC histology, hormone receptor status, bilateral BC, personal history of other malignant neoplasms and familial history of cancer) (Supplementary Table 1) neither in the additional 22 genes with GPVs (Supplementary Table 2).

Among patients without a GPV, the mean age at diagnosis of BC was 45.5 years. For patients with GPVs, the mean age at diagnosis was 44.2 years for *BRCA1*, 42.8 for *BRCA2*, 46.7 for *TP53*, 38.1 for *ATM* and 40.2 for *PALB2* carriers. Mean age in patients with GPVs located in low risk genes for BC was 46.3 years. Among patients diagnosed with BC before 45 years, 25.6% (47/183) had a GPV; between ages 46 and 60 years, 29.2% (31/106) and after 60 years of age, only 9.3% (3/32).

Regarding the diagnosis of other malignancies, excluding BC, 12.1% (39/321) of the women included had a diagnosis of another primary cancer and the most frequent was ovarian cancer (9/39 - 23%), followed by thyroid (7/39 - 18%) and colorectal cancer (7/39 - 18%).

Ancestry analysis

Genetic variants in target and off-target regions captured by the multigene panel were used to access the genetic ancestry of our cohort of 321 non-related women with BC (Supplementary Figure 1). According to the proportion of first and second most common ancestries, patients were divided into eleven categories. Most patients were classified as having predominant European ancestry (183/321 – 57.0%), or European admixed with a second ancestry (51/321 – 15.9%). African ancestry was observed as the major ancestry in only 12 patients (3.7%) and as the second ancestry in 36 patients (11.0%); and finally Asian ancestry was observed as major ancestry in 16 patients (5.0%) and Asian/Native American as second ancestries in 0.9% (3/321) (Supplementary Table 3).

TABLE 3 Description of the identified 83 germline pathogenic variants.

ID	Gene	Chr:Pos	Type	HGVS Nomenclature	dbSNP	MAF (gnomAD)	Clinical Significance (Clinvar)	ACMGClassification
P017	*ATM*	11:108196153	Frameshift	c.6691_6692insCTTTT, (p.Leu2231SerfsTer6)	ND	ND	ND	Likely pathogenic
P047	*ATM*	11: 108196143	Missense	c.6679C>T, (p.Arg2227Cys)	rs564652222	ND	Pathogenic	
P062	*ATM*	11: 108138057	Nonsense	c.2626C>T, (p.Gln876Ter)	ND	ND	ND	Likely pathogenic
P087	*ATM*	11: 108155009	Frameshift	c.3802delG,(p.Val1268Terfs)	rs587779834	ND	Pathogenic	
P134	*ATM*	11: 108203613	Nonsense	c.7913G>A, (p.Trp2638Ter)	rs377349459	0.000017	Pathogenic	
P233	*ATM*	11: 108203613	Nonsense	c.7913G>A, (p.Trp2638Ter)	rs377349459	0.000017	Pathogenic	
P185	*ATM*	11: 108175549	Nonsense	c.5644C>T, (p.Arg1882Ter)	rs786204433	ND	Pathogenic	
P211	*ATM*	11:108192104	Nonsense	c.6529C>T, (p.Gln2177Ter)	rs766706861	0.0000039	ND	Likely pathogenic
P044	*BLM*	15: 91310153	Frameshift	c.2207_2212delinsTAGATTC, (p.Tyr736fs)	rs113993962	0.00017	Pathogenic	
P009	*BRCA1*	17: 41251790	Splice Donor	c.547+2T>A	rs80358047	ND	Pathogenic	
P012	*BRCA1*	17: 41244068 - 41244071	Frameshift	c.3477_3480delAAAG, (p.Ile1159Metfs)	rs80357781	ND	Pathogenic	
P018	*BRCA1*	17: 41256984	Intrônica	c.213-11T>G, IVS5-11T>G	rs80358061	0.000011	Pathogenic	
P065	*BRCA1*	17: 41219623	Splice Donor	c.5074+2T>C, IVS17+2T>C	rs80358089	ND	Pathogenic	
P082	*BRCA1*	17: 41246532	Frameshift	c.1016dupA (p.Val340GlyfsTer6)	rs80357569	ND	Pathogenic	
P123	*BRCA1*	17: 41199683	Nonsense	c.5444G>A, p.(Trp1815Ter)	rs80356962	ND	Pathogenic	
P131	*BRCA1*	17: 41219623	Splice Donor	c.5074+2T>C, IVS17+2T>C	rs80358089	ND	Pathogenic	
P132	*BRCA1*	17: 41209082	Frameshift	c.5266dupC, (p.Gln1756ProfsTer74)	rs80357906	ND	Pathogenic	
P145	*BRCA1*	17: 41215889	Splice Donor	c.5152+2T>C	rs886040914	ND	Pathogenic	
P154	*BRCA1*	17: 41243513	Frameshift	c.4035delA, (p.Glu1346LysfsTer20)	rs80357711	ND	Pathogenic	
P157	*BRCA1*	17: 41209082	Frameshift	c.5266dupC, (p.Gln1756ProfsTer74)	rs80357906	ND	Pathogenic	
P200	*BRCA1*	17: 41209082	Frameshift	c.5266dupC, (p.Gln1756ProfsTer74)	rs80357906	ND	Pathogenic	
P248	*BRCA1*	17: 41209082	Frameshift	c.5266dupC, (p.Gln1756ProfsTer74)	rs80357906	ND	Pathogenic	
P228	*BRCA1*	17: 41203112	Frameshift	c.5300delG, p.(Cys1767PhefsTer26)	ND	ND	ND	Likely pathogenic
P229	*BRCA1*	17: 41203112	Frameshift	c.5300delG, p.(Cys1767PhefsTer26)	ND	ND	ND	Likely pathogenic
P289	*BRCA1*	17: 41245861	Nonsense	c.1687C>T (p.Gln563Ter)	rs80356898	0.00004	Pathogenic	
P314	*BRCA1*	17: 41246251	Frameshift	c.1297delG, (p.Ala433Profs)	rs80357794	ND	Pathogenic	
P002	*BRCA2*	13: 32912236 - 32912239	Frameshift	c.3744_3747delTGAG, (p.Ser1248ArgfsTer10)	rs80359403	ND	Pathogenic	
P013	*BRCA2*	13:32914109 - 32914113	Frameshift	c.5617_5621delGTAAT, (p.Val1873Ter)	ND	ND	ND	Likely pathogenic
P052	*BRCA2*	13: 32914942	Frameshift	c.6450dupA, (p.Val2151SerfsTer25)	rs80359595	ND	Pathogenic	
P090	*BRCA2*	13: 32911300 - 32911303	Frameshift	c.2808_2811del, (p.Ala938ProfsTer21)	rs80359351	0.00002	Pathogenic	
P109	*BRCA2*	13:32953937	Missense	c.9004G>A, (p.Glu3002Lys)	rs80359152	ND	Pathogenic	
P130	*BRCA2*		Deletion	deletion éxons 1 e 2	ND	ND	ND	Likely pathogenic
P136	*BRCA2*	13: 32890599	Missense	c.2T>G, (p.Met1Arg)	rs80358547	0.00001	Pathogenic	
P211	*BRCA2*	13: 32893302 - 32893303	RNA splicing	c.156_157insAlu	ND	ND	Pathogenic	
P238	*BRCA2*	13: 32911300 - 32911303	Frameshift	c.2808_2811del, (p.Ala938Profs)	rs80359351	0.00002	Pathogenic	
P245	*BRCA2*	13: 32900635	Splice Acceptor	c.517-1G>A, IVS6-1G>A	rs81002849	ND	Pathogenic	
P250	*BRCA2*	13: 32929114	Nonsense	c.7124T>G, (p.Leu2375Ter)	rs886040687	ND	Pathogenic	

(Continued)

TABLE 3 Continued

ID	Gene	Chr:Pos	Type	HGVS Nomenclature	dbSNP	MAF (gnomAD)	Clinical Significance (Clinvar)	ACMGClassification
P275	*BRCA2*	13: 32915148	Nonsense	c.6656C>G, (p.Ser2219Ter)	rs80358893	ND	Pathogenic	
P309	*BRCA2*	13: 32972521	Frameshift	c.9871delT, (p.Ser3291Leufs)	rs886040854	ND	Pathogenic	
P319	*BRCA2*	13: 32914174	Nonsense	c.5682C>G, (p.Tyr1894Ter)	rs41293497	ND	Pathogenic	
P107	*BRIP1*	17: 59793412	Nonsense	c.2392C>T, (p.Arg798Ter)	rs137852986	0.00015	Conflicting	Pathogenic
P003	*CHEK2*	22: 29091857	Frameshift	c.1100delC, (p.Thr367Metfs)	rs555607708	0.00182	Conflicting	Pathogenic
P004	*CHEK2*	22: 29121326	Missense	c.349A>G, (p.Arg117Gly)	rs28909982	0.00011	Pathogenic	
P033	*CHEK2*	22: 29121326	Missense	c.349A>G, (p.Arg117Gly)	rs28909982	0.00011	Pathogenic	
P048	*CHEK2*	22: 29121058	Missense	c.499G>A, (p.Gly167Arg)	rs72552322	0.000023	Conflicting	Likely pathogenic
P105	*CHEK2*	22: 29091857	Frameshift	c.1100delC, (p.Thr367Metfs)	rs555607708	0.00182	Conflicting	Pathogenic
P013	*CHEK2*	22:29121087	Missense	c.470T>C, (p.Ile157Thr)	rs17879961	0.0049	Conflicting	Likely pathogenic
P261	*FANCA*	16:89862330	Frameshift	c.983_986TCAC, (p.His330AlafsTer4)	rs772359099	0.000042	Pathogenic	
P076	*FANCD2*	3:10133904	Nonsense	c.3817C>T, (p.Arg1273Ter)	rs745930696	0.000015	ND	Likely pathogenic
P208	*FANCE*	6:35423630	Nonsense	c.355C>T, (p.Gln119Ter)	rs121434505	0.000011	Pathogenic	
P096	*FANCI*	15: 89849381	Frameshift	c.3493delG, (p.Asp1165Thrfs)	rs1060501884	ND	Pathogenic	
P264	*FANCI*	15:89828432	Nonsense	c.1804C>T, (p.Arg602Ter)	rs1432325198	0.000010	ND	Likely pathogenic
P030	*FANCL*	2:58388668	Inframe	c.1007_1009delTAT, (p.Ile336_Cys337delinsSer)	rs747253294	ND	Conflicting	Likely pathogenic
P043	*FANCM*	14:45618145 - 45618161	Frameshift	c.865_881delCTTATTGTTCCGCTTGG, (p.Leu289Ter)	ND	ND	ND	Likely pathogenic
P042	*FH*	1:241661228	Inframe	c.1431_1433dupAAA, (p.Lys477dup)	rs367543046	0.0010	Conflicting	Likely pathogenic
P080	*HNF1A*	12:121432118	Frameshift	c.872dupC, (p.Gly292ArgfsTer25)	rs587776825	ND	Pathogenic	
P296	*HNF1A*	12:121432118	Frameshift	c.872dupC, (p.Gly292ArgfsTer25)	rs587776825	ND	Pathogenic	
P071	*MUTYH*	1: 45798475	Missense	c.536A>G, (p.Tyr179Cys)	rs34612342	0.0015	Pathogenic	
P192	*MUTYH*	1: 45798475	Missense	c.536A>G, (p.Tyr179Cys)	rs34612342	0.0015	Pathogenic	
P155	*MUTYH*	1: 45797228	Missense	c.1187G>A, (p.Gly396Asp)	rs36053993	0.003	Pathogenic	
P203	*MUTYH*	1: 45797228	Missense	c.1187G>A, (p.Gly396Asp)	rs36053993	0.003	Pathogenic	
P285	*MUTYH*	1: 45797228	Missense	c.1187G>A, (p.Gly396Asp)	rs36053993	0.003	Pathogenic	
P058	*MUTYH*	1:45797760	Splice Acceptor	c.934-2A>G	rs77542170	0.0011	Conflicting	Likely pathogenic
P147	*MUTYH*	1:45797760	Splice Acceptor	c.934-2A>G	rs77542170	0.0011	Conflicting	Likely pathogenic
P023	*PALB2*	16: 23641218	Nonsense	c.2257C>T, (p.Arg753Ter)	rs180177110	0.000023	Pathogenic	
P254	*PALB2*	16: 23649427	Frameshift	c.72delG, (p.Arg26Glyfs)	rs180177142	ND	Pathogenic	
P266	*PALB2*	16: 23637594	Nonsense	c.2711G>A, (p.Trp904Ter)	rs1060502726	ND	Pathogenic	
P304	*PALB2*	16: 23646627	Nonsense	c.1240C>T, (p.Arg414Ter)	rs180177100	0.0000079	Pathogenic	
P036	*PHOX2B*	4:41748030	Frameshift	c.739delG, (p.Ala247ProfsTer62)	ND	ND	ND	Likely pathogenic
P284	*PMS2*	7:6045549	Missense	c.137G>T, (p.Ser46Ile)	rs121434629	0.00017	Likely pathogenic	
P001	*PRF1*	10:72358804	Missense	c.673C>T; (p.Arg225Trp)	rs28933973	0.000012	Pathogenic	
P166	*PRF1*	10:72358189	Frameshift	c.1288dupG, (p.Asp430GlyfsTer28)	rs1226526104	ND	ND	Likely pathogenic
P081	*RAD51C*	17:56798156	Frameshift	c.890_899del, (p.Leu297HisfsTer2)	rs1555602141	ND	Pathogenic	
P056	*RECQL4*	8:145741776	Nonsense	c.727C>T, (p.Gln243Ter)	rs1345625725	0.000031	ND	Likely pathogenic
P313	*RECQL4*	8:145738437	Frameshift	c.2547_2548delGT, (p.Phe850ProfsTer33)	rs778141083	0.000010	Pathogenic	
P069	*SBDS*	7:66459197	Splice Donor	c.258+2T>C	rs113993993	0.0038	Pathogenic	
P112	*SBDS*	7:66459197	Splice Donor	c.258+2T>C	rs113993993	0.0038	Pathogenic	

(Continued)

TABLE 3 Continued

ID	Gene	Chr:Pos	Type	HGVS Nomenclature	dbSNP	MAF (gnomAD)	Clinical Significance (Clinvar)	ACMGClassification
P283	*SBDS*	7:66459197	Splice Donor	c.258+2T>C	rs113993993	0.0038	Pathogenic	
P230	*SLX4*	16:3633330	Frameshift	c.4921dupG, (p.Val1641GlyfsTer15)	rs770425994	0.000027	ND	Likely pathogenic
P097	*TP53*	17: 7574017	Missense	c.1010G>A, (p.Arg337His)	rs121912664	0.00001	Pathogenic	
P160	*TP53*	17: 7574017	Missense	c.1010G>A, (p.Arg337His)	rs121912664	0.00001	Pathogenic	
P277	*TP53*	17: 7574017	Missense	c.1010G>A, (p.Arg337His)	rs121912664	0.00001	Pathogenic	
P256	*TP53*	17:7577121	Missense	c.817C>T, (p.Arg273Cys)	rs121913343	0.00001	Pathogenic	

Chr:Pos, chromosome position; HGVS, Human Genome Variation Society; dbSNP, Single Nucleotide Polymorphism database; MAF, Minor allele frequency; gnomAD, genome aggregation database; ND, not described; ACMG, American College of Medical Genetics and Genomics.

As molecular and cancer genomics studies are sparse for populations, such as those with African, as well as Asian/Native American ancestries we hypothesized that more VUS would be found in among non-Europeans. Also, if these indeed represent real cancer risk, some variations could have stronger associations with familial cancer history or early age at diagnosis. We therefore investigated whether family history of cancer, age at diagnosis, ancestry and ethnicity may be associated with number of VUS per patient. Using multivariate Poisson regression with feature selection process (see methods) we found evidence for Asian ancestry and family history (1st, 2nd and 3rd degree relatives) with any cancer other than breast and ovarian cancer as positive predictors of the number of VUS. For an individual with family history of cancer other than breast and ovarian cancer, their incidence rate ratios of having VUS is 1.35 [95% CI 1.06 – 1.75, p = 0.019] when all other variables are held constant. With a percentage increase in Asian and Native American ancestry we would expect number of VUS to

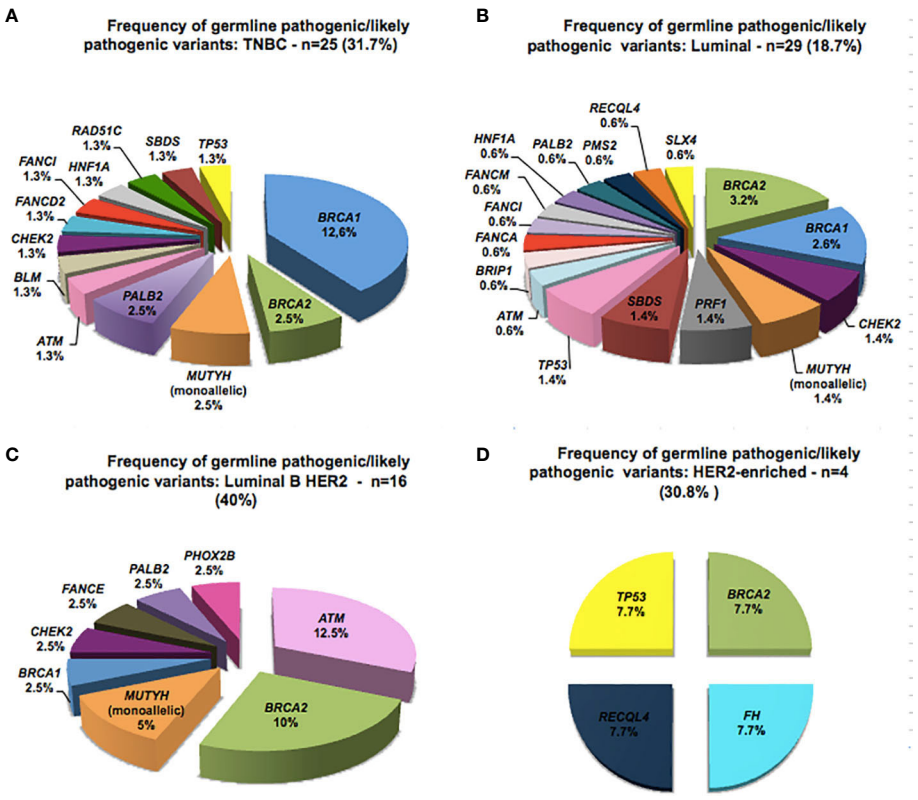


FIGURE 2
Spectrum of germline pathogenic variants detected according to the molecular subtype of breast cancer.

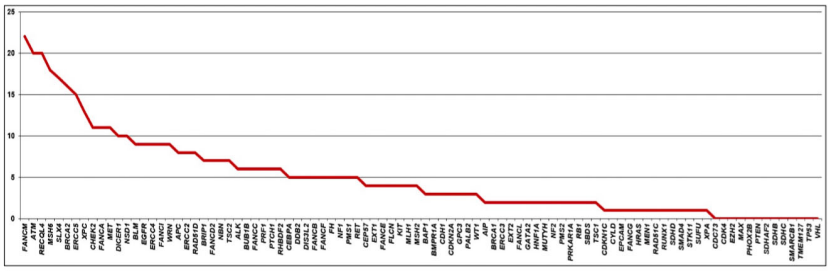


FIGURE 3
Distribution of variants of uncertain significance, according to the identified gene.

increase by a factor of 1.61 [95% CI 1.09 – 2.32, p = 0.013] and 4.58 [95% CI 1.30 – 13.76, p = 0.011] respectively per increase in percent (Supplementary Table 4). However, incidence ratios were higher when narrowed down to patients, who had VUS in any of the eight frequently mutated genes (top 10% VUS) in the series thus *FANCM, ATM, MSH6, RECQL4, BRCA2, ERCC5, SLX4* and *XPC*.

Regarding GPVs, using a multivariate binomial logistic regression model of all predictors followed by bootstrap backward stepwise feature selection process, we found that only the Ashkenazi Jewish ethnicity had positive association with presence or absence of GPVs. The odds ratio of having GPVs was 9.19 [95% 1.16 – 187.30] but reached no statistical significance (p = 0.056) (Supplementary Table 5). There was no difference between the ancestries groups and the frequency of GPV in the most frequent mutated genes (*BRCA1, BRCA2, TP53, PALB2, ATM, CHEK2* and *MUTYH*). Breast cancer molecular subtypes (TBNC and non-TNBC) and age of tumor onset were also similar among different ancestry categories (data not shown).

Discussion

The use of multigene panels for genetic counselling in hereditary BC is growing more and more. However, although the use of these panels can help in the diagnosis cancer predisposition syndromes for some families, the challenges of interpreting the results for meaningful genetic counseling still lingers, as cancer risk estimates and management strategies still have to be established for many genes.

In the present study, among 321 unrelated women with BC, the frequency of GPVs in *BRCA1/2* genes was 9.6% and in non-*BRCA1/2* cancer predisposition genes was 15.6%. Overall the analysis of the 94-genes panel contributed to identify GPVs in non-*BRCA1/2* in 50 patients, increasing the frequency of variants identification by almost 16%, similar to some previous studies in BC (33–36). Considering the 9 BC-genes, which were recently described as the most relevant BC predisposing genes (*BRCA1, BRCA2, ATM, BARD1, CHEK2, PALB2, RAD51C, RAD51D* and *TP53*) (9), 16.8% patients were detected as carriers GPV in at least one of these genes. The seven non-*BRCA1/2* genes contributed with 7.2% in the ability to detect a GPV as a genetic determinant of BC in these women, showing a main gain in terms of clinical value in analyzing of these 9 BC-genes instead of only *BRCA1/2*.

Our findings demonstrated a higher prevalence of GPVs in high-risk BC genes, such as *BRCA1*, *BRCA2*, *PALB2* and *TP53* as expected. Pathogenic variants in *TP53* were identified in four patients and three of them harbored the same variant - c.1010G>A (p.Arg337His/R337H), that was introduced in Brazil possibly a founder effect, and is now found in relatively high frequencies in the southeast and southern regions of the country (37). A recent study identified a variant in the tumor suppressor *XAF1* (E134*) in a subset of R337H carriers and proposes that the co-segregation of *XAF1*-E134* and *TP53*-R337H mutations leads to a more aggressive cancer phenotype than R337H alone (38). The analysis of variant E134* was positive in 2 patients (P097 and P277) and negative in patient P160. No GPV were identified in the other high-risk BC genes such as *PTEN*, *CDH1* and *STK11* and they are very rare, as demonstrated in a recent study (9).

As for variants of moderate risk for BC, GPVs in *ATM*, *CHEK2* and *RAD51C* were found in 4.6% of our series and corresponding to 18% of GPVs found, a finding consistent with other recent studies (33–36, 39). According to Tung et al. (2016), germline mutations in moderate risk BC susceptibility genes are identified in approximately 2% to 5% of individuals performing multigene panel (36). It is important to note that, after *BRCA1/2*, *ATM* was the most prevalent mutated gene among patients in our study. De Souza Timoteo et al. (2018) reported that germline mutations in moderate- and low-risk BC genes were detected in 3.8% of individuals, including *ATM, ATR, CDH1, MLH1* and *MSH6* (40).

Our results are similar to other Brazilian studies. De Souza Timoteo et al. (2018) evaluated 157 individuals (132 with breast and

25 cancer-unaffected) using three different types of multigene panel (40). Germline pathogenic variants were identified in twenty-seven individuals (17.2%), 24 with BC and three asymptomatic, and most of them in *BRCA1/2* genes (75%) (38). A recent study evaluated germline molecular data in hereditary BC in 224 patients and GPVs were detected in 20.5% (41). The frequency of GPV in a high-penetrance BC gene was 61% and frequency of moderate penetrance genes represented 15.2% of the positive results (41).

According to the guidelines of the NCCN, screening recommended for patients with moderate-risk BC GPVs such as *ATM* and *CHEK2* is annual mammogram and consider breast MRI with contrast due to increased risk of BC (20). There is insufficient evidence for risk-reducing mastectomy and should be based on family history. In our series, just one of the patients with a GPV in *CHEK2* had a contralateral BC and none of the patients with a GPV in the moderate-risk BC genes had other primary cancers (excluding BC) (20). Bilateral BC was not significantly associated with GPVs in our cohort.

Regarding the molecular subtype of BC, we observed the predominance of GPVs in *BRCA1* genes in TNBC tumors, as reported by others, including our own previous study with 131 Brazilian women with TNBC (42–45). For the luminal subtype, GPVs in *BRCA1* and *BRCA2* genes were found in only 5.8%, while GPVs in other genes correspond to 12.9%, highlighting the contribution of the multigene panel in luminal tumors. Moreover, we found that the proportion of *ATM* GPVs is significantly higher in Luminal B HER2 tumors, as previously reported in the literature (46).

Seven carriers with GPV in low-risk gene *MUTYH* (monoallelic), were found here, whose association with BC risk is still controversial. Some studies reported an increased risk of BC for monoallelic *MUTYH* mutation carriers (47, 48). However, other studies did not find statistical evidence for an increased risk of BC (49, 50).

Recently, it has been described the MINAS condition, which is characterized by the presence of two or more GPVs in genes related to cancer predisposition in the same individual (51). We found a frequency of 2.4% among patients who are carriers of two GPVs, with the following combinations *BRCA2*/*ATM* and *BRCA2*/*CHEK2*. An overlap of phenotypes associated with both genes was observed in theses cases.

According to previous studies, about a third of multigene panels identify at least one VUS in one or more genes (18, 19). A study of 2,158 women with BC referred to genetic testing using a 25-multigene panel, showed that VUS were found in 40% of individuals (19). Another similar study with 198 women who underwent 42 multigene panel showed that VUS was identified in 88% of them (18). In our series using a 94-genes panel, we found 77.5% of patients with VUS and showed that patients with Asian and Native American ancestry were associated to a higher number of VUS. It is expected that the panels containing a larger number of genes will result in a higher rate of patients presenting VUS. Also, as most VUS represent rare missense variants with low minor allele frequencies or not described in populational databases, it is anticipated that genetically less characterized populations, such as the Brazilian, will have more VUS. In this sense, a recent study evaluating more than a 100,000 multigene hereditary cancer genetic tests revealed that, compared to Europeans, Asian and Middle Eastern individuals were most likely to be identified with VUS (52).

It should be noted that there were limitations associated with our study. Copy number variation analysis with MLPA was only performed for *BRCA1/2*. The panel used is not validated for large deletion/duplication analysis. Patients selected to perform the multigene panel had criteria for HBOC, with personal or family history that suggested higher inherited cancer risk. At the same time, it is possible that we did not include BC patients who did not meet criteria for HBOC but could have GPVs in the other genes included in the panel. Another limitation of the present study was that we could not establish a genotype-phenotype correlation for moderate and low-risk BC genes, due to the small number of patients with GPVs in these genes.

In conclusion, our results indicate that although most of the GPV found in this study were in the *BRCA1/2* genes (9.6%), women who fulfill the clinical criteria for HBOC may benefit from multigene panel testing, because the panel allows to identify GPV in relevant BC predisposing genes (7.2%), including those who change the clinical management. This is the first study that analyzed multigene panel and its relation with molecular subtypes in Brazilian BC patients. Further studies are still needed to better comprehend the heritability of distinct subtypes of BC in Brazilian women, including those who do not fulfill clinical criteria for HBOC to correlate genotype-phenotype of moderate and low-risk BC genes.

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 Ethics Committee of the A. C. Camargo Cancer Center (protocol number 2483/18). The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization and design of the study: DP, MF, GT and DC; Data analysis: DP, GT, KS and DC; Carrying out of the

Ancestry algorithm: EN and IS; Ancestry analysis: SA and PP; Writing and editing the original draft preparation: DP, GT, MF, PP, WF and DC; Funding acquisition, DC. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo, grant number 2014/50943-1, Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number 465682/2014-6 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - 88887.136405/2017-00.

Acknowledgments

We acknowledge the patients who participated in the study, and the A.C. Camargo Laboratory of Genomic Diagnostic.

Conflict of interest

Author PP was employed by C2i Genomics.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.976959/full#supplementary-material

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* (2015) 1365:E359–86. doi: 10.1002/ijc.29210

2. Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. *Cancer* (1996) 7711:2318–24. doi: 10.1002/SICI1097-01421996060177:11<2318::AID-CNCR21>3.0.CO;2-Z

3. Kean S. The other breast cancer genes. *Science* (2014) 3436178:1457–9. doi: 10.1126/science.343.6178.1457

4. Anglian Breast Cancer Study Group. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. *Br J Cancer* (2000) 8310:1301–8. doi: 10.1054/bjoc.2000.1407

5. Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, et al. EMBRACE. cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J Natl Cancer Inst* (2013) 10511:812–22. doi: 10.1093/jnci/djt095

6. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: A combined analysis of 22 studies. *Am J Hum Genet* (2003) 725:1117–30. doi: 10.1086/375033

7. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol* (2007) 2511:1329–33. doi: 10.1200/JCO.2006.09.1066

8. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A population-based study of genes previously implicated in breast cancer. *N Engl J Med* (2021) 3845:440–51. doi: 10.1056/NEJMoa2005936

9. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, et al. Breast cancer risk genes - association analysis in more than 113,000 women. *N Engl J Med* (2021) 3845:428–39. doi: 10.1056/NEJMoa1913948

10. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer risks associated with germline PALB2 pathogenic variants: An international study of 524 families. *J Clin Oncol* (2020) 387:674–85. doi: 10.1200/JCO.19.01907

11. Euhus DM, Robinson L. Genetic predisposition syndromes and their management. *Surg Clin North Am* (2013) 932:341–62. doi: 10.1016/j.suc.2013.01.005

12. Slavin TP, Niell-Swiller M, Solomon I, Nehoray B, Rybak C, Blazer KR, et al. Clinical application of multigene panels: Challenges of next-generation counseling and cancer risk management. *Front Oncol* (2015) 5(5):208. doi: 10.3389/fonc.2015.00208

13. LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, et al. Utilization of multigene panels in hereditary cancer predisposition testing: Analysis of more than 2,000 patients. *Genet Med* (2014) 1611:830–7. doi: 10.1038/gim.2014.40

14. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn* (2012) 144:357–66. doi: 10.1016/j.jmoldx.2012.03.002

15. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* (2010) 10728:12629–33. doi: 10.1073/pnas.1007983107

16. Hall MJ, Forman AD, Pilarski R, Wiesner G, Giri VN. Gene panel testing for inherited cancer risk. *J Natl Compr Canc Netw* (2014) 12(9):1339–46. doi: 10.6004/jnccn.2014.0128

17. Rainville IR, Rana HQ. Next-generation sequencing for inherited breast cancer risk: counseling through the complexity. *Curr Oncol Rep* (2014) 163:371. doi: 10.1007/s11912-013-0371-z

18. Kurian AW, Hare EE, Mills MA, Kingham KE, McPherson L, Whittemore AS, et al. Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *J Clin Oncol* (2014) 3219:2001–9. doi: 10.1200/JCO.2013.53.6607

19. Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next- generation sequencing with a 25-gene panel. *Cancer* (2015) 1211:25–33. doi: 10.1002/cncr.29010

20. National Comprehensive Cancer Network. *Genetic/familial high-risk assessment: breast and ovarian. Li-fraumeni syndrome management. NCCN clinical practice guidelines in oncology; version 3* (2021). Available at: https://www.nccn.org/professionals/physician_gls/default.aspx#genetics_screening.

21. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus

recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med* (2015) 175:405–24. doi: 10.1038/gim.2015.30

22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* (2007) 81 (3):559–75. doi: 10.1086/519795

23. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* (1983) 32:307–17. doi: https://doi.org/10.2307/2987937

24. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* (2015) 5267571:68–74. doi: 10.1038/nature15393

25. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science* (2002) 2985602:2381–5. doi: 10.1126/science.1078311

26. Rosenberg NA. Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* (2006) 70Pt 6:841–7. doi: 10.1111/j.1469-1809.2006.00285.x

27. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* (2009) 19(9) 1655–64. doi: 10.1101/gr.094052.109

28. Oak N, Cherniack AD, Mashl RJTCGA Analysis Network, , Hirsch FR, Ding L, et al. Ancestry-specific predisposing germline variants in cancer. *Genome Med* (2020) 121:51. doi: 10.1186/s13073-020-00744-3

29. Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell* (2020) 375:639–54.e6. doi: 10.1016/j.ccell.2020.04.012

30. Foulkes WD. The ten genes for breast and ovarian cancer. *Nat Rev Clin Oncol* (2021) 185:259–60. doi: 10.1038/s41571-021-00491-3

31. Kilpivaara O, Vahteristo P, Falck J, Syrjäkoski K, Eerola H, Easton D, et al. CHEK2 variant I157T may be associated with increased breast cancer risk. *Int J Cancer* (2004) 111(4):543–7. doi: 10.1002/ijc.20299

32. Staalesen V, Falck J, Geisler S, Bartkova J, Børresen-Dale AL, Lukas J, et al. Alternative splicing and mutation status of CHEK2 in stage III breast cancer. *Oncogene* (2004) 2352:8535–44. doi: 10.1038/sj.onc.1207928

33. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol* (2017) 3(9):1190–6. doi: 10.1001/jamaoncol.2017.0424

34. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* (2016) 188:823–32. doi: 10.1038/gim.2015.166

35. Thompson ER, Rowley SM, Li N, McInerny S, Devereux L, Wong-Brown MW, et al. Panel testing for familial breast cancer: Calibrating the tension between research and clinical care. *J Clin Oncol* (2016) 3413:1455–9. doi: 10.1200/JCO.2015.63.7454

36. Tung N U, Lin N, Kidd J, Allen BA, Singh N, Wenstrup RJ, et al. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J Clin Oncol* (2016) 3413:1460–8. doi: 10.1200/JCO.2015.65.0747

37. Garritano S, Gemignani F, Palmero EI, Olivier M, Martel-Planche G, Le Calvez-Kelm F, et al. Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of southern Brazil: evidence for a founder effect. *Hum Mutat* (2010) 312:143–50. doi: 10.1002/humu.21151

38. Pinto EM, Figueiredo BC, Chen W, Galvao HCR, Formiga MN, Fragoso MCBV, et al. XAF1 as a modifier of p53 function and cancer susceptibility. *Sci Adv* (2020) 626:eaba3231. doi: 10.1126/sciadv.aba3231

39. Rosenthal ET, Evans B, Kidd J, Brown K, Gorringe H, van Orman M, et al. Increased identification of candidates for high-risk breast cancer screening through expanded genetic testing. *J Am Coll Radiol* (2017) 144:561–8. doi: 10.1016/j.jacr.2016.10.003

40. de Souza Timoteo AR, Gonçalves AÉMM, Sales LAP, Albuquerque BM, de Souza JES, de Moura PCP, et al. A portrait of germline mutation in Brazilian at-risk for hereditary breast cancer. *Breast Cancer Res Treat* (2018) 1723:637–46. doi: 10.1007/s10549-018-4938-0

41. Sandoval RL, Leite ACR, Barbalho DM, Assad DX, Barroso R, Polidorio N, et al. Germline molecular data in hereditary breast cancer in Brazil: Lessons from a large single-center analysis. *PLoS One* (2021) 162:e0247363. doi: 10.1371/journal.pone.0247363

42. Young SR, Pilarski RT, Donenberg T, Shapiro C, Hammond LS, Miller J, et al. The prevalence of BRCA1 mutations among young women with triple-negative breast cancer. *BMC Cancer* (2009) 9:86. doi: 10.1186/1471-2407-9-86

43. Evans DG, Howell A, Ward D, Lalloo F, Jones JL, Eccles DM. Prevalence of BRCA1 and BRCA2 mutations in triple negative breast cancer. *J Med Genet* (2011) 488:520–2. doi: 10.1136/jmedgenet-2011-100006

44. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, et al. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol* (2015) 334:304–11. doi: 10.1200/JCO.2014.57.1414

45. Brianese RC, Nakamura KDM, Almeida FGDSR, Ramalho RF, Barros BDF, Ferreira ENE, et al. BRCA1 deficiency is a recurrent event in early-onset triple-negative breast cancer: A comprehensive analysis of germline mutations and somatic promoter methylation. *Breast Cancer Res Treat* (2018) 1673:803–14. doi: 10.1007/s10549-017-4552-6

46. Renault AL, Mebirouk N, Fuhrmann L, Bataillon G, Cavaciuti E, Le Gal D, et al. Morphology and genomic hallmarks of breast tumours developed by ATM deleterious variant carriers. *Breast Cancer Res* (2018) 201:28. doi: 10.1186/s13058-018-0951-9

47. Win AK, Reece JC, Dowty JG, Buchanan DD, Clendenning M, Rosty C, et al. Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. *Int J Cancer* (2016) 1397:1557–63. doi: 10.1002/ijc.30197

48. Rennert G, Lejbkowicz F, Cohen I, Pinchev M, Rennert HS, Barnett-Griness O. MUTYH mutation carriers have increased breast cancer risk. *Cancer* (2012) 1188:1989–93. doi: 10.1002/cncr.26506

49. Beiner M, Zhang W, Zhang S, Gallinger S, Sun P, Narod SA. Mutations of the MYH gene do not substantially contribute to the risk of breast cancer. *Breast Cancer Res Treat* (2009) 1143:575–8. doi: 10.1007/s10549-008-0042-1

50. Out AA, Wasielewski M, Huijts PE, van Minderhout IJ, Houwing-Duistermaat JJ, Tops CM, et al. MUTYH gene variants and breast cancer in a Dutch case-control study. *Breast Cancer Res Treat* (2012) 1341:219–27. doi: 10.1007/s10549-012-1965-0

51. Whitworth J, Skytte A, Sunde L, Lim DH, Arends MJ, Happerfield L, et al. Multilocus inherited neoplasia alleles syndrome: A case series and review. *JAMA Oncol* (2016) 23:373–9. doi: 10.1001/jamaoncol.2015.4771

52. Roberts ME, Susswein LR, Janice Cheng W, Carter NJ, Carter AC, Klein RT, et al. Ancestry-specific hereditary cancer panel yields: Moving toward more personalized risk assessment. *J Genet Couns* (2020) 294:598–606. doi: 10.1002/jgc4.1257

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 02 September 2022
DOI 10.3389/fonc.2022.949098

OPEN ACCESS

EDITED BY
Parvin Mehdipour,
Tehran University of Medical Sciences, Iran

REVIEWED BY
Gianluca Aimaretti,
Università del Piemonte Orientale, Italy
C. Christofer Juhlin,
Karolinska Institutet (KI), Sweden
Flavia Prodam,
University of Eastern Piedmont, Italy

*CORRESPONDENCE
Valentina Cirello
valentina.cirello@unimi.it

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

RECEIVED 20 May 2022
ACCEPTED 08 August 2022
PUBLISHED 02 September 2022

CITATION
Colombo C, Pogliaghi G, Tosi D, Muzza M, Bulfamante G, Persani L, Fugazzola L and Cirello V (2022) Thyroid cancer harboring *PTEN* and *TP53* mutations: A peculiar molecular and clinical case report. *Front. Oncol.* 12:949098. doi: 10.3389/fonc.2022.949098

Thyroid cancer harboring *PTEN* and *TP53* mutations: A peculiar molecular and clinical case report

Carla Colombo[1,2], Gabriele Pogliaghi[2,3], Delfina Tosi[4], Marina Muzza[3], Gaetano Bulfamante[4], Luca Persani[1,3,5], Laura Fugazzola[1,2] and Valentina Cirello[2,3]*

[1]Division of Endocrine and Metabolic Diseases, Istituto Auxologico Italiano Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy, [2]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy, [3]Laboratory of Endocrine and Metabolic Research, Istituto Auxologico Italiano IRCCS, Milan, Italy, [4]Unit of Human Pathology, Department of Health Sciences Santi Paolo e Carlo Medical School, University of Milan, Milan, Italy, [5]Department of Biotechnology and Translational Medicine, University of Milan, Milan, Italy

To date, the molecular mechanisms that underline aggressiveness and resistance to tyrosine kinase inhibitors in some thyroid carcinomas (TCs) are not known yet. We report the case of a young patient with a metastatic poorly differentiated (PDTC) and follicular thyroid carcinoma (FTC) refractory to conventional therapies and to Sorafenib. The patient, despite an initial partial response, died of progressive disease 21 months after diagnosis. The genetic analysis performed on the primary tumor and on lymph nodes and distant metastases allowed to identify a frameshift mutation (p.P248Tfs*5) in the *PTEN* gene, never described in TC. This mutation was present in the primary tumor and, with a lower allelic frequency, in metastases diagnosed after treatment with Sorafenib. Mutations in *TP53* (p.C135Y and c.920-2A>G previously detected in anaplastic carcinomas and p.M133R never found in TC) were also detected in the primary tissue together with a mono-allelic expression of the p.C135Y mutant at RNA level. At metastatic sites level, we found only the *TP53* splicing mutation c.920-2A>G. The presence of defects in mismatch repair (MMR) proteins and genomic instability was also evaluated. The primary tumor showed a partial expression of MMR proteins together with a strong genomic instability. In conclusion, we demonstrated that the rare combination of somatic *PTEN* and *TP53* mutations in a patient with a metastatic FTC, together with the presence of tumor heterogeneity and genomic instability, might be associated with a high tumor aggressiveness and resistance to treatments.

KEYWORDS

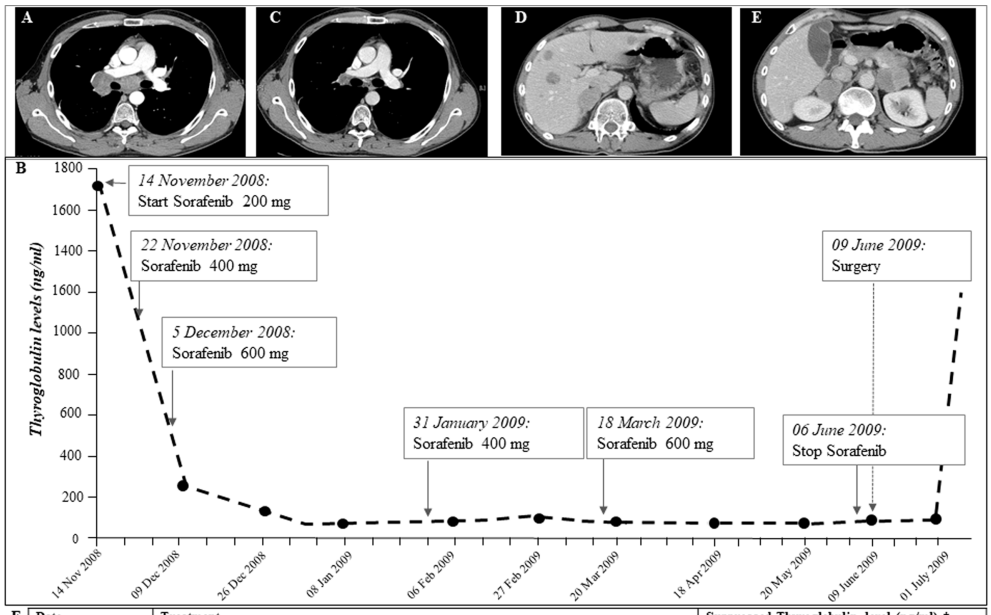
aggressive follicular thyroid cancer, *PTEN*, *TP53*, mismatch repair proteins, microsatellite instability, tyrosine kinase inhibitor, Sorafenib

Introduction

Well-differentiated thyroid carcinomas (WDTCs) are efficiently treated by surgery and radioiodine. On the contrary, poorly differentiated thyroid cancer (PDTC) and anaplastic thyroid cancer (ATC) are refractory to radioiodine therapy, and, in recent decades, tyrosine kinase inhibitors (TKIs) with angiogenetic and molecular targets were developed and used for these cases (1). The molecular mechanisms that generate thyroid cancer (TC) dedifferentiation are still unclear. Recently, next-generation sequencing (NGS) studies unraveled PDTCs and ATCs mutational landscapes, supporting the model of multistep tumorigenesis whereby PDTCs and ATCs arise from WDTCs through stepwise accumulation of additional genetic abnormalities, with prognostic and possible therapeutic relevance (2, 3). To date, one of the best characterized genetic alterations leading to the development of poorly and undifferentiated thyroid cancers is the loss of p53 tumor suppressor. The *TP53* gene codifies for a master regulatory protein, also known as "guardian of genome", involved in different cellular processes such as apoptosis, DNA repair, cell cycle arrest, and cellular senescence (4). The p53 protein has a key role in the maintenance of genetic stability and, thus, in preventing tumor development. *TP53* mutations, usually located in the region between exons 5 and 8, have been described in about 50% of human cancers. Whereas WDTC are rarely (<10%) *TP53* mutated, more than 70% of PDTC/ATCs are associated with *TP53* mutations (5). Almost all p53 mutations impaired p53 transcriptional activity and are not only important for tumor progression but also in the response to chemotherapy, to radioiodine therapy, and to TKIs treatment (5, 6). Interestingly, two mismatch repair genes, *MLH1* and *PMS2*, have been identified as targets for p53 in normal fibroblasts (7). The mismatch repair (MMR) system recognizes mismatched bases in double-stranded DNA and initiates the repair process. The identification of MLH1 and PMS2 as direct targets for p53 defines a signaling pathway that couples two important cellular guardian pathways, growth arrest, and apoptosis (7). Another tumor suppressor gene often mutated in human cancers, and also in aggressive TCs, is *PTEN* (phosphatase and tensin homolog). PTEN, through its lipid phosphatase activity inhibiting the PI3K/AKT pathway, regulates many cellular processes, including proliferation, survival, energy metabolism, cellular architecture, and motility (8). Mutations result in a non-functional or absent PTEN protein and are relatively common in ATCs, followed by PDTCs, and uncommon in follicular thyroid cancer (FTC) (9). Although TKIs are now available for aggressive TCs treatment (1, 10), additional strategies are currently being investigated, using the ability to modulate epigenetic changes in cancer DNA, restore the transcriptional activity of mutant p53, and block signal transduction downstream of different p53 family members (11). Immune checkpoint inhibitors, already used in colorectal cancer patients with high level of microsatellite instability or with defects in one of the MMR genes (12), might be also explored in the future. In this study, we reported the peculiar clinical and molecular characterization of a 35-year-old male patient who died of a metastatic FTC refractory to surgery and radioiodine and Sorafenib treatment and harboring *PTEN* and *TP53* mutations in a context of tumor heterogeneity and genomic instability.

Case description

In January 2008, a 35-year-old man without a family history of benign/malignant thyroid diseases or other tumors had noticed the appearance of a rapidly growing nodule in the neck. The patient had never had any relevant diseases or tumors, thus excluding the DICER1 syndrome, and had therefore never undergone external radiotherapy at the neck level. An ultrasound of the neck showed a hypoechogenic thyroid nodule of about 35 mm in the right lobe. Therefore, the patient underwent a thyroid needle aspiration for cytological examination. The patient was submitted, for a suspected cytological result, to total thyroidectomy. According to the WHO 2022 classification, histological examination showed a 35-mm right lobe extensively invasive, necrotic, and angioinvasive PDTC (80%) with widely invasive FTC areas with pleomorphic nuclei (20%) (pT3NX, according to the 7th TNM edition) (13, 14). The tissue analyzed shows Turin criteria (14): solid/trabecular/insular growth pattern, no nuclear cytology features of PTC, presence of tumor necrosis, mitotic count of 8/2 mm^2, convoluted nuclei, and absence of anaplastic features. In addition, Ki-67 immunostaining showed a proliferation index >5% in both the primary tumor and the distant metastasis (Supplementary Figure 1). Moreover, at diagnosis, a CT scan showed the presence of a small lung suspicious nodule and multiple metastatic lymph nodes (laterocervical, paratracheal, and mediastinal) and particularly a 5-cm lymph node metastasis located in the right hilar region across the main bronchus. In April 2008, 1,850 MBq of 131I was administered, and the total body scan showed two small thyroid residues and laterocervical adenopathy, showing instead a radioiodine refractoriness of mediastinal and lung metastases. In August 2008, 3 months after initial treatments, the patient was submitted to a thoracic surgery in order to remove progressive lung metastases and several enlarging mediastinal lymph nodes, conditioning dyspnea, chest pain, dysphagia, and fatigue. However, the metastases were unresectable, only a few lymph nodes were debulked, and the upper lobe of the right lung was removed. Histological examination showed massive lymph node and lung metastases of PDTC. Thus, in November 2008, due to the presence of progressive, symptomatic, iodine-refractory, unresectable distant metastases (Figure 1A), after informed



F

Date	Treatment	Suppressed Thyroglobulin level (ng/ml) *
January 2008	Total Thyroidectomy	40 (post-surgery)
April 2008	1850 MBq of 131I	55
August 2008	Surgical removal of metastatic mediastinal lymph nodes and upper lobe of right lung	200
November 2008	Sorafenib starting	1720
June 2009	Right pneumonectomy and mediastinal lymphadenectomy	49
August 2009	Disease progression	1550
September 2009	Disease-related death	4000

FIGURE 1
(A) CT scan performed on November 2008, before Sorafenib is started. Large lymph node metastases (between 2 and 5.5 cm in diameter) localized in the right hilar region across the pulmonary artery and the main bronchus. **(B)** Thyroglobulin (Tg) biomarker trend during Sorafenib treatment and after surgical removal of metastases performed after Sorafenib treatment (anti-Tg antibodies persistently negative): Tg values were significantly reduced in the first few weeks of Sorafenib therapy and remained low throughout treatment on different doses. In July 2009, after surgical removal of lymph node and lung metastases, Tg values rose suddenly. **(C, D)** CT scan performed on July 2009, after following Sorafenib withdrawal and surgical removal of lymph node and lung metastases. Ten liver metastases of 1–2.5 cm size and adrenal metastases (5 cm on the left and 4 cm on the right) were observed. **(E)** CT scan performed on January 2009, during Sorafenib treatment, showing significant reduction in lymph node metastases (maximum diameter 1 cm). **(F)** Chronological description of serum suppressed thyroglobulin levels and treatments carried out for thyroid carcinoma.

consent of the patient, we started treatment with Sorafenib, the only TKI available at the time (1, 15). Due to major drug-related side effects that developed (diarrhea, pruritus, fatigue, weight loss, hand–foot syndrome, musculoskeletal pain, and tachycardia), the maximum dosage reached by the patient was 600 mg/day, which was maintained without any withdrawal during 7 weeks. During Sorafenib treatment, lymph node and lung metastases showed a significant volume reduction, and serum thyroglobulin (Tg) decreased from 1,700 to 55 ng/ml (Figure 1B). The reduction of Sorafenib to 400 mg daily was associated with an immediate increase in Tg values, without a substantial reduction in side effects. The dose was then increased to 600 mg/day with a new decrease in Tg levels (Figure 1B). In June 2009, given the significant reduction of metastases burden (Figure 1C), the patient underwent right pneumonectomy and mediastinal lymphadenectomy. The histological examination showed lymph node and lung PDTC metastases. In July 2009, after surgery, serum Tg levels increased progressively, and a total body CT scan, negative for neck and chest metastases, showed multiple liver and adrenal metastases (Figures 1D–F). The patient's general conditions deteriorated rapidly, so no further treatments were possible, and he died in September 2009.

Molecular and Protein expression Analyses

At the time of diagnosis, the only molecular analyses available in our laboratory concerned the search for *BRAF* and *RAS* mutations and *ret/PTC* rearrangements, which were negative in this patient. Therefore, now that the molecular analysis in our laboratory has been extended to numerous genes involved in thyroid carcinogenesis by means of both mass spectrometry and Sanger sequencing, we have decided to re-analyze this particular case. The molecular analysis of the DNA obtained from the formalin-fixed paraffin-embedded

(FFPE) primary TC revealed the presence of a frameshift variant in exon 7 of *PTEN* (c.741dupA, p.P248Tfs*5) (Figure 2A) and a missense mutation in exon 5 of *TP53* gene (c.404G>A, p.C135Y) in heterozygosis, both absent in the corresponding contralateral normal thyroid tissue (*data not shown*). Analyzing different frozen primary TC sections, we confirmed the presence of the *PTEN* p.P248Tfs*5 mutation in all specimens (*data not shown*). On the other hand, a heterogeneous *TP53* molecular profile was observed in these samples: one harbored the *TP53* p.C135Y mutant together with a splicing mutation in intron 8 (c.920-2A>G), the second harbored the *TP53* p.C135Y variant together with another missense variant (c.398T>G, p.M133R), the third sample had only the p.M133R mutation, and the last had no *TP53* mutations (Figure 2B). Evaluating the presence of these variants in patient's FFPE lung and lymph node metastatic samples, the *PTEN* frameshift was detected in all tissues obtained pre- and post-Sorafenib treatment but, in the latter, at very low allelic frequency (Figure 2A). On the contrary, the two *TP53* missense mutations in exon 5 were absent in all metastatic samples analyzed, while the splicing variant c.920-2A>G was present in lung metastases obtained both pre- and post-treatment and in the lymph node metastasis before starting Sorafenib (Figure 2C). Interestingly, at RNA level the *TP53* p.C135Y mutation was found in homozygosis in the primary TC tissue (Figure 2D). The DNA recovered from the sample harbored the p.C135Y mutation in heterozygosis.

We then investigated the expression of p53 at protein level by immunohistochemistry. A peculiar pattern was observed in tissues analyzed: p53 was abnormal/over-expressed with high nuclear expression in primary TC (both FTC and PDTC areas) and normal in the contralateral normal thyroid tissue (Figure 3A). As far as metastases are concerned, all metastatic tissues had an abnormal/cytoplasmatic p53 staining with low nuclear expression (Figures 3B–D) with the exception of the lymph node obtained after Sorafenib treatment showing an abnormal/over-expressed p53 pattern with high nuclear expression (Figure 3E).

Finally, evaluating the expression of mismatch repair proteins (MMRs), a positive nuclear staining was observed for all these markers (MHL1, MSH2, MSH6, and PMS2) in the tissues analyzed (Figures 4A–C), although not all nuclei in the primary tumor showed the expression of MMR antigens (Figure 4A). It is interesting to note that there is an almost total loss of MMR proteins in the tumor area corresponding to PDTC (Supplementary Figure S2). The immunostaining score for each protein analyzed is reported in Supplementary Table S1.

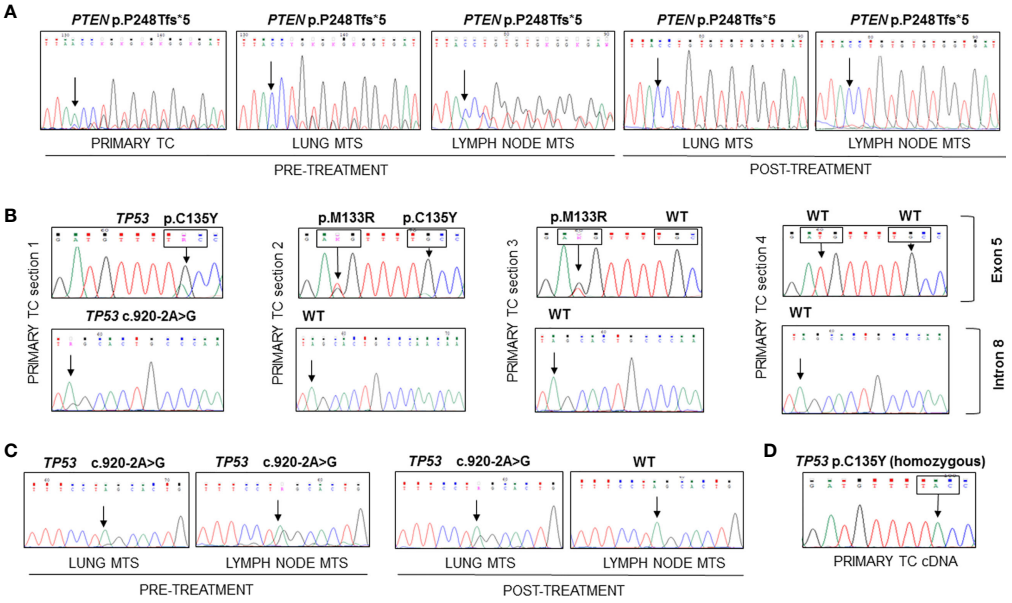


FIGURE 2
Sequencing of the PCR amplicons corresponding to exon 7 of PTEN and to exons 5 and 9 of TP53 genes and of RT-PCR amplicon corresponding to exon 5 of TP53 in FFPE and frozen samples. **(A)** The PTEN p.P248Tfs*5 mutation was found in heterozygosity in the DNA extracted from FFPE primary TC and lung and lymph node metastases samples obtained before and after Sorafenib treatment. **(B)** Different mutational patterns observed for TP53 in four sections of primary TC. The first sample showed the presence of TP53 p.C135Y mutation in exon 5 and c.920-2A>G splicing variant in intron 8, the second harbored two TP53 mutations (p.M133R and p.C135Y), the third had only the p.M133R mutant, while the last had no mutations. **(C)** The c.920-2A>G splicing variant was found in all metastases analyzed with the exception of the lymph node metastasis obtained after Sorafenib treatment. **(D)** The analysis of TP53 transcript encompassing exon 5 showed the presence of the C135Y mutation in homozygous state in the cDNA obtained from frozen primary TC. TC, thyroid cancer; MTS, metastasis.

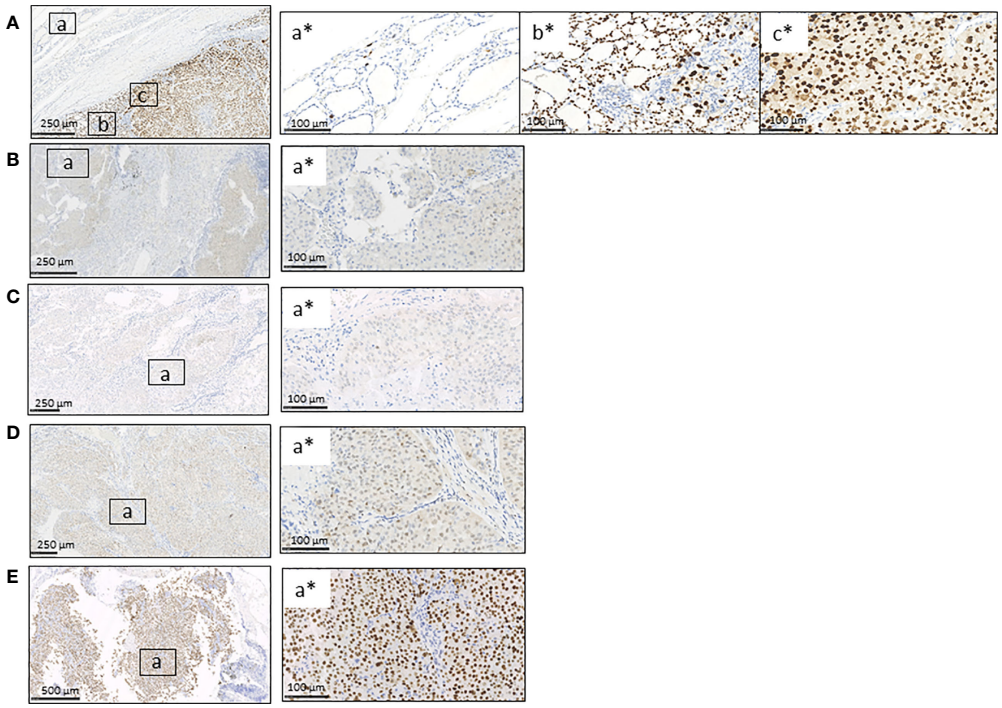


FIGURE 3
Immunohistochemistry for p53 protein in FFPE primary TC, contralateral normal thyroid tissue, and lung and lymph node metastases obtained before and after Sorafenib treatment. **(A)** The immunostaining for p53 was wild type in the contralateral normal thyroid tissue (inset a*) and abnormal/over-expressed with high nuclear expression in primary TC, both FTC/PDCT (inset b*) and PDTC areas (inset c*). **(B–D)** Lung metastases obtained before **(B)** and after Sorafenib treatment **(D)** and lymph node metastasis before TKI **(C)** showed an abnormal/cytoplasmatic p53 staining with low nuclear expression (for each is shown a selected inset area). **(E)** The lymph node metastasis obtained after the TKI treatment showed an abnormal/over-expressed p53 pattern with high nuclear expression. FTC, follicular thyroid cancer; PDTC, poorly differentiated thyroid cancer; NT, normal thyroid; TKI, tyrosine kinase inhibitor. Scale bar are shown for each images.

Interestingly, the analysis of microsatellite instability (MSI) status showed no amplification of BAT25 and BAT26 loci in the primary TC (Supplementary Table S2). Copy number variation (CNV) analysis showed the presence of a loss of heterozygosity (LOH) for both loci in the primary TC (Figure 4D). On the other hand, a high MSI for D2S123 and D5S346 loci was found only in the lymph node metastasis obtained after Sorafenib treatment, as shown in Figure 4E.

Discussion

To date, the molecular mechanisms associated with TC aggressiveness and resistance to TKIs treatments are not well understood. In the present study, we reported the case of a 35-year-old male patient with metastatic poorly differentiated (PDTC) and follicular thyroid carcinoma (FTC) treated with surgery, radioiodine, and Sorafenib. The patient, despite an initial partial response, died of progressive disease 21 months after diagnosis.

Molecular analyses of primary tumor and metastatic tissues showed the co-occurrence of a *PTEN* frameshift variant (p.P248Tfs*5) together with three *TP53* mutations (p.C135Y, p.M133R, and c.920-2A>G) in some, but not all, samples analyzed. We are tempted to speculate that TP53 mutations occur in poorly differentiated tumor area, since 80% of the primary TC is composed by PDTC. The co-occurrence of both *PTEN* and *TP53* mutations in thyroid and other cancers has already been reported (16, 17). In particular, the *PTEN* p.P248Tfs*5 variant was previously identified at somatic level in several solid tumors (COSM4986), but never in sporadic TC. On the other hand, the *TP53* p.C135Y variant and the splicing mutation c.920-2A>G were already described in ATC [3, 18] and in other solid and hematopoietic tumors (COSM10801 and COSM33650). Finally, the *TP53* p.M133R variant was reported at germline level in Li–Fraumeni syndrome and at somatic level in solid and hematopoietic tumors (COSM43730), but never in TC. It is possible to speculate that TP53 mutations.

Interestingly, the *PTEN* mutation is constantly found in primary TC, and all metastases analyzed were consistent with its clonal origin. It is worth to note that the allelic frequency of this

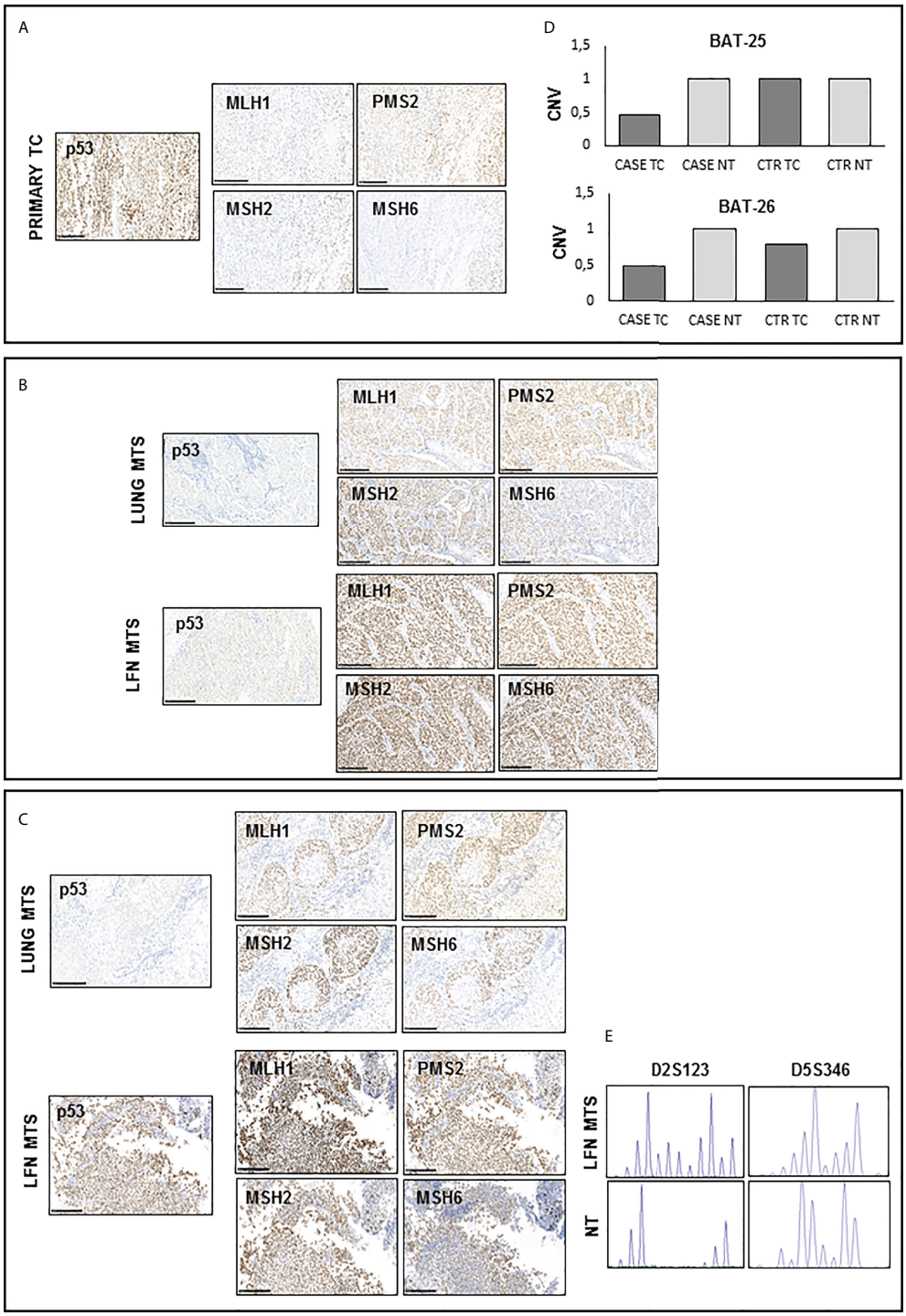


FIGURE 4
Immunohistochemistry for p53 and DNA mismatch repair (MMR) proteins and microsatellite instability (MSI) detection in available samples before and after Sorafenib treatment **(A, C, D)** A positive nuclear staining was observed for MHL1, MSH2, MSH6, and PMS2 proteins in all analyzed tissues samples. However, primary TC shows light staining and subclonal loss of expression in all MMR proteins. **(B)** CNV analysis shows the presence of a LOH for both BAT-25 and BAT-26 loci in the primary TC of our patient. **(E)** Peak analysis of D2S123 and D5S346 loci, performed using Genemapper 5 software, is clearly differently shaped in the lymph node obtained after Sorafenib treatment with respect to those of and contralateral normal thyroid tissue DNA, indicating a high microsatellite instability. TC, thyroid cancer; MTS, metastasis; NT, normal thyroid; CNV, copy number variation.

mutation is lower in metastases obtained after Sorafenib treatment with respect to other samples, indicating a potential effect of the TKI treatment on tumor clones harboring the *PTEN* frameshift. On the other hand, a variable mutational pattern for *TP53* was observed in primary TC and metastases samples (sections with either one or two or three mutations or entirely wild type for TP53). We excluded that these mutations are passenger ones, as they were previously reported in TC and other cancers and known to be non-functional pathogenic variants (p.C135Y and p.M133R, https://tp53.isb-cgc.org) or predicted to be likely pathogenic/pathogenic (c.920-2A>G, https://www.ncbi.nlm.nih.gov/clinvar). The finding of three mutations affecting the *TP53* gene is not surprising, since FTC exhibits remarkable genomic instability evidenced by *TP53* hyper-mutability (5). *TP53* mutations seem to be subclonal, each present in only a subset of malignant cells, contributing to heterogeneity within the tumor and potentially to treatment resistance (19, 20). Indeed, preclinical and clinical evidence suggests that cancer harboring *TP53* mutations are often resistant to TKI inhibitors (6, 21–25). Intriguingly, the p.C135Y mutation known to be a non-functional, dominant negative hot spot mutant (https://tp53.isb-cgc.org) (26) was found in homozygosis at RNA level in primary TC, suggesting a monoallelic expression (MAE) of the mutated allele. LOH of cancer-associated genes at DNA level is a common and important mechanism in carcinogenesis, but MAE at RNA level is a much less understood phenomenon. MAE may precede or enhance a mutation by expression of only the mutant or disease-related allele, having a role in tumor progression and clinical implications. High rate of MAE was previously observed in progressive brain tumors harboring *TP53* mutations (27, 28), but never in TC. The finding of the *TP53* p.M133R mutation in our patient is also intriguing, since germline mutations affecting the codon 133 cause the loss of the Δ133p53 isoforms and are frequently implicated in the development of Li–Fraumeni and Li–Fraumeni-like cancer predisposing syndromes (29). Although the precise functions of these isoforms is still poorly understood, syndromic forms of breast cancer are strongly associated with the loss of codon 133, indicating that the expression of Δ133p53 isoforms is critical for regulating p53 activity and carcinogenesis in some tissues (30). The presence of these two *TP53* missense mutations within exon 5 in primary TC is in agreement with the nuclear p53 protein expression pattern. On the other hand, the *TP53* c.920-2A>G splicing mutant is likely unable to enter the nucleus and accumulate in the cytoplasm, and, indeed, an abnormal/cytoplasmatic p53 staining was observed in all tissues harboring this mutation. The lymph node metastasis obtained post-Sorafenib is the only tissue that, despite an abnormal/over-expressed p53 staining with high nuclear expression, does not harbor *TP53* mutations. For this sample, we cannot exclude the presence of mutations in intronic/regulatory regions or a LOH of the *TP53* gene as far as the presence of alterations in other proteins of the DNA repair pathway that finally cause the inactivation of p53 protein. The possible involvement of defects in the DNA MMR proteins was also suspected, since the patient had multiple mutations. In our study, MMR proteins resulted to be functional in all tissue samples, although the primary TC showed a partial loss of MMR expression as already reported in a FTC case (31). Defects in MMR proteins are responsible for genomic instability, which can be evidenced by alterations in microsatellites markers. The primary TC, in which we detected a partial loss of MMR proteins, showed an LOH for BAT-25 and BAT-26 microsatellites. LOH for these loci were not shared by any metastatic site, but surprisingly, the lymph node metastasis obtained after Sorafenib treatment showed a MSI-High (MSI-H) for other two microsatellites (D2S123 and D5S346), but a normal MMR proteins expression. The discordant pattern of CNV for microsatellites between primary and metastatic sites may be explained by independent clonal evolution selected during the metastatic process. Moreover, the finding of MSI-H and MMR-proficient in the lymph node metastasis after Sorafenib is not surprising. Indeed, it is possible that some missense mutations in MMR genes can lead to functional inactivation of the corresponding protein without affecting its stability, antigenicity and expression level (32) or that some MSI-H tumors derive from alterations of MMR pathway-related proteins are not detectable by current technologies (33). It is exciting to find that the two tumor samples with abnormal nuclear expression of p53 protein (primary tumor and lymph node metastases after Sorafenib) show a high genomic instability, highlighting a strong relationship between MMR and the role of p53 in regulation of the cell-cycle arrest/apoptosis decision processes when DNA damage overwhelms a critical threshold. As recently reported (34, 35), subclonal expansions seem common in thyroid cancer cases with aberrant DNA repair with a selection of highly aggressive clones that will progress as what we observed in our patient.

Unfortunately, the patient's general conditions deteriorated rapidly after surgery in June 2009, and thus, distant liver and adrenal metastases were not available for further molecular characterization. It is well-known that TKIs exert their effect through a cytostatic action, and, once started, a continuative treatment is needed to maintain a response to the disease. When the TKI treatment is stopped, as it happened for our patient, the escape phenomenon is observed, and the progression of the disease can become even more rapid.

Conclusion

In conclusion, we demonstrated the presence of genomic heterogeneity and instability in a patient with metastatic poorly differentiated (PDTC) and follicular thyroid carcinoma (FTC) refractory to all treatments. The rare combination of *PTEN* and *TP53* mutations seems to be associated with a particular tumor aggressiveness and maybe with a possible resistance to TKI. This

case report highlights the importance of characterizing both primary tumor and metastases at molecular level to predict the response to treatments. Indeed, the tumor heterogeneity can evolve during tumor progression or as a consequence of drug-dependent selection of a pre-existing or newly acquired resistant clones. For this reason, further studies are needed, and new therapeutic strategies will be explored, such as drugs able to restore the transcriptional activity of mutant p53 or immune checkpoint inhibitors useful in cancers with high level of microsatellite instability or with defects in MMR genes.

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

This study was reviewed and approved by Istituo Auxologico Italiano. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CC provided clinical details. GP performed the molecular analysis of other genes, the analysis of microsatellites status, and of copy number variation. DS performed immunohistochemistry. MM performed PTC mass array. GB revised immunohistochemical results. VC conceived and designed the study; she performed molecular analysis of TP53 gene and transcript. VC, CC, LP, and LF wrote, edited, and reviewed the manuscript. All authors were involved in analyzing the data, writing the paper, and had final approval of the submitted and published versions.

Funding

This study was partially funded by Ricerca Corrente Istituto Auxologico Italiano IRCCS (PTC-array, 05C825_2018 and THYCANC, 2022_03_08_03) and by the Italian Ministry of University and Research (PRIN 2017-2017YTWKWH).

Acknowledgments

The author acknowledge the support of the APC central fund of the university of Milan.

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.2022.949098/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Hematoxylin and eosin (H&E) and Ki-67 stainings performed on the primary thyroid cancer section showing Turin criteria. H&E staining showed the presence of **(A)** pleomorphic nuclei typical of follicular thyroid carcinoma, **(B)** convoluted nuclei, **(C)** tumor necrosis, **(D)** and a mitotic count of 8/2 mm^2 (black circle).

SUPPLEMENTARY FIGURE 2
Representative images of the immunohistochemistry for the DNA mismatch repair proteins MSH6 and MHL1 obtained for the primary thyroid cancer. **(A)** Both MHS6 and MLH1 were expressed in all nuclei of follicular cells in the tumor area corresponding to 20% FTC (**A** and **C**, respectively). On the other hand, the almost total loss of both MSH6 and MLH1 expression was observed in the nuclei of follicular cells in the tumor area corresponding to 80% PDTC (**B** and **D**, respectively).

References

1. Fugazzola L, Elisei R, Fuhrer D, Jarzab B, Leboulleux S, Newbold K, et al. European Thyroid association guidelines for the treatment and follow-up of advanced radioiodine-refractory thyroid cancer. *Eur Thyroid J* (2019) 8:227–45. doi: 10.1159/000502229

2. Kunstman JW, Juhlin CC, Goh G, Brown TC, Stenman A, Healy JM, et al. Characterization of the mutational landscape of anaplastic thyroid cancer *via* whole-exome sequencing. *Hum Mol Genet* (2015) 24:23182329. doi: 10.1093/hmg/ddu749

3. Landa I, Ibrahimpasic T, Boucai L, Sinha R, Knauf JA, Shah RH, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest* (2016) 126:1052–66. doi: 10.1172/JCI85271

4. Jänicke RU, Graupner V, Budach W, Essmann F. The do's and don'ts of p53 isoforms. *Biol Chem* (2009) 390:951–63. doi: 10.1515/BC.2009.093

5. Shahedian B, Shi Y, Zou M, Farid NR. Thyroid carcinoma is characterized by genomic instability: evidence from p53 mutations. *Mol Genet Metab* (2001) 72:155–63. doi: 10.1006/mgme.2000.3114

6. Canale M, Petracci E, Delmonte A, Chiadini E, Dazzi C, Papi M, et al. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. *Clin Cancer Res* (2017) 23:2195–202. doi: 10.1158/1078-0432.CCR-16-0966

7. Chen J, Sadowski I. Identification of the mismatch repair genes PMS2 and MLH1 as p53 target genes by using serial analysis of binding elements. *Proc Natl Acad Sci USA* (2005) 102:4813–8. doi: 10.1073/pnas.0407069102

8. Worby CA, Dixon JE. PTEN. *Annu Rev Biochem* (2014) 83:641–69. doi: 10.1146/annurev-biochem-082411-113907

9. Liu Z, Hou P, Ji M, Guan H, Studeman K, Jensen K, et al. Highly prevalent genetic alterations in receptor tyrosine kinases and phosphatidylinositol 3-kinase/akt and mitogen-activated protein kinase pathways in anaplastic and follicular thyroid cancers. *J Clin Endocrinol Metab* (2008) 93:3106–16. doi: 10.1210/jc.2008-0273

10. Tirrò E, Martorana F, Romano C, Vitale SR, Motta G, Di Gregorio S, et al. Molecular alterations in thyroid cancer: From bench to clinical practice. *Genes (Basel)* (2019) 10:709. doi: 10.3390/genes10090709

11. Manzella L, Stella S, Pennisi MS, Tirrò E, Massimino M, Romano C, et al. New insights in thyroid cancer and p53 family proteins. *Int J Mol Sci* (2017) 18:1325. doi: 10.3390/ijms18061325

12. André T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in microsatellite-Instability-High advanced colorectal cancer. *N Engl J Med* (2020) 383:2207–18. doi: 10.1056/NEJMoa2017699

13. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. American Joint committee on cancer (AJCC) cancer staging manual. *7th Ed Chicago: Springer Inc* (2010) 7:97–100.

14. Baloch ZW, Asa SL, Barletta JA, Ghossein RA, Juhlin CC, Jung CK, et al. Overview of the 2022 WHO classification of thyroid neoplasms. *Endocr Pathol* (2022) 33:27–63. doi: 10.1007/s12022-022-09707-3

15. Brose MS, Nutting CM, Jarzab B, Elisei R, Siena S, Bastholt L, et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet* (2014) 384:319–28. doi: 10.1016/S0140-6736(14)60421-9

16. Sadow PM, Dias-Santagata D, Zheng Z, Lin DT, Le LP, Nucera C. Identification of insertions in PTEN and TP53 in anaplastic thyroid carcinoma with angiogenic brain metastasis. *Endocr Relat Cancer* (2015) 22:L23–8. doi: 10.1530/ERC-15-0198

17. Wei S, LiVolsi VA, Montone KT, Morrissette JJD, Baloch ZW. PTEN and TP53 mutations in oncocytic follicular carcinoma. *Endocr Pathol* (2015) 26:365–9. doi: 10.1007/s12022-015-9403-6

18. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* (2015) 7:283ra54. doi: 10.1126/scitranslmed.aaa1408

19. Gasparyan M, Lo MC, Jiang H, Lin CC, Sun D. Combined p53- and PTEN-deficiency activates expression of mesenchyme homeobox 1 (MEOX1) required for growth of triple-negative breast cancer. *J Biol Chem* (2020) 295:12188–202. doi: 10.1074/jbc.RA119.010710

20. Chang GC, Hsu SL, Tsai JR, Liang FP, Lin SY, Sheu GT, et al. Molecular mechanisms of ZD1839-induced G1-cell cycle arrest and apoptosis in human lung adenocarcinoma A549 cells. *Biochem Pharmacol* (2004) 68:1453–64. doi: 10.1016/j.bcp.2004.06.006

21. Munsch D, Watanabe-Fukunaga R, Bourdon JC, Nagata S, May E, YonishRouach E, et al. Human and mouse fas (APO-1/CD95) death receptor genes each contain a p53-responsive element that is activated by p53 mutants unable to induce apoptosis. *J Biol Chem* (2000) 275:3867–72. doi: 10.1074/jbc.275.6.3867

22. Rho JK, Choi YJ, Ryoo BY, Na II, Yang SH, Kim CH, et al. p53 enhances gefitinib-induced growth inhibition and apoptosis by regulation of fas in non-small cell lung cancer. *Cancer Res* (2007) 67:1163–9. doi: 10.1158/0008-5472.CAN-06-2037

23. Li J, Lin B, Li X, Tang X, He Z, Zhou K. Biomarkers for predicting response to tyrosine kinase inhibitors in drug-sensitive and drug-resistant human bladder cancer cells. *Oncol Rep* (2015) 33:951–7. doi: 10.3892/or.2014.3639

24. MacLaine NJ, Wood MD, Holder JC, Rees RW, Southgate J. Sensitivity of normal, paramalignant, and malignant human urothelial cells to inhibitors of the epidermal growth factor receptor signaling pathway. *Mol Cancer Res* (2008) 6:53–63. doi: 10.1158/1541-7786.MCR-07-0134

25. Hsu SP, Lin PH, Chou CM, Lee WS. Progesterone up-regulates p27 through an increased binding of the progesterone receptor-A-p53 protein complex onto the non-canonical p53 binding motif in HUVEC. *J Steroid Biochem Mol Biol* (2019) 185:163–71. doi: 10.1016/j.jsbmb.2018.08.011

26. Savova V, Vigneau S, Gimelbrant AA. Autosomal monoallelic expression: genetics of epigenetic diversity? *Curr Opin Genet Dev* (2013) 23:642–8. doi: 10.1016/j.gde.2013.09.001

27. Walker EJ, Zhang C, Castelo-Branco P, Hawkins C, Wilson W, Zhukova N, et al. Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res* (2012) 72:636–44. doi: 10.1158/0008-5472.CAN-11-2266

28. Kouidou S, Malousi A, Maglaveras N. Li-Fraumeni and Li-fraumeni-like syndrome mutations in p53 are associated with exonic methylation and splicing regulatory elements. *Mol Carcinog* (2009) 48:895–902. doi: 10.1002/mc.20537

29. Kouidou S, Malousi A, Kyventidis A, Fragou A, Maglaveras N. G:C > A:T mutations and potential epigenetic regulation of p53 in breast cancer. *Breast Cancer Res Treat* (2007) 106:351–60. doi: 10.1007/s10549-007-9514-y

30. Genutis LK, Tomsic J, Bundschuh RA, Brock PL, Williams MD, Roychowdhury S, et al. Microsatellite instability occurs in a subset of follicular thyroid cancers. *Thyroid* (2019) 29:523–9. doi: 10.1089/thy.2018.0655

31. Baretti M, Le DT. DNA Mismatch repair in cancer. *Pharmacol Ther* (2018) 189:45–62. doi: 10.1016/j.pharmthera.2018.04.004

32. Li K, Luo H, Huang L, Luo H, Zhu X. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int* (2020) 20:16. doi: 10.1186/s12935-019-1091-8

33. Paulsson JO, Backman S, Wang N, Stenman A, Crona J, Thutkawkorapin J, et al. Whole-genome sequencing of synchronous thyroid carcinomas identifies aberrant DNA repair in thyroid cancer dedifferentiation. *J Pathol* (2020) 250:183–94. doi: 10.1002/path.5359

34. Pstrąg N, Ziemnicka K, Bluyssen H, Wesoły J. Thyroid cancers of follicular origin in a genomic light: in-depth overview of common and unique molecular marker candidates. *Mol Cancer.* (2018) 17:116. doi: 10.1186/s12943-018-0866-1

35. Wagle N, Grabiner BC, Van Allen EM, Amin-Mansour A, Taylor-Weiner A, Rosenberg M, et al. Response and acquired resistance to everolimus in anaplastic thyroid cancer. *N Engl J Med* (2014) 371:1426–33. doi: 10.1056/NEJMoa1403352

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 September 2022
DOI 10.3389/fonc.2022.963910

OPEN ACCESS

EDITED BY
Giovana Tardin Torrezan,
A.C.Camargo Cancer Center, Brazil

REVIEWED BY
Raymond H. Kim,
University of Toronto, Canada
Julie Dutil,
Ponce Health Sciences University,
Puerto Rico

*CORRESPONDENCE
Ana Carolina Rathsam Leite
anarathsam@gmail.com

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

RECEIVED 08 June 2022
ACCEPTED 08 August 2022
PUBLISHED 05 September 2022

CITATION
Leite ACR, Suzuki DA, Pereira AAL, Machado NP, Barroso-Sousa R, Correa TS, Moura FC, Morbeck IAP, Gumz BP, Faria LDBB, Fernandes GdS and Sandoval RL (2022) What can we learn from more than 1,000 Brazilian patients at risk of hereditary cancer? *Front. Oncol.* 12:963910. doi: 10.3389/fonc.2022.963910

What can we learn from more than 1,000 Brazilian patients at risk of hereditary cancer?

Ana Carolina Rathsam Leite*, Daniele Assad Suzuki, Allan Anderson Lima Pereira, Natalia Polidorio Machado, Romualdo Barroso-Sousa, Tatiana Strava Correa, Fernanda Cesar Moura, Igor Alexandre Protzner Morbeck, Brenda Pires Gumz, Luiza Dib Batista Bugiato Faria, Gustavo dos Santos Fernandes and Renata Lazari Sandoval

Department of Oncology, Hospital Sírio-Libanês, Brasília, Distrito Federal, Brazil

Background: Identifying individuals at a higher risk of developing cancer is a major concern for healthcare providers. Cancer predisposition syndromes are the underlying cause of cancer aggregation and young-onset tumors in many families. Germline genetic testing is underused due to lack of access, but Brazilian germline data associated with cancer predisposition syndromes are needed.

Methods: Medical records of patients referred for genetic counseling at the Oncogenetics Department at the Hospital Sírio-Libanês (Brasília, DF, Brazil) from July 2017 to January 2021 were reviewed. The clinical features and germline findings were described. Detection rates of germline pathogenic/likely pathogenic variant (P/LPV) carriers were compared between international and Brazilian guidelines for genetic testing.

Results: A total of 1,091 individuals from 985 families were included in this study. Most patients (93.5%) had a family history of cancer, including 64% with a family member under 50 with cancer. Sixty-six percent of patients (720/1091) had a personal history of cancer. Young-onset cancers (<50 years old) represented 62% of the patients affected by cancer and 17% had multiple primary cancers. The cohort included patients with 30 different cancer types. Breast cancer was the most prevalent type of cancer (52.6%). Germline testing included multigene panel (89.3%) and family variant testing (8.9%). Approximately 27% (236/879) of the tested patients harbored germline P/LPVs in cancer susceptibility genes. *BRCA2*, *BRCA1*, and *TP53* were the most frequently reported genes, corresponding to 18.6%, 14.4%, and 13.5% of the positive results, respectively. Genetic testing criteria from international guidelines were more effective in identifying carriers than the Brazilian National Agency of Supplementary Health (ANS) criteria (92% vs. 72%, p<0.001). Forty-six percent of the cancer-unaffected patients who harbored a germline P/LPV (45/98) would not be eligible for genetic testing according to

ANS because they did not have a family variant previously identified in a cancer-affected relative.

Conclusion: The high detection rate of P/LPVs in the present study is possibly related to the genetic testing approach with multigene panels and cohort's characteristics, represented mainly by individuals with a personal or family history of young-onset cancer. Testing asymptomatic individuals with suspicious family history may also have contributed to a higher detection rate. A significant number of carriers would not have been identified using ANS criteria for genetic testing.

KEYWORDS

hereditary cancer, cancer predisposition, multigene analyses, genetic testing access, cancer risk assessment

Introduction

Carriers of cancer predisposition syndromes (CPSs) are at a higher risk of developing cancer. Familial aggregation, early-onset cancer, and the risk of multiple primary cancers are shared characteristics among CPSs (1, 2). Genetic counseling, modified surveillance, and risk-reduction strategies are essential in these scenarios. Therefore, health professionals involved in comprehensive health care, especially in the diagnosis and treatment of cancer, must be able to identify individuals at risk of hereditary cancer.

Clinical criteria used to be the main diagnostic tool for CPSs (3–5). Nevertheless, the discovery of cancer susceptibility genes (6) and the decreasing costs of DNA sequencing created a pathway for genetic testing implementation in the diagnostic framework. Genetic testing criteria have evolved rapidly in recent years (7–9). Despite this progress, global disparities exist, and access remains a critical concern (10).

Genetic counseling access, genetic testing costs, and lack of epidemiological hereditary cancer data are barriers to cancer predisposition assessments in Brazil and other Latin American countries (11). In Brazil, health insurance coverage for genetic testing was initiated in 2018. Although this coverage does not include all clinical scenarios eligible for genetic testing according to current international guidelines, it was the beginning of genetic testing access, at least for the Brazilian population with health insurance. Unfortunately, only 25% of Brazilians have health insurance, therefore genetic testing is not yet available for most citizens who depend on the public health system (12).

Epidemiological data are paramount to understand demands and opportunities for cost-effective interventions and resource allocation. The present study explores regional epidemiological data from Brazilian patients at risk for hereditary cancer. We also aimed to compare national and international guidelines criteria for germline genetic testing.

Methods

Individual patient data were retrospectively collected from medical records of patients referred to the Oncogenetics Department at Hospital Sírio-Libanês (Brasília, Federal District, Brazil) for genetic counseling between July 2017 and January 2021. Patients with a personal history of cancer and/or family history of cancer were included in the analysis. A waiver of informed consent was approved by the Institutional Review Board of Hospital Sírio-Libanês.

Data were anonymized by removing all patient identifiers. The collected data included sex, age at cancer diagnosis, cancer type, number of primary cancers, family history of cancer, and germline genetic test results. Family cancer history was obtained through pedigree analysis and/or information from proband's medical records. Any cancer in first-, second-, or third-degree relatives was considered a positive family history of cancer. The concept of limited family structure proposed by Weitzel et al. (13), was adapted for this study. Limited family structure was defined as fewer than two first- or second-degree relatives surviving past 45 in either lineage, maternal or paternal. Patients with an unknown family history were also classified as having a limited family structure.

Criteria for germline genetic testing were revised according to national and international guidelines: (i) testing criteria published by the Brazilian National Agency of Supplementary Health (ANS); and (ii) NCCN Clinical Practice Guidelines on Oncology: genetic/familial high-risk assessment for breast, ovarian, and pancreatic cancer (version 2.2021) (14) and genetic/familial high-risk assessment for colorectal cancer

(version 1.2021) (15). For hereditary diffuse gastric cancer, we used updated clinical practice guidelines proposed by Gullo etal. (16). Information about the commercial laboratory that performed the germline test, the testing methodology, and the number of genes evaluated were also collected. The classification of the variants described in this paper are those reported by the respective laboratories.

Statistical analyses

Data were tested for normality using the Kolmogorov–Smirnov and Shapiro–Wilk tests. Values are expressed as medians and percentiles for non-normal continuous variables and as means and standard deviations for normal continuous variables. Categorical data are presented as absolute values and percentages and were tested using the Pearson X2 and Fisher exact tests. Quantitative data were compared by applying Student's t-test to compare the two groups for normally distributed variables and the Mann–Whitney U test for non-normally distributed ones. Statistical significance was set at $P \leq 0.05$. Statistical analyses were performed using SPSS version 21.0 (IBM, NY, USA).

Results

General characteristics of the studied population

In total, 1,091 individuals from 985 families were included in this study. Female patients represented 83.6% (912/1091) of the cohort. At the first genetic counseling session, 66.0% (720/1091) of patients had a personal history of cancer. Thirty-one percent of patients (346/1091) were cancer-unaffected, 2.7% were under investigation for a malignant disease, 10.0% had a recent diagnosis of cancer, 13.0% were receiving oncological treatment (radiotherapy or chemotherapy), 34.6% had already completed cancer treatment, and 5% had metastatic disease. Disease status information was unavailable for 35 patients.

Most patients (93.5%, 1020/1091) had one or more family members affected by cancer. Most of these family members had cancer before the age of 50 years (64.0%, 653/1020). Twenty-three patients (2.1%) had a limited family structure or unavailable family history. One hundred and three patients (9.4%) were referred for genetic testing because of a previous identification of a familial germline pathogenic variant in a blood relative.

Of the 82.5% (900/1091) of patients who fulfilled international guidelines for genetic testing, 60.0% (655/900) were eligible for ovarian, breast, and pancreatic cancer genetic testing, 7.1% (78/900) for Lynch syndrome, 1.6% (18/900) for Li-Fraumeni syndrome, 1.4% (15/900) for adenomatous polyposis syndromes, and 2.8% (31/900) for other CPSs. Considering the ANS criteria, 57.5% (627/1091) of patients were eligible for genetic testing.

Cancer-affected patients

Approximately 62.0% (446/720) of patients with a personal history of cancer were under the age 50. The median age at first cancer diagnosis was 46 years (interquartile range [IQR], 37-55 years).

The cohort included patients with 30 different cancer types (Supplementary Material 1). Breast cancer was the most prevalent tumor (52.9%, 466/880), followed by colorectal cancer (8.3%), ovarian cancer (5.9%), thyroid cancer (3.7%), sarcoma (3.4%), renal cancer (3.2%), prostate cancer (2.6%), pancreatic cancer (2.4%), endometrial cancer (2.2%), neuroendocrine tumor (2.2%), melanoma (2.2%), and gastric cancer (2.0%).

Seventeen percent of patients (127/720) had multiple primary cancers. Of those, most patients had two primary cancers (81.0%, 103/127). A higher risk of multiple primary cancers was associated with young-onset cancers; however, this association was not statistically significant (p=0.263).

Germline testing results

Among the 879 patients who underwent germline genetic testing, 89.3% (n=785) underwent multigene panel testing. The remaining patients were tested using the following strategies: family variant testing (8.9%), BRCA1/BRCA2 gene testing (1.4%), and whole-exome sequencing (0.5%) (Figure 1). The four patients who underwent whole-exome sequencing were under investigation for CPSs and other disorders of genetic background, such as hereditary neuropathy, inborn errors of metabolism, and premature ovarian insufficiency.

One hundred and twenty pathogenic or likely pathogenic variants (P/LPVs) in 34 genes were detected in 26.9% (236/879) of the tested patients. According to ACMG's list of medically actionable genetic findings (17) and/or the list proposed by Desmond et al. (18), the prevalence of actionable P/LPVs would be 24.9% (n=219).

Variants of uncertain significance (VUS) were described in 402 patients (45.7%). Table 1 summarizes the clinical characteristics of patients who underwent germline testing. Most P/LPVs were found in cancer-affected patients (58.5% vs. 41.5%; p<0.001). Family members affected by cancer under the age of 50 and a history of multiple primary cancers were independently associated with germline P/LPVs identification (p< 0.05).

Among 269 cancer-unaffected patients who underwent genetic testing, 36.4% (n=98) harbored a P/LPV. Fifty-four

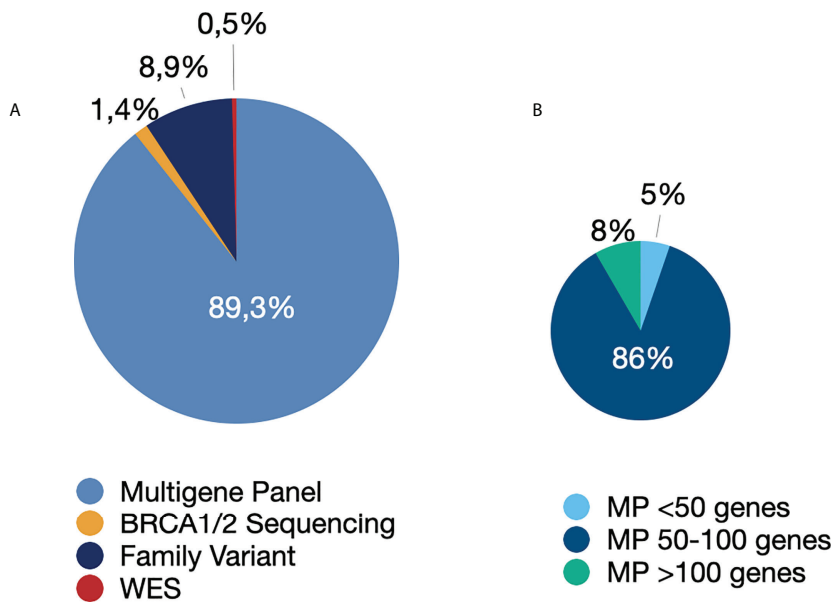


FIGURE 1
Genetic tests performed. **(A)** Type of genetic test performed. **(B)** Number of genes in the multigene panel. Abbreviations: MP, multigene panel; WES, whole-exome sequencing.

percent (53/98) of those patients were tested due to previous identification of the variant in another family member (cascade testing).

Heterozygous P/LPVs in BRCA2 (18.6%), BRCA1 (14.4%), TP53 (13.6%), MUTYH (9.7%), and CHEK2 (5.9%) were the most frequently reported genetic findings. Fourteen patients harbored a monoallelic P/LPV associated with recessive disorders (NTHL1, RECQL4, ERCC3, FANCA, and BLM). Table 2 shows the distribution of germline P/LPVs according to the cancer type.

Thirteen patients harbored two P/LPVs. One was homozygous for a pathogenic variant of MUTYH. Four patients (1.7%, 4/236) had an overlap of high/moderate penetrance P/LPVs for autosomal dominant CPSs (Supplementary Material 2). Eight patients harbored monoallelic variants in high/moderate penetrance cancer genes associated with autosomal dominant inheritance and a second variant in a gene associated with a recessive disorder (MUTYH, FANCA, NTHL1) or low penetrance cancer (TYR).

Testing criteria

Sixty-two percent of the tested patients (550/879) fulfilled both the international guidelines and ANS criteria, 19.9% fulfilled only the international guidelines criteria, and 8.1% underwent germline testing despite not meeting testing criteria.

Genetic testing criteria from international guidelines were more effective in identifying P/LPV carriers than the ANS criteria (92% vs. 72%, p<0.001). Both approaches would have missed some diagnoses, including approximately 10% (19/191) of patients screened by international guidelines and 14% (66/464) of those screened by ANS criteria.

Discussion

This study analyzed data of 1,091 Brazilian individuals from 985 different families, referred to genetic counseling due to personal and/or family history of cancer. This is the largest single center Brazilian cohort, from the Center-West of the country, that underwent germline genetic testing with multigene panels for hereditary cancer. Germline genetic tests guide high-risk surveillance, risk-reduction recommendations, and cancer treatment (15, 16, 19–23). The detection rate of germline P/LPVs varies according to criteria selection for testing and testing approaches (24–26). Although most Brazilian patients have limited access to hereditary cancer risk assessment (12, 27), our results provide some insights on genetic testing for hereditary cancer in Brazil.

The present cohort comprised a highly selected population with access to private healthcare and molecular testing. Most patients referred to genetic counseling met clinical criteria for germline testing (82.5%). A personal history of multiple primary cancers and family history of cancer under 50 were important predictors of a positive test result, in line with previous studies (28–30). International criteria more effectively identified P/LPV carriers in cancer susceptibility genes than ANS criteria (92% vs. 72%, p<0.001). Both criteria missed 10-14% of P/LPV carriers.

TABLE 1 Clinical profile of the patients who underwent germline testing for CPSs.

	Positive result N (%)	Negative/VUS resultN (%)	p value
Personal history of cancer Yes	138 (58.5)	472 (73.4)	<0.001
No	98 (41.5)	171 (26.6)	
Age 1st cancer diagnosis < 18 yrs	1 (0.7)	7 (1.5)	0.456
19- 35 yrs	26 (19.0)	79 (16.8)	
36- 45 yrs	42 (30.7)	137 (29.1)	
46- 49 yrs	12 (8.8)	68 (14.4)	
> 50 yrs	56 (40.9)	180 (38.2)	
Total	137	471	
N° of primary cancers 0	98 (41.5)	171 (26.6)	0.039
1	106 (45)	397 (61.7)	
2	25 (10.6)	64 (10)	
3	4 (1.7)	11 (1.7)	
4	2 (0.8)	0	
5	1 (0.4)	0	
Total	236	643	
Family history of cancer Yes	224 (94.9)	595 (92.5)	0.349
Negative	7 (3)	34 (5.3)	
Unknown	5 (2.1)	14 (2.2)	
Total	236	643	
Relatives affected < 50 yrs Yes	165 (69.9)	367 (57.1)	0.001
No	66 (28)	261 (40.6)	
Unknown	5 (2.1)	15 (2.3)	
Total	236	643	
Fulfill international criteria Yes	217 (91.9)	530 (82.4)	<0.001
No	19 (8.1)	113 (17.6)	
Total	236	643	
Fulfill ANS criteria Yes	170 (72)	380 (59.1)	<0.001
No	66 (28)	263 (40.9)	
Total	236	643	

VUS, variant of uncertain clinical significance; yrs, years; ANS, Brazilian National Agency of Supplementary Health.

Other studies have already highlighted that current testing criteria may not be able to identify all carriers (31–33). For this reason, some authors advocate for universal screening in some clinical scenarios (9, 34–36). Nevertheless, before advocating for universal genetic testing, we must ensure equitable access to interventions associated with positive test results (37).

In addition, cost-effectiveness of genetic testing may be affected by cascade testing, which involves identifying asymptomatic family members at risk (38). Our study demonstrated that 36.4% of the tested cancer-unaffected patients harbored a P/LPV in a CPS gene. Among these patients, only 54% had a previously identified family variant that made them eligible to pursue genetic testing according to ANS criteria. ANS only endorses germline testing for asymptomatic patients who have a relative with previous identification of a germline P/LPV. This finding should prompt discussion among Brazilian regulatory agencies and medical societies who are involved in revising national guidelines.

Approximately 27% of all tested patients harbored one or more PV/LPVs in cancer susceptibility genes, which is similar to study results from India (39) but higher than other studies (9, 40, 41). We attributed this difference to a highly selected sample including patients with previous identification of a family variant (9%), patients that did not meet genetic testing criteria

TABLE 2 Distribution of tumors according to germline P/LPVs identification.

	Number of tumors	P/LPVs (%)	Gene (N° of patients harboring germline P/LPVs)																																
			APC (4)	*APC (I1307K)* (2)	*ATM* (8)	*AXIN2* (2)	*BARD1* (3)	*BLM* (2)	*BRIP1* (3)	*BRCA1* (34)	*BRCA2* (44)	*CDH1* (1)	*CHEK2* (14)	*ERCC3* (1)	*FANCA* (1)	*FH* (1)	*MEN1* (2)	*MITF* (4)	*MLH1* (2)	*MSH2* (9)	*MSH6* (1)	*MUTYH mono* (23)	*MUTYH biallelic* (1)	*NF1* (2)	*NTHL1* (7)	*PALB2* (8)	*PMS2* (1)	*PRKAR1A* (1)	*RAD50* (2)	*RAD51C* (9)	*RAD51D* (2)	*RECQL4* (3)	*SDHA* (1)	*SDHB* (4)	*TP53* (32)	*TYR* (2)
			N° of tumors according to germline P/LPVs																																
Breast	407	23.3%			3	1	4		2	18	21	1	12					1			1	7			2	4			1	2	1	1		13	1
Colorectal	60	26.7%				1		1					3						3	5		1								1	1				
Ovarian	48	33.3%								5	3				1	1				2			1		2				1						
Thyroid	29	20.7%									1		3						1			1													
Sarcoma	24	29.2%									1											1												3	
Renal	17	17.6%											1									1												1	
Prostate	18	33.3%								1	2		1							1											1				
Pancreatic	20	20.0%			1								1												1						1				
Endometrial	16	43.8%								1	2							2				1						1							
NET	13	15.4%															2																		
Melanoma	14	21.4%																				1		1										1	
Gastric	12	8.3%																		1															
Other cancers*	61	32.8%		1	1			1		1	5		1			1							1									1	1	6	
No cancer	-	-	4	1	3	1			1	9	20		8	1				2		3		10		1	2	1	1		5		2		3	15	1

P/LP, pathogenic/likely pathogenic; VUS, variant of uncertain significance; NET, neuroendocrine tumor.

*Other cancers included testicular, lymphoma, leukemia, schwannoma, urothelial, bladder, adrenocortical, head and neck, skin (non-melanoma), soft tissue, lung, central nervous system, uterus, parotid, hepatocarcinoma, pheochromocytoma, appendix, gallbladder, and multiple myeloma.

Adapted from Samadder et al. (9).

Color shading is related to the frequency of each alteration, darker shades represent higher frequencies.

10.3389/fonc.2022.963910

(8.1%) and the use of large multigene panels (95% ≥ 50 genes). Universal genetic testing is indicated for some cancer types (e.g. epithelial ovarian cancer) and has been debated for other clinical scenarios (e.g. breast cancer, colorectal cancer) (7, 9, 42). Higher rates of positive genetic test results are achieved with the universal testing approach in comparison to the criteria-based (9, 31). In addition, Tsaousis et al. (43) demonstrated that depending on the number of genes included in the multigene panel, the identification of PVs can increase from 15.1% to 24.7%, and the higher range may be attributed to 4.5% of PVs in low-risk/limited data genes. In our cohort, fourteen patients harbored a monoallelic P/LPV associated with recessive disorders (NTHL1, RECQL4, ERCC3, FANCA, and BLM) and one patient harbored PV in a low penetrance cancer gene (TYR). The actionability in carriers of recessive disorders are related to reproductive risks.

The most frequently mutated genes in our cohort were BRCA2 (18.6%), BRCA1 (14.4%), TP53 (13.6%), and monoallelic MUTYH (9.7%). Interestingly, a recently published nationwide Brazilian study, with the largest breast cancer patient cohort (n= 1663) submitted to genetic testing, also described these genes as the most mutated among patients with positive genetic test results (44). In contrast to the cohort from Guindalini's paper, we included patients with different types of cancer. However, our sample was enriched by breast cancer patients; therefore, a high prevalence of P/LPVs in BRCA1/2 was expected. The high rate of P/LPVs in the TP53 gene described in our study is possibly related to the founder effect that the p.Arg337His (p.R337H) variant exerts in Brazil, and a possible selection bias associated with referrals to our team of specialists in Li-Fraumeni syndrome. This variant is found in up to 0.3% of the population of the southern and southeastern regions of the country (45), and in lower frequencies in other Brazilian regions (46, 47). Unlike BRCA1/2 and TP53, monoallelic MUTYH P/LPVs are not associated with breast cancer but may predict earlier colorectal screening in families affected by colorectal cancer.

Hereditary cancer awareness is growing rapidly. Professional education in hereditary cancer risk assessment, including multidisciplinary team training, strategies to optimize genetic counseling referrals, and genetic testing access, improve CPSs identification rates (11, 48). In Brazil and other low- or middle-income countries, the socioeconomic barrier impact health care access. Despite the worldwide advocacy for broad genetic testing access (49, 50), uninsured patients remain a concern. Providing access to genetic testing without assurance of all subsequent preventive and treatment opportunities may bring more harm than benefit (37). Continuous efforts in private and public settings should be made to pursue equitable hereditary cancer diagnosis and management.

Despite some limitations related to the retrospective nature of this study, as well as, the fact that it consisted of a highly selected sample from a single center, our results might form the basis for prospective studies and national collaborative efforts to achieve higher quality data that will impact policy makers.

Conclusion

The Brazilian ANS testing criteria should be revised to consider inclusion criteria for germline testing of cancer-unaffected patients with a suspected family history of CPS. Multigene panels provide high rates of P/LPV detection and should be considered a first-tier strategy. Hereditary cancer awareness among health care providers, genetic counseling training, and education for the proper interpretation of genetic test results, including understanding their clinical validity and utility, should be available in private and public settings.

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

The authors confirm contribution to the paper as follows: study conception and design: AL and RS; data collection: AL and RS; analysis and interpretation of results: AL and RS. Draft manuscript preparation: AL, DS, AP, NM, RB-S, TC, FM, IM, BG, LF, GF, and RS. All authors reviewed the results and approved the final version of the manuscript.

Conflict of interest

Author DAS participates in clinical studies sponsored by the company Lilly, and as a speaker at events for AstraZeneca, MSD, Novartis, Roche, Pfizer, Advisory, GSK, Roche. Author AALP reported research involvement with MSD; received honoraria from Servier, MERK, MSD, Novartis, Lilly, AstraZeneca; reported logistic support for educational meetings from AmGen/MSD/Roche; and had advisory role from Servier, Bayer, MERK. Author RB-S reported receiving speaker bureau fees from Agilant, AstraZeneca, Daichi-Sankyo, Eli Lilly, Pfizer, Novartis, Merck, and Roche; he has also served as a consultant/advisor for AstraZeneca, Daichi-Sankyo, Eli Lilly, Libbs, Roche, Merck and has received support for attending medical conferences from AstraZeneca, Eli Lilly, Daichi-Sankyo, and Merck. Author TSC participates in clinical studies sponsored

by the company Lilly, Pfizer, Novartis, BMS; and reported receiving speaker bureau fees from AstraZeneca, Novartis, Pfizer, Novartis, Libs. Author IAPM is employed as an advisory board by BMS, Bayer, Astellas, Janssen, MSD, ROCHE, AstraZeneca, participates in clinical studies sponsored by the company BMS, Astellas, Janssen; and reported receiving speaker bureau fees from Novartis, Janssen, ROCHE, Astellas, MSD, BMS, IPSEN, AMGEM, AstraZeneca. Author GdSF is employed as an advisory board by BMS, MSD, ROCHE, Astellas, Boeringher, Bayer, and reported receiving speaker bureau fees from BMS, MSD, ROCHE. DAS participates in clinical studies.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.963910/full#supplementary-material

References

1. Lindor NM, McMaster ML, Lindor CJ, Greene MH, National Cancer Institute DConcise handbook of familial cancer susceptibility syndromes - second edition. *J Natl Cancer Institute. Monogr* (2008) 38:1–93. doi: 10.1093/jncimonographs/lgn001

2. Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment. *CA: A Cancer J Clin* (2011) 61(5):327–59. doi: 10.3322/caac.20128

3. King MC, Rowell S, Love SM. Inherited breast and ovarian cancer. what are the risks? what are the choices? *JAMA* (1993) 269(15):1975–80. doi: 10.1001/jama.1993.03500150087033

4. Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, et al. A cancer family syndrome in twenty-four kindreds. *Cancer Res* (1988) 48 (18):5358–62.

5. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of lynch syndrome: 1895-2015. *Nat Rev Cancer* (2015) 15(3):181–94. doi: 10.1038/nrc3878

6. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* (2014) 505(7483):302–8. doi: 10.1038/nature12981

7. Desai N. v., Yadav S, Batalini F, Couch FJ, Tung NM. Germline genetic testing in breast cancer: Rationale for the testing of all women diagnosed by the age of 60 years and for risk-based testing of those older than 60 years. *Cancer* (2021) 127 (6):828–33. doi: 10.1002/cncr.33305

8. Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, et al. Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. *JAMA* (2017) 318(9):825. doi: 10.1001/jama.2017.11137

9. Samadder NJ, Riegert-Johnson D, Boardman L, Rhodes D, Wick M, Okuno S, et al. Comparison of universal genetic testing vs guideline-directed targeted testing for patients with hereditary cancer syndrome. *JAMA Oncol* (2020) 7(2):312. doi: 10.1001/jamaoncol.2020.6252

10. Yip CH, Evans DG, Agarwal G, Buccimazza I, Kwong A, Morant R, et al. Global disparities in breast cancer genetics testing, counselling and management. *World J Surg* (2019) 43(5):1264–70. doi: 10.1007/s00268-018-04897-6

11. Achatz MI, Caleffi M, Guindalini R, Marques RM, Nogueira-Rodrigues A, Ashton-Prolla P. Recommendations for advancing the diagnosis and management of hereditary breast and ovarian cancer in Brazil. *J Global Oncol* (2020) 6:439–52. doi: 10.1200/JGO.19.00170

12. Ashton-Prolla P, Seuanez HN. The Brazilian hereditary cancer network: historical aspects and challenges for clinical cancer genetics in the public health care system in Brazil. *Genet Mol Biol* (2016) 39(2):163–5. doi: 10.1590/1678-4685-GMB-2014-0373

13. Weitzel JN, Lagos VI, Cullinane CA, Gambol PJ, Culver JO, Blazer KR, et al. *Limited family structure and BRCA gene mutation status in single cases of breast cancer* . Available at: http://jama.jamanetwork.com/.

14. Daly MB, Pal T, Berry MP, Buys SS, Dickson P, et al. Genetic/Familial High-Risk Assessment. In: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2020)

15. Gupta S, Weiss JM, Burke CA, Chen L-M, Chung DC, Clayback KM, et al. *NCCN guidelines version 1.2021 Genetic/Familial high-risk assessment: Colorectal* (2021). Available at: https://www.nccn.org/home/.

16. Gullo I, Figueiredo J, Seruca R, Carneiro P, Paredes J, Ribeiro AS, et al. Hereditary diffuse gastric cancer: Updated clinical practice guidelines. *Policy Rev Lancet Oncol* (2020) 21:e386–e397. doi: 10.1016/S1470-2045(20)30219-9

17. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing 2016 update (ACMG SF v2.0): A policy statement of the American college of medical genetics and genomics. *Genet Medicine : Off J Am Coll Med Genet* (2017) 19(2):249–55. doi: 10.1038/gim.2016.190

18. Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, et al. Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. *JAMA Oncol* (2015) 1(7):943–51. doi: 10.1001/jamaoncol.2015.2690

19. Stoffel EM, Mangu PB, Gruber SB, Hamilton SR, Kalady MF, et al. Hereditary colorectal cancer syndromes: American Society of clinical oncology clinical practice guideline endorsement of the familial risk-colorectal cancer: European society for medical oncology clinical practice guidelines. *J Clin Oncol* (2015) 33(2):209–17. doi: 10.1200/JCO.2014.58.1322

20. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RWAmerican College of Gastroenterology. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol* (2015) 110(2):223–62 quiz 263. doi: 10.1038/ajg.2014.435

21. Schienda J, Stopfer J. Cancer genetic counseling-current practice and future challenges. *Cold Spring Harb Perspect Med* (2020) 10(6):a036541. doi: 10.1101/cshperspect.a036541

22. Konstantinopoulos PA, Norquist B, Lacchetti C, Armstrong D, Grisham RN, et al. Germline and somatic tumor testing in epithelial ovarian cancer: ASCO guideline. *J Clin Oncol* (2020) 38(11):1222–1245. doi: 10.1200/JCO.19.02960

23. Daly MB, Pal T, Buys SS, Dickson P, Domchek S, et al. Genetic/Familial High-Risk Assessment. In: Breast, Ovarian, and Pancreatic, Version 1.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2021). Available at: https://www.nccn.org/home/member-.

24. Kemp Z, Turnbull A, Yost S, Seal S, Mahamdallie S, Poyastro-Pearson E, et al. Evaluation of cancer-based criteria for use in mainstream BRCA1 and BRCA2 genetic testing in patients with breast cancer. *JAMA Network Open* (2019) 2(5): e194428. doi: 10.1001/jamanetworkopen.2019.4428

25. LaDuca H, Polley EC, Yussuf A, Hoang L, Gutierrez S, Hart SN, et al. A clinical guide to hereditary cancer panel testing: Evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genet Med* (2020) 22(2):407–15. doi: 10.1038/s41436-019-0633-8

26. Muller C, Nielsen SM, Hatchell KE, Yang S, Michalski ST, Hamlington B, et al. Underdiagnosis of hereditary colorectal cancers among Medicare patients: Genetic testing criteria for lynch syndrome miss the mark. *JCO Precis Oncol* (2021) 5:1103–1111. doi: 10.1200/PO.21.00132

27. Palmero EI, Kalakun L, Schüler-Faccini L, Giugliani R, Regla Vargas F, Rocha JCC, et al. Cancer genetic counseling in public health care hospitals: the experience of three Brazilian services. *Community Genet* (2007) 10(2):110–9. doi: 10.1159/000099089

28. Cybulski C, Nazarali S, Narod SA. Multiple primary cancers as a guide to heritability. *Int J Cancer* (2014) 135(8):1756–63. doi: 10.1002/ijc.28988

29. Frank TS, Deffenbaugh AM, Hulick M, Gumpper K. Hereditary susceptibility to breast cancer: Significance of age of onset in family history and contribution of BRCA1 and BRCA2. *Dis Markers* (1999) 15(1–3):89–92. doi: 10.1155/1999/291023

30. Saam J, Moyes K, Landon M, Williams K, Kaldate RR, Arnell C, et al. Hereditary cancer-associated mutations in women diagnosed with two primary cancers: An opportunity to identify hereditary cancer syndromes after the first cancer diagnosis. *Oncology* (2015) 88(4):226–33. doi: 10.1159/000368836

31. Beitsch PD, Whitworth PW, Hughes K, Patel R, Rosen B, Compagnoni G, et al. Underdiagnosis of hereditary breast cancer: Are genetic testing guidelines a tool or an obstacle? *J Clin Oncol* (2019) 37(6):453–60. doi: 10.1200/JCO.18.01631

32. Hampel H, Pearlman R, Beightol M, Zhao W, Jones D, Frankel WL, et al. Assessment of tumor sequencing as a replacement for lynch syndrome screening and current molecular tests for patients with colorectal cancer. *JAMA Oncol* (2018) 4(6):806–13. doi: 10.1001/jamaoncol.2018.0104

33. Yang S, Axilbund JE, O'Leary E, Michalski ST, Evans R, Lincoln SE, et al. Underdiagnosis of hereditary breast and ovarian cancer in Medicare patients: Genetic testing criteria miss the mark. *Ann Surg Oncol* (2018) 25(10):2925–31. doi: 10.1245/s10434-018-6621-4

34. Plichta JK, Sebastian ML, Smith LA, Menendez CS, Johnson AT, et al. Germline Genetic testing: What the breast surgeon needs to know. *Ann Surg Oncol* (2019) 26(7):2184–2190. doi: 10.1245/s10434-019-07341-8

35. Asphaug L, Melberg HO. The cost-effectiveness of multigene panel testing for hereditary breast and ovarian cancer in Norway. *MDM Policy Pract* (2019) 4 (1):238146831882110. doi: 10.1177/2381468318821103

36. Manchanda R, Sun L, Brentnall A, Patel S, Buist DSM, Bowles EJA, et al. A cost-effectiveness analysis of multigene testing for all patients with breast cancer. *JAMA Oncol* (2019) 5(12):1718–30. doi: 10.1001/jamaoncol.2019.3323

37. Caffrey RG. Advocating for equitable management of hereditary cancer syndromes. *J Genet Couns* (2022) 31(3):584–9. doi: 10.1002/jgc4.1548

38. Ladabaum U, Wang G, Terdiman J, Blanco A, Kuppermann M, Boland CR, et al. Strategies to identify the lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. *Ann Internal Med* (2011) 155(2):69–79. doi: 10.7326/0003-4819-155-2-201107190-00002

39. Singh J, Thota N, Singh S, Padhi S, Mohan P, Deshwal S, et al. Screening of over 1000 Indian patients with breast and/or ovarian cancer with a multi-gene panel: prevalence of BRCA1/2 and non-BRCA mutations. *Breast Cancer Res Treat* (2018) 170(1):189–96. doi: 10.1007/s10549-018-4726-x

40. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A population-based study of genes previously implicated in breast cancer. *New Engl J Med* (2021) 384(5):440–51. doi: 10.1056/NEJMoa2005936

41. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol* (2017) 3(4):464–71. doi: 10.1001/jamaoncol.2016.5194

42. Jiang W, Li L, Ke C-F, Wang W, Xiao B-Y, Kong L-H, et al. Universal germline testing among patients with colorectal cancer: Clinical actionability and optimised panel. *J Med Genet* (2022) 59(4):370–6. doi: 10.1136/jmedgenet-2020-107230

43. Tsaousis GN, Papadopoulou E, Apessos A, Agiannitopoulos K, Pepe G, Kampouri S, et al. Analysis of hereditary cancer syndromes by using a panel of genes: novel and multiple pathogenic mutations. *BMC Cancer* (2019) 19(1):535. doi: 10.1186/s12885-019-5756-4

44. Guindalini RSC, Viana DV, Kitajima JPFW, Rocha VM, López RVM, Zheng Y, et al. Detection of germline variants in Brazilian breast cancer patients using multigene panel testing. *Sci Rep* (2022) 12(1):4190. doi: 10.1038/s41598-022-07383-1

45. Custódio G, Parise GA, Kiesel Filho N, Komechen H, Sabbaga CC, Rosati R, et al. Impact of neonatal screening and surveillance for the TP53 R337H mutation on early detection of childhood adrenocortical tumors. *J Clin Oncology : Off J Am Soc Clin Oncol* (2013) 31(20):2619–26. doi: 10.1200/JCO.2012.46.3711

46. Felix GES, Abe-Sandes C, Machado-Lopes TMB, Bomfim TF, Guindalini RSC, Santos VCS, et al. Germline mutations in BRCA1, BRCA2, CHEK2 and TP53 in patients at high-risk for HBOC: Characterizing a northeast Brazilian population. *Hum Genome Variation* (2014) 1:1–8. doi: 10.1038/hgv.2014.12

47. Volc SM, Ramos CRN, Galvão HCR, Felicio PS, Coelho AS, Berardineli GN, et al. The Brazilian TP53 mutation (R337H) and sarcomas. *PLoS One* (2020) 15(1): e0227260. doi: 10.1371/journal.pone.0227260

48. Robson ME, Storm CD, Weitzel J, Wollins DS, Offit KAmerican Society of Clinical Oncology. American Society of clinical oncology policy statement update: genetic and genomic testing for cancer susceptibility. *J Clin Oncology : Off J Am Soc Clin Oncol* (2010) 28(5):893–901. doi: 10.1200/JCO.2009.27.0660

49. Frey MK, Finch A, Kulkarni A, Akbari MR, Chapman-Davis E. Genetic testing for all: Overcoming disparities in ovarian cancer genetic testing. *Am Soc Clin Oncol Educ Book* (2022) 42:471–82. doi: 10.1200/EDBK_350292

50. Samadder NJ, Riegert-Johnson D, Boardman L, Rhodes D, Wick M, Okuno S, et al. Comparison of universal genetic testing vs guideline-directed targeted testing for patients with hereditary cancer syndrome. *JAMA Oncol* (2021) 7(2):230. doi: 10.1001/jamaoncol.2020.6252

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fonc.2022.910494

OPEN ACCESS

EDITED BY
Manoj Kumar Kashyap,
Amity University Gurgaon, India

REVIEWED BY
Thommas Mutemi Musyoka,
Kenyatta University, Kenya
Aneesha Acharya,
Dr. D. Y. Patil Dental College & Hospital, India

*CORRESPONDENCE
Anshuman Dixit
anshuman@ils.res.in

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

RECEIVED 05 April 2022
ACCEPTED 15 August 2022
PUBLISHED 20 September 2022

CITATION
Kumari P, Kumar S, Sethy M, Bhue S, Mohanta BK and Dixit A (2022) Identification of therapeutically potential targets and their ligands for the treatment of OSCC. *Front. Oncol.* 12:910494. doi: 10.3389/fonc.2022.910494

Identification of therapeutically potential targets and their ligands for the treatment of OSCC

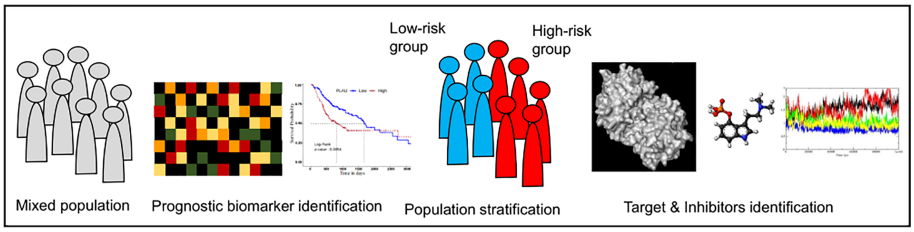
Pratima Kumari[1,2], Sugandh Kumar[1], Madhusmita Sethy[1], Shyamlal Bhue[1,2], Bineet Kumar Mohanta[1,2] and Anshuman Dixit[1]*

[1]Computational Biology and Bioinformatics Laboratory, Institute of Life Sciences, Bhubaneswar, India, [2]Regional Centre for Biotechnology (RCB), Faridabad, India

Recent advancements in cancer biology have revealed molecular changes associated with carcinogenesis and chemotherapeutic exposure. The available information is being gainfully utilized to develop therapies targeting specific molecules involved in cancer cell growth, survival, and chemoresistance. Targeted therapies have dramatically increased overall survival (OS) in many cancers. Therefore, developing such targeted therapies against oral squamous cell carcinoma (OSCC) is anticipated to have significant clinical implications. In the current work, we have identified drug-specific sensitivity-related prognostic biomarkers (*BOP1*, *CCNA2*, *CKS2*, *PLAU*, and *SERPINE1*) using gene expression, Cox proportional hazards regression, and machine learning in OSCC. Dysregulation of these markers is significantly associated with OS in many cancers. Their elevated expression is related to cellular proliferation and aggressive malignancy in various cancers. Mechanistically, inhibition of these biomarkers should significantly reduce cellular proliferation and metastasis in OSCC and should result in better OS. It is pertinent to note that no effective small-molecule candidate has been identified against these biomarkers to date. Therefore, a comprehensive *in silico* drug design strategy assimilating homology modeling, extensive molecular dynamics (MD) simulation, and ensemble molecular docking has been applied to identify potential compounds against identified targets, and potential molecules have been identified. We hope that this study will help in deciphering potential genes having roles in chemoresistance and a significant impact on OS. It will also result in the identification of new targeted therapeutics against OSCC.

KEYWORDS

chemotherapy, chemoresistance, prognosis, drug discovery, OSCC



GRAPHICAL ABSTRACT

Introduction

Oral squamous cell carcinoma (OSCC) constitutes a major subset of head and neck squamous cell carcinoma (HNSCC) and accounted for an estimated 0.37 million cases in the year 2020 (1). The high morbidity and mortality of OSCC pose a great challenge to its management (1, 2). The overall 5-year survival rate in OSCC is comparatively lower (~50%) than that in many other cancers (3, 4). For example, in India, 0.13 million new cases were detected, whereas 75,000 patients died in 2020 (1). Current treatment modalities include surgery, radiotherapy, chemotherapy, or their combinations. These are successful in patients with primary tumors, whereas patients with high-risk features (invasion/perineural invasion, metastasis, T3/T4 stage, or involvement of two or more lymph nodes) show less improvement. There is no clinical evidence to support the likely outcome in the case of high-risk oral cancer (5, 6).

Patient response to chemotherapy has been linked to tumor lineage and genetics. Changes in cellular gene expression in response to small-molecule treatments can provide insights into cellular processes governing the clinical outcome. Identifying the responsive features and mechanism of action is of immense value in cancer therapy and can be critical for the development of novel medicines. Furthermore, the heterogeneous response of patients to cancer therapies and the frequent development of drug resistance highlight the importance of a therapeutic response (7, 8). However, an accurate prediction of a therapeutic response and the identification of new anticancer drugs have remained a challenging task.

In recent years, the increasing understanding of genomics and the advent of next-generation sequencing (NGS) with advancements in bioinformatics approaches have made it possible to identify potential molecular targets for the betterment of chemotherapy. Numerous cancer studies with the help of NGS were able to identify novel and rare somatic mutations in more efficient and accurate ways. In a variety of cancers such as bladder cancer, renal cell carcinoma, small-cell lung cancer, prostate cancer, acute myelogenous leukemia, and chronic lymphocytic leukemia, researchers were able to accurately identify genetic alterations. On the other hand, together with NGS, the bioinformatics approach was successful in exploiting the heterogeneous nature of cancer to develop cancer diagnostic, prognostic, and predictive markers (9). Gene-based approaches are being used for the development of new therapeutic agents. The targeted approaches have been proven highly useful in the development of therapies for many cancers (10).

Doxorubicin is known to inhibit the topoisomerase-II (TOP2) activity in eliciting its antineoplastic effect. It is one of the most effective anticancer drugs widely used in the treatment of several cancers including breast cancer, lung cancer, gastric cancer, ovarian cancer, thyroid cancer, non-Hodgkin's and Hodgkin's lymphoma, multiple myeloma, sarcoma, and pediatric cancer (11). High toxicity and early resistant phenotype have limited the use of doxorubicin (12). Thus, it is imperative to study the molecular changes associated with doxorubicin resistance. Therefore, in the current work, we aimed to explore gene expression changes in the response to

Abbreviations: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; COAD, colon adenocarcinoma; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSCC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; AML, acute myeloid leukemia; LGG, brain lower-grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumor; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; MESO, mesothelioma; UVM, uveal melanoma; GEM, gemcitabine; GC, gastric cancer; OC, ovarian cancer, TMZ, temozolomide; NSCLC, non-small cell lung carcinoma.

doxorubicin treatment in OSCC using various datasets and the ways to reduce the chances of emergence of such resistance by molecularly targeted therapies. We have also identified potential ligands that may increase the effect of doxorubicin and/or can delay the progression of drug-related resistant phenotypes or sensitize oral cancer cells to chemotherapy.

Materials and methods

Data collection and differential gene expression analysis

In the present study, data were collected from two sources: The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) and Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). Since we wanted to analyze OSCC data, we further filtered out the OSCC samples according to the International Classification of Diseases (ICD) code. The ICD classifies diseases based on the site of disease occurrence (Supplementary Table S1). Finally, 319 OSCC and 44 normal adjacent tissue (NAT) samples were obtained. Additionally, two more datasets were obtained from GEO (13, 14): 1) mRNA expression profiling data of 27 OSCC patients (GEO ID-GSE23558) and 2) mRNA expression profiling data of the doxorubicin-treated SCC25 cell line (GSE58074). Two replicates of each (mock and treatment) were taken for expression analysis. The clinical details of TCGA patients are given in Supplementary Table S2.

The gene expression analysis was performed using R studio (http://www.rstudio.com/) version 3.4.4 using the limma-voom library for TCGA samples, whereas GEO samples were analyzed using GEO2R. Genes with |log2FC ≥1| and p-value<0.05 were considered significantly differentially expressed genes (DEGs). The ggplot2, complex heatmap, and circular library in R were used for volcano plot and heatmap. Common DEGs in all of the three datasets were considered genes of interest for further analysis.

Functional enrichment analysis

Gene Ontology (GO) and pathway enrichment of common genes present in all of the datasets were used to identify enriched biological events as a result of doxorubicin perturbation. We have used Reactome (https://reactome.org) (15, 16) and GO (http://geneontology.org/) (17, 18) online databases for enrichment analysis of the gene sets. Enriched biological pathways and GO terms were considered significant if p-value ≤0.05.

Cox proportional hazards regression

Cox proportional hazards regression (Coxph) is a semiparametric model used to predict the outcome of disease based on one or more predictors in survival time (time-to-event) through the hazard ratio (HR) function. It assumes that the effects of predictor variables have an additive effect on the hazard, i.e.,

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_n x_n) \quad (1)$$

where $h(t)$ is the hazard at time t for a subject with a set of predictors $x_1 \ldots x_n$, $h_0(t)$ is the baseline hazard function, and $\beta_1,.,\beta_n$ are the model coefficients describing the effect of the predictors on the overall hazard. HR is used for the interpretation of the Cox model. The HR examination shows selective factors that influence the rate of an event happening (e.g., death) at a particular point in time. An HR above 1 indicates a predictor that is positively associated with the event probability (death) and thus negatively associated with the length of survival, indicating worse prognosis; a negative HR indicates a protective effect of the predictors with which it is associated; while an HR equal to 1 means no effect. The DEGs were subjected to a Cox regression analysis using the "survival" package in R (https://github.com/therneau/survival).

Identification of biomarker signature and validation of prognosis-related genes

Machine learning (ML) is becoming popular in cancer biology in identifying prognostic, diagnostic, and therapeutic biomarkers. Therefore, we implemented two commonly used machine learning algorithms, *viz.*, random forest (RF) and partial least square (PLS) regression method, to identify prognostic targets. RF is a frequently used algorithm for the identification of prognostic biomarkers in many diseases including cancer (19, 20).

Survival analysis

To evaluate the reliability of the predicted prognosis signature, Kaplan–Meier (K-M) plots were generated to estimate the survival of the patients based on their median mRNA expression. An mRNA expression above the median value was considered high, whereas an expression below the median was counted as low expression. The R package "survival" was used to plot the patient's overall survival (OS), and significance was calculated based on the log-rank p-value. To further reflect the sensitivity and specificity of signature mRNAs, we employed a time-dependent receiver

operating characteristic (ROC) curve analysis *via* the R package "survivalROC."

Risk assessment and validation of the risk model

The risk score is an additive model of the mRNA expression level multiplied by their Cox regression coefficient. The risk score was calculated as follows:

$$\begin{aligned}\text{Risk score} &= \text{Coefficients of signature Gene A} \times \text{mRNA expression of A} + \ldots\ldots\ldots. \\ &+ \text{Coefficients of signature Gene N} \times \text{mRNA expression of N} \quad (2)\end{aligned}$$

The median risk score was used to divide the patients into high- and low-risk groups.

Use of the protein–protein interaction network to understand the potential signature biomarker

Next, we constructed the protein–protein interaction (PPI) network to understand the importance of key targets in the human interactome. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) was used to create a PPI network of identified genes with a cutoff score of >0.4, which was equivalent to medium confidence. Subsequently, a cluster analysis using MCODE (21) was done to decipher the modules in the created network. To understand the functional significance of the modules, pathway analysis was done using Reactome (www.reactome.org).

Identification of candidate small-molecule drugs

Protein structures

The X-ray crystal structures of SERPINE1 [Protein Data Bank (PDB): 4AQH], CCNA2 (PDB: 1H1R], PLAU (PDB: 1OWE), and CKS2 (PDB: 5HQ0) were used in the current studies. The three-dimensional (3D) structure of BOP1 is not available in the PDB.

Homology modeling

The 3D structure of BOP1 was modeled by a homology modeling approach using modeler v9.20. Delta-blast was used to identify suitable templates for modeling against the PDB database. The template was selected based on the expected value (E-value), bit score, and percentage of query coverage and identity. The carboxy-terminal domain of Erb1 of *Escherichia coli* (PDB-ID: 4U7A chain A, identity 40.1%, coverage 83%) was finally selected as the template to model the structure of BOP1. Twenty models were generated for each protein, and the final model was selected based on the lowest molpdf score.

System setup and molecular dynamics simulation

In the current study, the preliminary topology and coordinates for all proteins were generated in VMD v1.93, whereas the simulations were run in NAMD v2.14. All protein structures (BOP1, PLAU, SERPINE1, CKS2, and CCNA2 after stripping cocrystallized ligands) were prepared and solvated in a rectangular water box (TIP3P water model) with a buffering distance of 10 Å. Ions were added to ensure the electroneutrality of the solvated system. The SETTLE algorithm was used for the water molecules. The associated system topology and coordinates were generated by applying charmm34 force field parameters for molecular dynamics (MD) simulation. Prior to the simulation, the system was properly minimized with a stepwise minimization protocol. Firstly, the water molecules and ions were minimized that was then followed by hydrogen atoms and the side chains of the complex. The side chains were minimized for 40,000 steps, whereas the backbone atoms and the bond lengths of hydrogen atoms were kept fixed. Thereafter, all of the atoms were allowed to relax freely, and the whole system was energy-minimized for 40,000 steps with nominal restraints on Cα atoms (10 kcal/mol) to prevent any abrupt change in the structure. Subsequently, an equilibration protocol was followed where the system was heated gradually from 0 to 310 K in steps of 30 K with a canonical ensemble [constant volume, constant temperature (NVT)]. At each step, a 20-picosecond (ps) simulation was run to allow the system to adjust to the temperature. Once the system attained 310 K, an isobaric and isothermic ensemble [constant pressure, constant temperature (NPT)] was applied for a period of 100 ps with a constant pressure of 1.0 bar using Langevin dynamics. Finally, the applied restraints on Cα atoms were removed, and the system was equilibrated for 1 ns at 310 K using the Langevin piston coupling algorithm. During the whole simulation, the Particle Mesh Ewald (PME) algorithm was used to calculate the long-range electrostatic interactions with fixed periodic boundary conditions. The covalent bond interactions involving hydrogen were constrained using the SHAKE algorithm. Once the system was simulated with a constant 310 K temperature and 1.0 bar pressure, then the production run was done for a time period of 100 ns. The analyses of the MD trajectories were performed to analyze the structure and dynamic behavior of all proteins during MD. The trajectories were analyzed for root mean

square deviation (RMSD) and radius of gyration (Rg); these analyses were performed using VMD and in-house Perl and Tcl scripts. Five additional equidistant frames were generated from each trajectory for ensemble docking studies.

To study the stability of the protein–ligand complex, another simulation was run for 300 ns with selected ligands docked in the protein targets. The ligand parameters were generated using the Antechamber module of AMBER12 molecular simulation package (www.ambermd.org). The complex was solvated in a box of water with 10 Å buffering distance. AMBER parameters and coordinate files were generated using the tleap module of AMBER12. Equilibration and simulation were done using NAMD v2.14 as described above. The trajectories were analyzed for RMSD, hydrogen bonds, and salt bridges to assess the stability and strength of the interactions between the ligand and the protein.

Molecular docking

The frames obtained from the MD simulation for the selected proteins were prepared using the protein preparation wizard module of Schrödinger molecular modeling software v9.3. OPLS-2005 force field was used for energy minimization, whereas Schrödinger's LigPrep module was employed to generate 3D conformers and energy minimization of the potential small-molecule drugs from the US Food and Drug Administration (US FDA) library. A maximum of 32 conformers were generated per ligand. Before docking, the grid box (active site) was defined as residues within 5 Å of cocrystallized ligands. The active site of the modeled proteins or proteins without cocrystallized ligands was predicted through the sitemap algorithm as implied in Schrödinger. Docking was performed using standard precision (SP) mode with flexible ligand sampling in Schrödinger's Glide module. The average docking score was calculated based on the glide score in the five frames. The flowchart of methodology is given in Figure 1.

Results

Identification of differentially expressed genes

We identified 3,976, 5,418, and 1,241 DEGs in TCGA, GSE23558, and GSE58074 datasets, respectively. In TCGA samples, 2,163 genes were overexpressed and 1,813 genes were underexpressed, whereas 2,221 and 3,197 genes were found overexpressed and underexpressed, respectively, in GSE23558. In GSE58074, which contained doxorubicin-treated cell line data, 724 genes were overexpressed and 517 genes were underexpressed (Figure 2A and Supplementary Table S3). Among the three datasets, 168 common DEGs were identified. These were considered for further analysis (Figure 2B). DEGs are represented in a volcano plot with $\log_2$FC and -$\log_{10}$ p-value (Supplementary Figure S1). The heatmap represents the expression of common DEGs from the three datasets, with 134, 132, and 85 overexpressed genes and 34, 36, and 83 underexpressed genes in TCGA, GSE23558, and GSE58074,

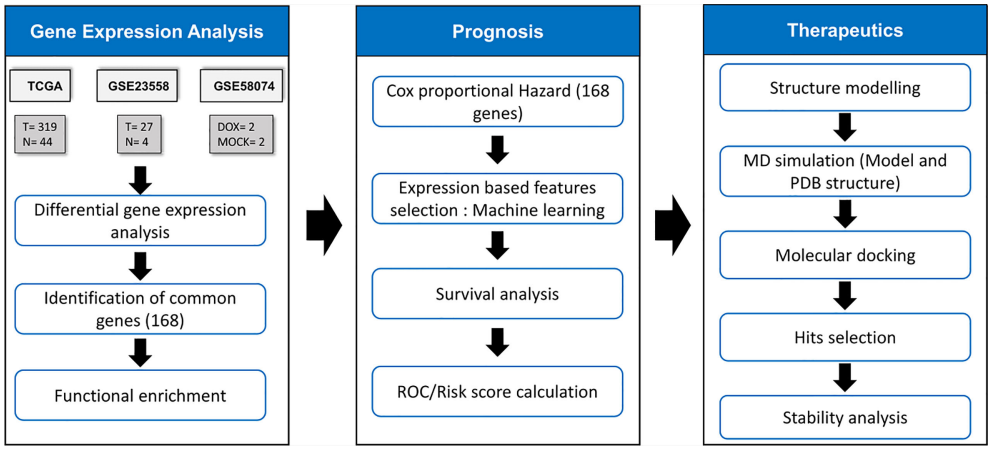


FIGURE 1
Methodology for identification of drug response related signature and identification of their inhibitors. The TCGA samples had 319 OSCC & 44 normal samples. GSE23558 had 27 OSCC patients' data. GSE58074 examined the effect of doxorubicin on SCC25 cell lines to check for molecular markers underlying doxorubicin response..

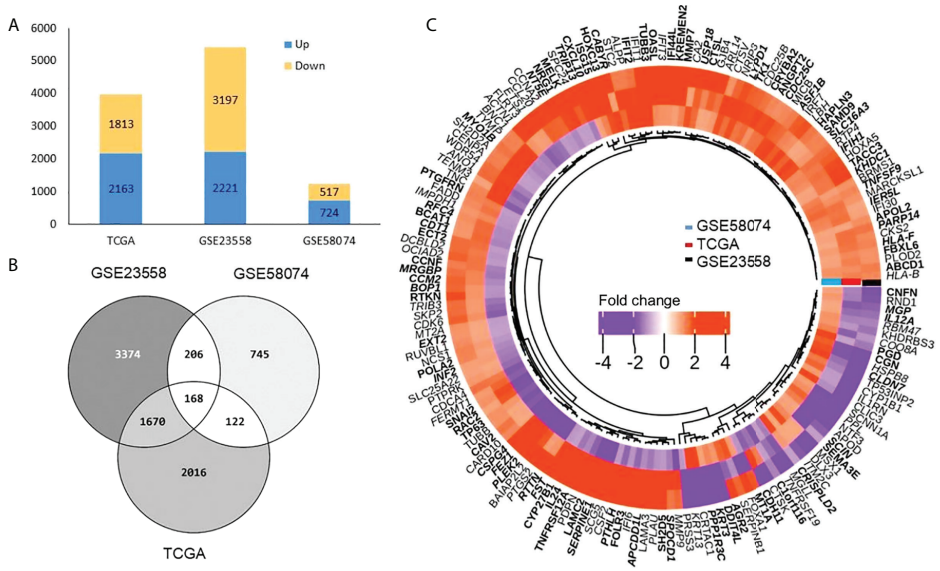


FIGURE 2
Differentially expressed genes **(A)** Bar graph representing total DEGs both up (blue) and down (yellow) in three datasets. **(B)** Venn diagram showing number of DEGs (common and unique) among TCGA, GSE23558 and GSE58074. A total of 168 common DEGs were obtained. **(C)** Heatmap shows DEGs related to doxorubicin response (GSE58074) and non-treated samples in OSCC from TCGA and GSE23558. Three types of distinct expression pattern are discernible (1) 52 genes overexpressed (orange) in all the three data sets, (2) 76 genes underexpressed in GSE58074 (purple) while overexpressed in TCGA (orange) and GSE23558 (orange), and (3) 26 genes underexpressed in TCGA (purple) and GSE23558 (purple) while overexpressed in GSE58074 (orange).

respectively (Figure 2C and Supplementary Table S3). From the heatmap, it is clearly visible that genes that are overexpressed in cancer prior to any treatment get downregulated after treatment and *vice versa*, whereas some of the genes present do not respond to the treatment, as there are no changes in their expression patterns.

Functional enrichment analysis

The common DEGs (n = 168) were studied for functional enrichment to investigate their involvement in various cellular processes. The top 10 enriched GO functions and pathways are shown in Figure 3. Biological process (BP) terms related to immune response, biological stress, cell proliferation, and survival were found to be significantly enriched. Enriched cellular component (CC) terms include extracellular related components, protein complex, membrane-bound components, and cellular vesicles. Molecular function (MF) enriched terms included various signaling pathways involved in ligand–receptor complex, enzymatic functions, G protein–related function, kinases, and other protein-binding functions. Pathway enrichment shows that most of the signaling related to cell cycle and its regulation such as G2/M transition, nucleotide salvage pathways, and DNA damage bypass. The detailed pathway and GO term information is provided in Supplementary Table S4 and Supplementary Figure S2.

Exploration of prognostic biomarkers

The prognostic significance of the selected DEGs (168 genes) was assessed through univariate Coxph regression analysis. A total of 59 genes were found to be significantly (p-value<0.05) associated with OS (Figure 4A and Supplementary Table S3). As indicated earlier, the higher values of HR (HR >1) indicate worse prognosis, whereas negative HR values (<-1) show favorable prognosis. Next, machine learning algorithms, *viz.*, RF and PLS regression, were employed to identify the potential predictive prognostic signature genes. We selected the top predictive genes based on the robustness of the prediction. A cutoff percentile score of ≥80 was used; nine genes were selected from RF, whereas seven genes were predicted by PLS (a total of 16) (Supplementary Table S5). Then, we performed survival

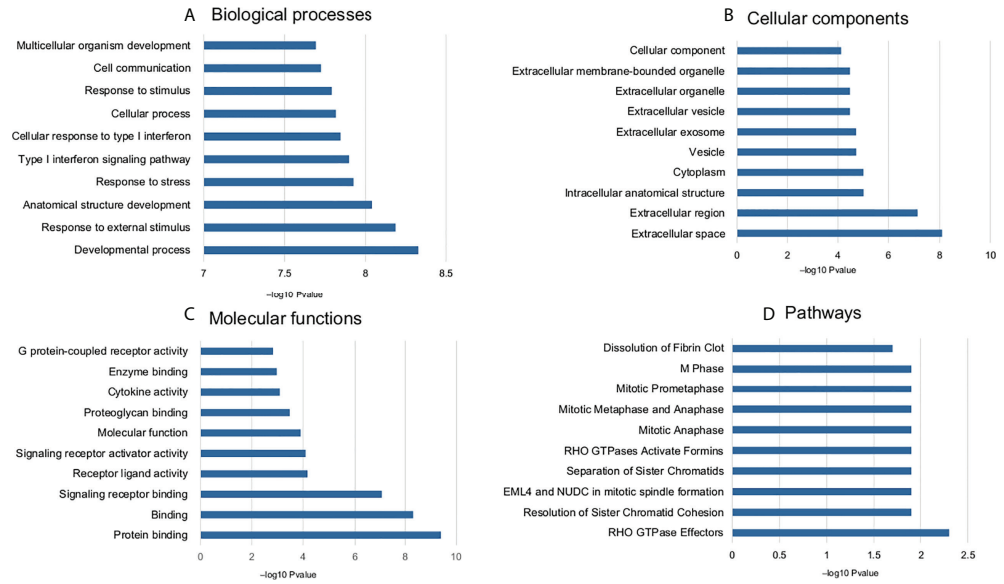


FIGURE 3
Enriched GO terms and pathways (top 10) common DEGs associated with doxorubicin response. **(A)** biological process, **(B)** cellular components, **(C)** molecular function, and **(D)** pathways.

analysis to identify the potential prognostic markers. We found that five genes, *viz.*, *BOP1*, *PLAU*, *SERPINE1*, *CCNA2*, and *CKS2*, were significantly related to the survival of the patients. The risk assessment score was calculated with the help of Cox regression analysis that resulted in the following equation (Eq. 2) and was used to determine the risk score (high and low):

$$\begin{aligned}\text{Risk score} = {} & (3.207785 \times \text{expression}_{BOP1}) \\ & + (1.986074 \times \text{expression}_{PLAU}) \\ & - (1.962924 \times \text{expression}_{SERPINE1}) \\ & + (2.913185 \times \text{expression}_{CCNA2}) \\ & + (4.164371 \times \text{expression}_{CKS2}) \end{aligned} \quad (2)$$

Importantly, low-risk patients have a low expression of these mRNAs as compared with those of high-risk patients (Figure 4B). It can be seen clearly in Figures 4D–H that the low–risk score group had better OS; therefore, the expression of the identified mRNAs has significant impact on the OS of the patients. To further assess the accuracy of the prognostic model, we constructed an ROC curve to assess the impact of the expression of these genes on patients' OS for 1, 3, and 5 years. The ROC curve is shown in Figure 4C, and the area under the curve (AUC) was 0.773, 0.806, and 0.928 for 1, 3, and 5 years, respectively.

Assessment of the Five Prognostic Signature Genes Across Cancers

The Cox regression and Machine learning (ML) analysis showed that overexpression of the identified genes is related to poor prognosis. Furthermore, we investigated the expression of these genes across different cancers and OS of patients to assess their clinical importance. For this, we have used GEPIA2 database (22). GEPIA2 retrieves data from TCGA and the Genotype-Tissue Expression (GTEx) portal. The expression analysis showed that these five genes were significantly upregulated (|log2FC >1|, p-value<0.05) in most of the tumor tissues as compared with those in normal tissues (Figure 5A and Supplementary Table S6). Survival analysis across various cancers showed that a high expression of these five genes is related to poor survival in most of the cancers (Figure 5B). Moreover, we also checked the expression of these genes in response to drug treatment. For this, we analyzed the expression array data available in the NCBI-GEO dataset (Supplementary Table S7). Two types of analysis were done: 1) expression analysis in drug-sensitive cells and 2) expression analysis in drug-resistant cells (Figures 5C, D). We found that the mRNA

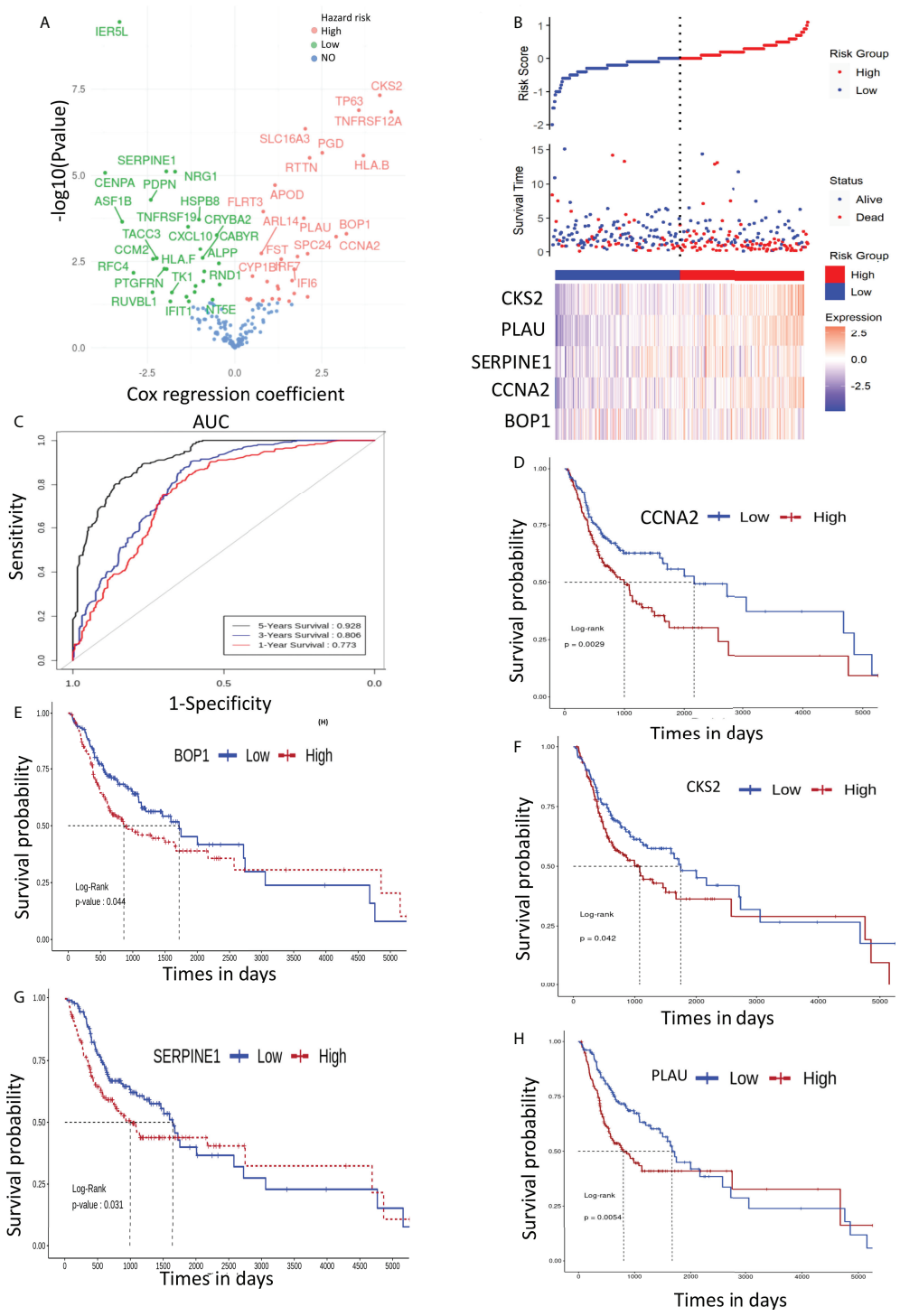


FIGURE 4
(A) Volcano plot of cox regression showing hazard ratio (HR) for 168 genes. **(B)** Patient stratification according to risk score to predict the survival time of patients with high- and low-expression level of prognostic genes. **(C)** ROC depicting the effect of selected genes on overall survival (1, 3, and 5 years). **(D–H)** depict the effect of signature gene expression on overall survival of OSCC patients..

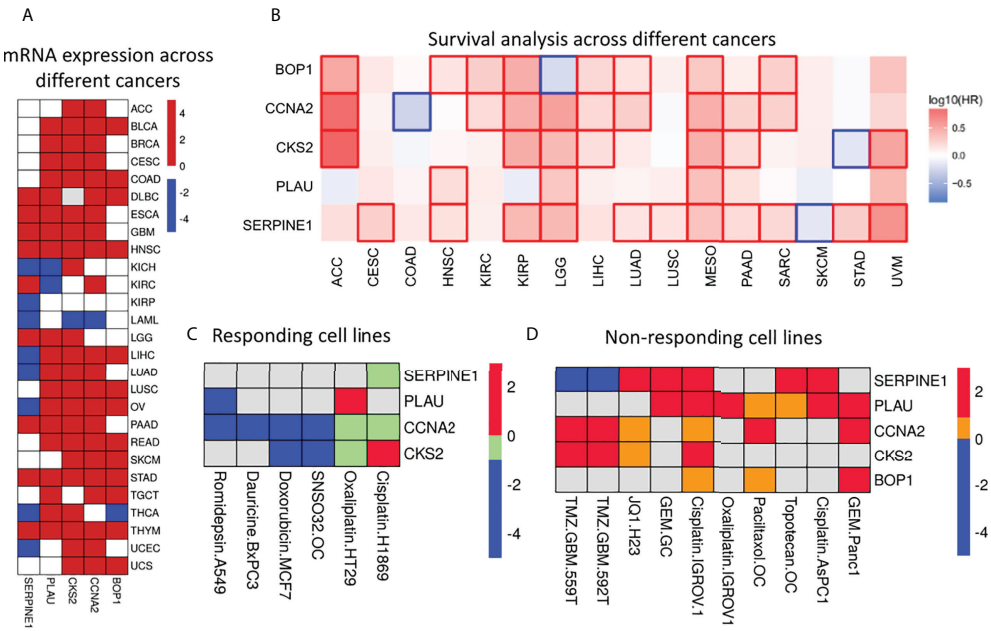


FIGURE 5
Expression & significance of selected prognostic genes. **(A)** Differential gene expression analysis across different cancers. Red, blue and white squares depict over, under and insignificant expression respectively. It is clear that these genes are upregulated in most of the cancers. **(B)** Survival map: Red and blue squares indicate poor survival due to over and under expression respectively. The figure clearly indicates the expression of these genes have significant effect on survival of the patients across cancers. **(C)** Heatmap represents expression level of these genes in drug sensitive cells. These genes were found to be significantly differentially expressed ($|\log_2 Fc| > 1$ and pvalue < 0.05). The down- (blue) and up- (red) regulated genes in response to drug; green square indicates down-regulated genes with significant pvalue but $|\log_2 Fc| < 1$. Grey square indicated no differential expression. **(D)** The expression of selected genes in drug resistant cells. Red and blue squares indicate significant ($|\log_2 Fc| > 1$ and pvalue < 0.05) up- and down-regulated genes; yellow square indicates up-regulated genes with significant pvalue ($|\log_2 Fc| < 1$); grey square indicated no differential expression of the genes.

expression level decreased for most of the genes upon drug treatment. For example, *CCNA2* mRNA levels were found to be decreased significantly with fold change -1.07, -4.33, -1.06, and -4.17 when treated with romidepsin, SNAO32, dauricine, and doxorubicin in A549, primary ovarian cancer, BxPC3, and MCF7 cell lines, respectively. Similarly, *PLAU*, *CKS2*, and *SERPINE1* mRNA levels were also found to be decreased in response to various drugs, whereas in the case of the drug-resistant cell line, an increase in mRNA expression levels of *PLAU*, *CKS2*, *CCNA2*, *BOP1*, and *SERPINE1* was observed. For example, PLAU showed increased mRNA expression with fold changes of 2.14, 1.8, 1.5, 1.4, and 1.2 in Panc1 gemcitabine-, MKN28 gemcitabine-, AsPc1 cisplatin-, IGROV1 cisplatin-, and IGROV1 oxaliplatin-resistant cell lines, respectively. Additionally, we analyzed the presence of genetic alterations (e.g., amplification, mutation, deletion, structural variant) in these five genes using CBioPortal (https://www.cbioportal.org/), which is an online consortium for cancer genomics. We investigated the TCGA Pan-Cancer Atlas dataset for the genomic alteration study (23, 24). It clearly showed that amplification is the most common alteration observed in *BOP1*, *PLAU*, *SERPINE1*, and *CKS2* across the cancers, whereas in *CCNA2*, mutation and deep deletion are prevalent (Supplementary Figure S3).

Protein–protein interaction network

The PPIs are at the heart of various molecular mechanisms. Therefore, a PPI network was constructed to understand the interactions of selected proteins with other proteins in the human interactome for better understanding of their regulatory roles (Figure 6). We identified three distinct clusters in the network; the proteins in those clusters are involved in 1) the regulation of rRNA processing; 2) the regulation of the cell cycle; and 3) angiogenesis, growth factors, and transcription factors such as Suppressor of Mothers against Decapentaplegic (SMAD) SMAD2/SMAD3. The overall network analysis indicates that these genes are connected to many important genes, and targeting them will affect cellular processes playing critical roles in the pathogenesis of OSCC.

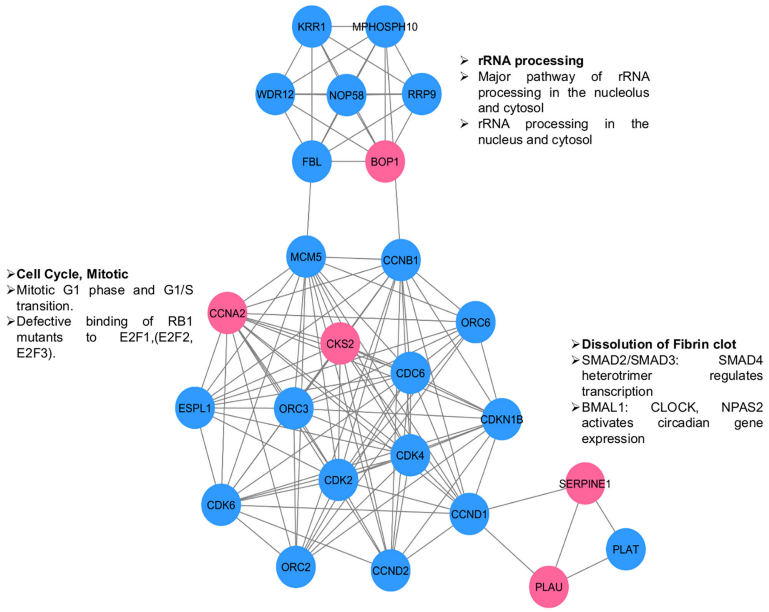


FIGURE 6
PPI network for selected genes (pink circles) to understand their interactions & significance. Other interacting proteins are depicted in blue circles. Three clusters are visible. The enriched functions are shown along the clusters.

Molecular dynamics simulations

The MD analysis was performed to assess the flexibility of the binding site that is not discernible through PDB structures. The generated structural ensembles were used for the identification of small-molecule binders. MD simulations were done for 100 ns on each protein (total 500 ns), and the stability of the simulation was evaluated using RMSD. The RMSD values reveal the structural changes that occurred during the MD. The RMSD plots for all proteins indicated that each system was stabilized quickly and then remained stable throughout the simulation time, as evidenced by the movement of the RMSD curve within 2 Å. These plots suggested that each system was quite stable for the docking study. Five equidistant frames for each protein at 20-ns distance (from 100-ns simulation), representing the dynamics of the protein structure, were extracted. Docking studies were then performed on these frames.

Screening of potential compounds from the US Food and Drug Administration–approved library

The binding affinity of US FDA-approved drugs with each of the proteins was assessed through molecular docking. The average glide docking score (average of five frames) was used to identify potential binders of the individual proteins. The top 20 compounds for each protein, based on their glide score, are given in Supplementary Table S8. Details for individual proteins are given below.

BOP1 **is a** RNA-binding protein involved in ribosome biogenesis, cell cycle, and cell proliferation (25). The docking analysis indicated saquinavir to be the best binder with an average docking score of -10.9 (Figure 7A). The active site of BOP1 is surrounded by multiple beta sheets forming a barrel-like structure. The molecule saquinavir is ensconced in a pocket lined by Trp182, Pro104, Pro229, Pro63, Thr181, Pro368, Gly184, Leu266, Val268, and Val309. Analysis of docking poses indicates that saquinavir has several hydrogen-bonding interactions with residues, *viz.*, Trp182, Val268, and Val309.

CCNA2 (cyclin A2) binds with both CDK2 and CDK1. It is required for entry into the S and M phases of the cell cycle. The overexpression of CCNA2 leads to cell growth and proliferation. Diacetolol was found to be the best binder, with an average docking score of -7.6. It is a beta-blocker used as an antihypertensive and antiarrhythmic agent. The docking study indicated that the ligand diacetolol binds in a cavity lined by residues His233, Gln337, Ser340, Tyr347, Tyr350, Pro352, Val354, Ile355, Ala356, Cys390, Asp393, and Leu394 of CCNA2. It makes HB interaction with Tyr347 (Figure 7B). It can be an attractive chemotherapeutic option for OSCC.

PLAU encodes a serine protease (uPA) that converts plasminogen to plasmin. uPA is involved in the degradation of the basement membrane and extracellular matrix (26). *PLAU* is

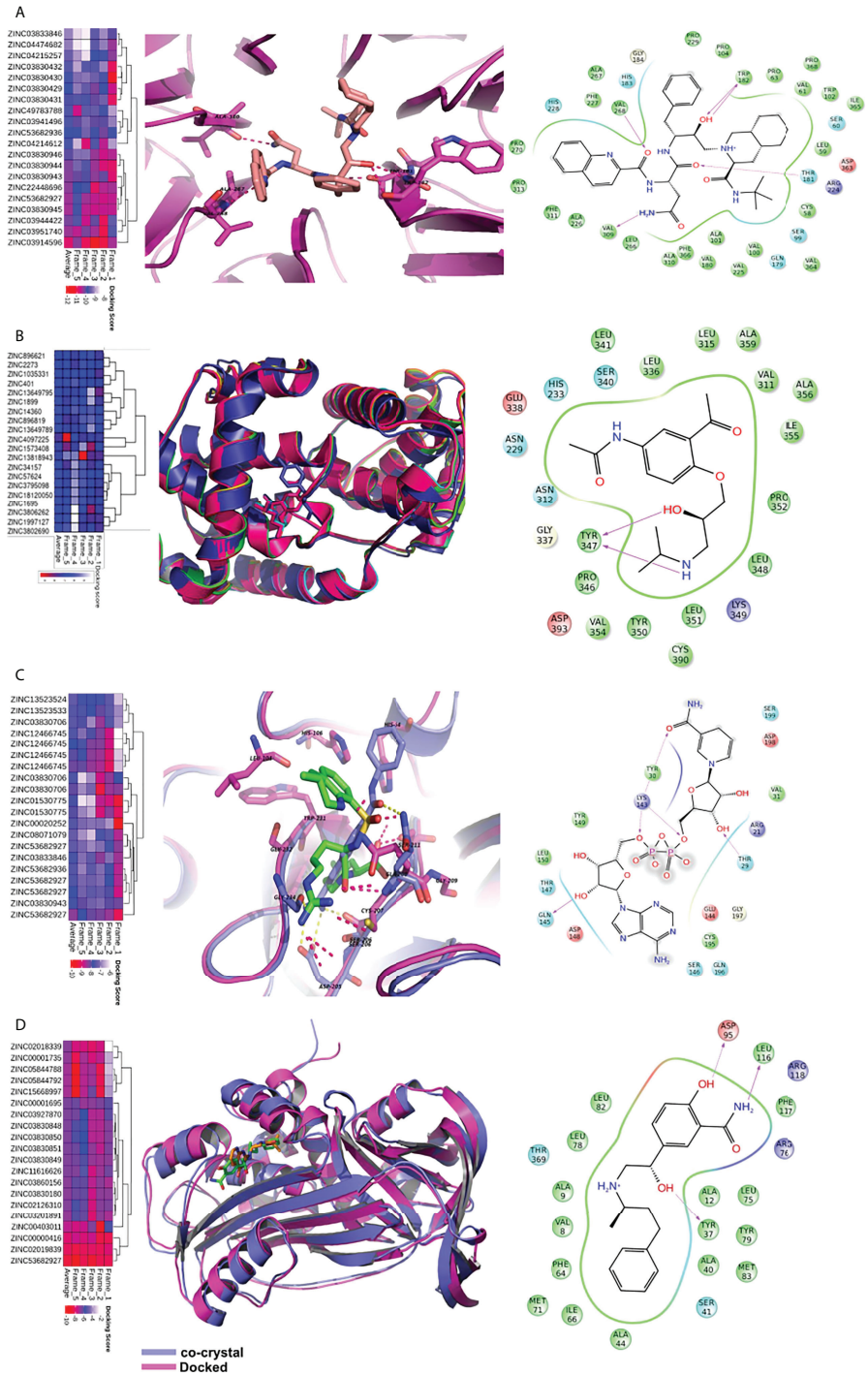


FIGURE 7
Virtual screening of FDA approved drugs against identified proteins. Top scoring 20 molecules for each of the five proteins are shown in heatmap (figures **A–D**). The first column depicts the heatmap of docking scores in five MD frames. The average docking score is also depicted for better understanding. In the second column, proteins are shown in cartoon while the ligands are shown in sticks. The dotted lines depict hydrogen bonds. The third column depicts the interaction map of the ligand receptor interactions. The pink arrows show hydrogen bonds. The arrowhead depicts the HB-acceptor molecule. Pi-pi stacking interactions are depicted by green line. **(A)** The docking of saquinavir with BOP1. It makes hydrogen bonding interactions with backbone of TRP-182, VAL-268, VAL-309, and sidechain of THR-181. **(B)** The binding of molecules in CCNA2. The diacetolol is shown in blue sticks. The ligand makes hydrogen bonds with Tyr347 **(C)** The binding of ligands in PLAU. NADH is shown in binding site of PLAU. It makes hydrogen bonding interactions with THR30, TYR31, LYS154, and SER157. **(D)** The binding of ligands in SERPINE1, and docked molecule labetalol. The drug forms hydrogen bonds with sidechain of ASP-95, TYR37.

one of the potential biomarkers for HNSCC and several other cancers. Nicotinamide adenine dinucleotide hydrogen (NADH), argatroban, diminazene (DIZE), and pentamidine are among the molecules that showed good binding affinity with PLAU. NADH binding to PLAU is shown in Figure 7C. It makes hydrogen-bonding interactions with Thr30, Tyr31, Lys154, and Ser157. NADH, due to its role in energy production, may help against wastage and weakness of cancer patients. It is also used to improve mental alertness. It can be given orally and has cleared clinical trials as a chemotherapeutic agent for other illnesses (27).

SERPINE1 (PAI-I) inhibits the plasminogen activator uPA/uPAR complex that promotes cell matrix degradation and cell migration. Overexpression of *SERPINE1* is highly associated with poor survival in primary tumor, lymph node, and head and neck cancer metastasis. The docking study showed that NADH, reproterol, and labetalol bind to *SERPINE1* with high affinity. The molecule labetalol makes hydrogen-bonding interactions with Tyr37, Ser41, Asp95, Phe117, and Arg118. The phenyl ring sits in the vicinity of hydrophobic residues such as Phe64, Ile66, and Phe117 (Figure 7D). We could not find a good binder for CKS2. The overall analysis indicates that the identified small molecules hold potential as possible therapeutics for OSCC and other cancers.

Stability analysis of the drug–ligand complex

Post docking MD simulation was run to check the stability of the ligands inside the binding cavity of the proteins. The stability of the simulation was analyzed by RMSD. The hydrogen bond interactions between ligand and protein were also calculated to check their strength.

The RMSD analysis of the MD simulation of the BOP1–saquinavir complex indicated that the simulation stabilized at about 10 ns. At about 40 ns, there is a slight change in ligand RMSD (<1 Å). Thereafter, the trajectory is very smooth, and both the protein and ligand RMSDs show a very stable trajectory (Figure 8A). The hydrogen bond analysis showed that saquinavir has two hydrogen-bonding interactions (with Thr149 and Trp150) with >50% occupancy (Table 1), which indicates that the HB interactions are strong and the BOP1–ligand complex is highly stable.

The RMSD during the MD simulation of the PLAU–NADH complex initially increases until about 10 ns; afterward, its movement is confined in a small window (<2 Å) (Figure 8B). This clearly indicated that the simulation is stable. The ligand RMSD was also calculated. It shows a close trend as that of the protein, again indicating good stability of the ligand inside the

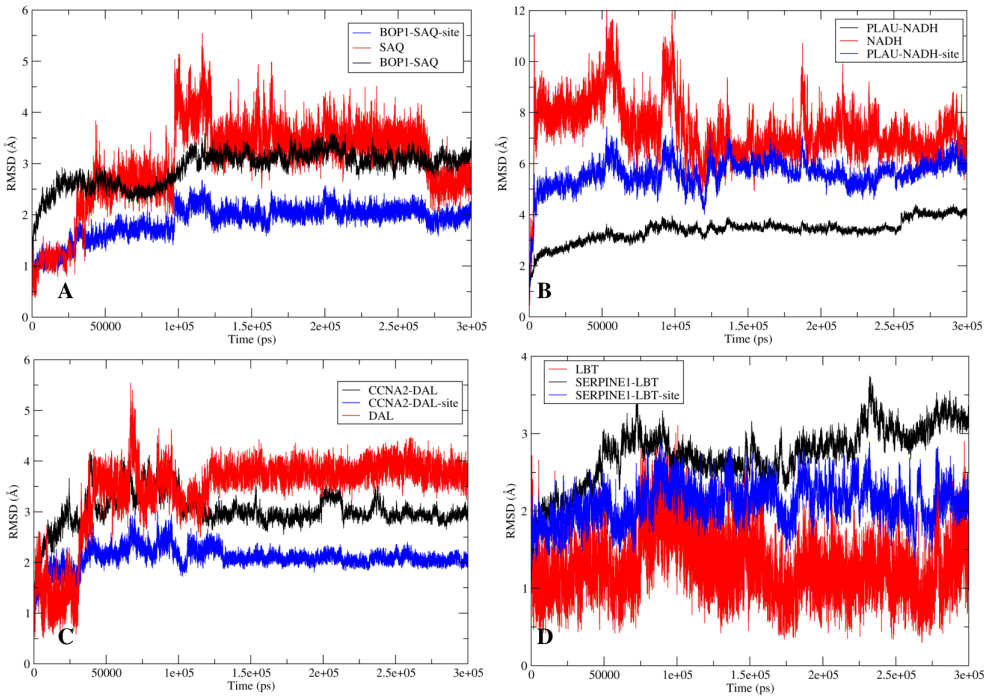


FIGURE 8
The stability analysis: The root mean square deviation (RMSD) analysis for the protein-ligand complexes. The RMSD of protein is shown in red while the RMSD of ligand is shown in blue. **(A)** RMSD for BOP1-Saquinavir complex. **(B)** RMSD for PLAU-NADH complex. **(C)** RMSD for CCNA2-Diacetolol complex. **(D)** RMSD for SERPINE1-Labetalol complex.

TABLE 1 Hydrogen bond analysis between the ligand and the protein.

Protein BOP	S. No.	Residue	Ligand	Occupancy*
	1	THR149	Saquinavir	63.43%
	2	TRP150	Saquinavir	63.51%
	3	PHE195	Saquinavir	31.31%
PLAU				
	1	LYS154	NADH	71.71%
	2	TYR31	NADH	22.49%
	3	SER157	NADH	16.89%
CCNA2				
	1	TYR176	Diacetolol	19.89%
	2	LYS175	Diacetolol	14.47%
SERPINE1				
	1	ASP95	Labetalol	60.24%
	2	TYR37	Labetalol	16.19%

*Defined as the HB interaction present in x% of frames out of a total of 100,000 frames.

binding cavity. Hydrogen bonds of the ligand with Lys154 and Tyr31 showed occupancy of about 71% and 23%, respectively (Table 1). All of these observations point to the high stability of the NADH–PLAU complex. The RMSD analysis of the MD simulation of the CCNA2–diacetolol complex indicated that the simulation stabilized quickly, as indicated by the movement of the RMSD in a narrow window. Both the protein and ligand RMSDs show a stable trajectory; however, we could not find a strong hydrogen bonding interaction between diacetolol and CCNA2 (Figure 8C and Table 1).

The MD simulation of Serpine1 indicated that after an initial increase, the RMSD gets stabilized quickly, as indicated by a rangebound (<2 Å) movement of the RMSD curve (Figure 8D). The ligand RMSD was also calculated, and it showed movement in a very narrow window, indicating the stability of the simulation and that of the complex. The HB analysis indicated that the hydrogen bond between the ligand and Asp95 is highly stable, as measured by an occupancy of >60% (Table 1). Overall analysis indicates that the Serpine1–labetalol complex is highly stable.

Discussion

In recent years, prognosis-based gene signature identification has been of immense interest for the prediction of outcome or for evaluation of the course of a disease (28, 29). Therapeutic biomarker prediction models are currently in focus to identify predictive factors of the response to chemotherapy (30, 31). Non-responding cancer cells are either refractory to chemotherapy or have acquired resistance during the course of the treatment. Both are strongly related to molecular alteration of the targets. The major challenge is acquired chemoresistance during treatment, which eventually leads to cancer regrowth, even if the tumor initially responds to the chemotherapeutic agent. Moreover, reports suggest changes in gene and protein expression levels in cancer tissues and cell lines after chemotherapy. For instance, a xenograft study has shown that drug treatment (5-fluorouracil and cisplatin) first reduces the expression level of drug-specific sensitivity-related genes followed by their upregulation in the regrowth phase and emergence of chemoresistance in esophageal cancer (32). Chemoresistance, either intrinsic or acquired, contributes to the low survival of the patients, necessitating the need for identification of drug-specific sensitivity-related markers and their modulators to enhance the sensitivity of cancer cells to subdue/delay chemoresistance. One such example is inhibition of the Phosphoinositode-3-kinase/AKT serine/threonine kinase (PI3K/AKT) pathway to increase the sensitivity and reverse acquired resistance of esophageal cancer cells to chemotherapeutic drugs (33). Thus, it is imperative to identify the therapeutic targets that not only are advantageous in predicting the clinical outcome but also can resolve the emerging chemotherapy-related issues as well.

In our study, using different genomic data, we examined the changes in gene expression before and after treatment to investigate the key genes, pathways, biomarkers, and risk gene signature. Interestingly, we found that most of the genes that were highly expressed in OSCC patients are found to be significantly downregulated after doxorubicin treatment. We also get another cluster of genes whose expression does not change significantly after doxorubicin treatment. In fact, the expression of some of them increases further. In total, 168 overlapping DEGs were found from the three datasets and were considered for further study. Five gene signatures (*SERPINE1*, *PLAU*, *BOP1*, *CKS2*, and *CCNA2*) were proposed in this study through Cox regression and machine learning. The overexpression of these genes increases the risk of adverse

outcome. Application of Cox regression resulted in an equation that can be used for risk stratification of OSCC patients. From the survival analysis, we were able to predict that increased expression of these genes is related to poor prognosis of the patients. A similar study found SMA and SERPINE1 to be significantly associated with prognosis in OSCC (34); SERPINE1, PLAU, and ACTA1 acted as both diagnostic and prognostic markers (35). Another study by Liu et al. (36) identified an eight-gene prognostic signature in HNSCC that included PLAU. CCNA2 is identified as an independent indicator of worse OS and may serve as a reliable biomarker to identify high-risk subgroups with poor prognosis in OSCC (37). Moreover, these five genes were highly expressed and reported to be associated with cancer progression in various cancers including HNSCC (35, 38), glioblastoma multiforme (GBM) (39), epithelial ovarian cancer (40, 41), gastric cancer (42, 43), breast cancer (44, 45), bladder cancer (46, 47), esophageal cancer (48), colorectal cancer (49, 50), hepatocellular carcinoma (51), melanoma (52), and non-small cell lung carcinoma (NSCLC) (53). In cancer pathogenesis, these genes perturbed numerous cellular mechanisms such as extracellular matrix (ECM) modulation, epithelial-to-mesenchymal transition (EMT), cell migration, and angiogenesis (39, 47–49, 54–59). Interestingly, from a genetic study, we also found that gene amplification in *SERPINE1*, *PLAU*, *BOP1*, and *CKS2* is the primary cause of their overexpression, except in *CCNA2*, where mutation is the predominant cause in HNSCC. Experimental evidence showed that amplification of these genes occurs at both mRNA and protein levels in several cancers such as breast cancer (60, 61), prostate cancer (62), rectal cancer (63), hepatocellular cancer (51), NSCLC (64), gastric cancer (65), and tongue cancer (66).

Overexpression of these markers is indicative of poor prognosis and corresponds to chemotherapy resistance. We showed that initially during chemotherapy, there is downregulation of these genes in the responsive cancer (Figure 5C), but as treatment continues, these genes are upregulated to give rise to chemoresistant phenotypes (Figure 5D). We examined the mRNA expression of these five genes with different drug treatments, which include doxorubicin, cisplatin, oxaliplatin, gemcitabine, topotecan, temozolomide, paclitaxel, JQ1, romidepsin, dauricine, and SNSO32, in various cancers to support our outcome. All of the drugs that we used in our study have DNA as their target and inhibit DNA replication or transcription. In another analysis with the drug olaparib, which is a poly(ADP-ribose) polymerase 1 (PARP1) inhibitor (GSE165585), we found that our genes of interest were not significantly expressed (Supplementary Table S3). In support of our results, we found that PARP1 inhibitors help in the sensitization of temozolomide-resistant glioblastoma cancer (67). We found that after drug treatment, gene expression changes from high to low or low to high or remains unchanged. Further additional new gene expression was also seen. We found that during the period of treatment responsiveness, our genes change their expression level from low to high again, which is the same as the prior treatment. Other studies have also determined the involvement of these genes in resistance generation, for example, *SERPINE1* is upregulated in cisplatin-resistant oral cancer cell lines (SCC9, SCC4, and H357) (68) and paclitaxel-resistant breast cancer (44). In the context of therapeutic potential, SERPINE1 has been identified as a potential therapeutic target, as it acts as a pro-proliferative oncogenic factor (69). *PLAU* is found to be upregulated in cisplatin-resistant oral cancer cell lines (68). Both *PLAU* and *SERPINE1* were found to be highly expressed in breast cancer patients with adjuvant endocrine therapy and related to shorter disease-free survival and OS (70). Whereas *in vivo*, BOP1 downregulation was reported to inhibit paclitaxel resistance and Cancer stem cells (CSC)-like phenotype in triple-negative breast cancer (TNBC) cells (71). In estrogen receptor–positive (ER+) breast cancer, tamoxifen resistance is correlated with the overexpression of *CCNA2* (72). CKS2 in complex with SSPB1 regulates mitochondrion DNA replication in cervical cancer and can be indicative of chemoradioresistance (73). In the present study, we found that these genes are upregulated and are related to resistance, and they are involved in vital biological processes (Figure 9); therefore, targeting these genes can be of immense therapeutic benefit.

Thus, it is imperative to identify modulators that can check aberrant expression of these genes and increase the barrier toward the emergence of chemoresistance. In this milieu, we also identified small-molecule ligands that can enhance the efficacy/sensitivity of chemotherapeutic agents. A combination of molecules targeting proteins can provide a potent therapeutic option with reduced changes of emergence of chemoresistance. Many of the identified molecules are already reported to be effective against cancers such as colorectal, breast, and lung. Some of the selected proteins are involved in mutually exclusive pathways, as evident from the network analysis, and thus have different mechanisms of action. Simultaneously targeting them can be advantageous to both primary tumor and advanced metastatic tumors. Most of the top-scoring molecules for BOP1 were HIV protease inhibitors, e.g., saquinavir, indinavir, nelfinavir, and ritonavir (Norvir). They are reported to induce cell death in both the chemosensitive and chemoresistant ovarian cancer cell lines in a dose-dependent manner (74). The lopinavir/ritonavir combination is reported to have significant inhibition on cell growth and migration, whereas it enhanced radiosensitivity in HNSCC cell lines (75). The anticancer potential of protease inhibitors is already reported in various previous publications. Some of these molecules are in clinical trials. Therefore, they may present attractive options as targeted chemotherapeutic agents against chemosensitive and chemoresistant OSCC as well (76–78). Moreover, BOP1 inhibition provides additional advantage for non-cancerous cells by inducing a cytoprotective nucleolar stress response and reducing damage to normal tissues from anticancer

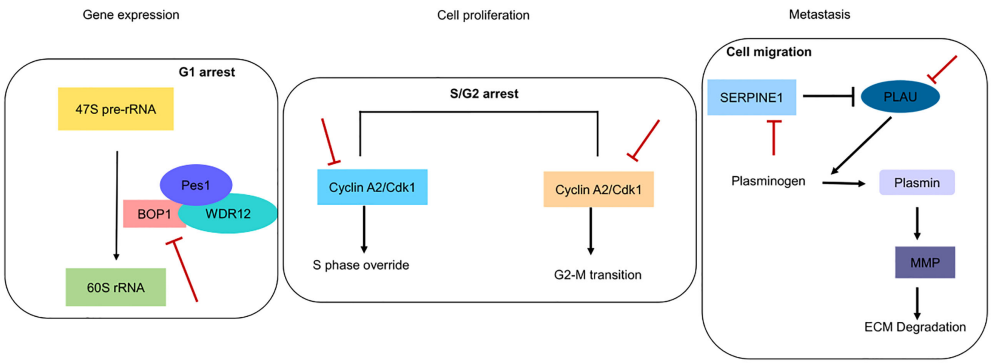


FIGURE 9
Pictorial depiction of identified genes in biological pathways. Red blunt lines indicate that their pharmacological inhibition can inhibit major functions involved in gene expression, cell proliferation and metastasis.

drugs such as camptothecin or methotrexate. It is reported that *BOP1* inhibition together with camptothecin results in selective killing of p53 null cells, producing a synergistic effect (79, 80). *PLAU* encodes uPA, commonly associated with cancer progression *via* apoptosis inhibition and breakdown of the ECM. It also promotes angiogenesis (81). PLAU, as a top identified molecule, inhibits various biological functions such as NADH due to its role in energy production and may help against wastage and weakness of cancer patients. It is also used to improve mental alertness. It can be given orally and has cleared clinical trials as a chemotherapeutic agent for other illnesses (27). Argatroban, which is an antithrombotic agent, inhibits the metastasis in breast cancer (82) and melanomas (83). It has significant antineoplastic effect on gliomas as well (84). It downregulates the MAPK/ERK and STAT phosphorylation that results in a reduction of interleukin (IL)-6, IL-12, and tumor necrosis factor (85). The antiparasitic pentamidine has been shown to be effective against various human cancers such as melanoma (86), breast cancer (87), lung cancer, ovarian cancer, and cervical cancer (88, 89). However, the mechanism of its antineoplastic action has not been elucidated fully. Overall, these results infuse great confidence in our analysis. Labetalol, which we have identified as a SERPINE1 binder, blocks both alpha and beta adrenoceptors and has been used for the treatment of hypertension. Alpha-blockers have been reported to increase recurrence-free survival (RFS) (90). Recently, it has been suggested that beta-blockers hinder mechanisms that initiate tumorigenesis, angiogenesis, and metastasis. Beta-blockers have shown good antineoplastic activity in various cancer cell lines. They are also reported to increase the effect of anticancer chemotherapy (91, 92). Therefore, labetalol, having a mix of alpha- and beta-blocker activities, can be a potential candidate for the treatment of OSCC. Reproterol is a β2-agonist used as an antiasthmatic drug. Thus, we suggest that targeting BOP1 and/or PLAU can be advantageous against both primary and metastatic tumors. Their inhibitors can also be combined together with fluorouracil and methotrexate (93). We have also identified CCNA2 ligands, whereas no direct small-molecule inhibitors for CCNA2 are yet known. While targeting the CCNA2 function, we will be able to target the S/G2/G2-M phase of the cell cycle.

Conclusion

In this study, we have identified changes in the gene expression level as a result of treatment in OSCC. Applying machine learning techniques and Cox regression, we constructed a five-gene-based prognostic signature that can stratify the patients (high and low risk). It is evident that overexpression of these genes is related to poor prognosis and reduced survival in many cancers including OSCC. They are also related to the emergence of chemoresistance against many drugs. Changes in their expression levels, pretreatment, posttreatment, and resistant cell lines make them suitable for targeting. We have also identified potential molecules that can bind to these proteins with high affinity. Since the identified proteins are involved in disparate processes, a combination of molecules targeting them can provide a potent therapeutic option with reduced chances of chemoresistance. We hope that this study provides new avenues for the design of better chemotherapeutic agents especially against chemoresistance in OSCC.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

The work was designed and conceptualized by PK and AD. PK, SK, AD, SB, MS, and BM analyzed the data and performed the work. The manuscript was prepared and written by PK, SK, and AD. All authors read and approved the final manuscript.

Acknowledgments

The authors acknowledge the Director, Institute of Life Sciences (ILS), Bhubaneswar and the Department of Biotechnology, Government of India for providing necessary high-performance computing facility. PK and SK are thankful for the University Grant Commission (UGC) for their fellowship. SB and BM is thankful for CSIR for their fellowship. The ILS high performance computing facility is duly acknowledged for providing computational facility for this work. This work is supported by an ILS core grant.

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.2022.910494/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) 7:1–9. doi: 10.3322/caac.21660

2. Jitender S, Sarika G, Varada HR, Omprakash Y, Mohsin K. Screening for oral cancer. *J Exp Ther Oncol* (2016) 11(4):303–7. doi: 10.1007/s12032-021-01548-0

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590

4. Nocini R, Capocasale G, Marchioni D, Zotti F. A snapshot of knowledge about oral cancer in Italy: A 505 person survey. *Int J Environ Res Public Health* (2020) 17(13):4889. doi: 10.3390/ijerph17134889

5. Hartner L. Chemotherapy for oral cancer. *Dental Clinics North Am* (2018) 62 (1):87–97. doi: 10.1016/j.cden.2017.08.006

6. Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefèbvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *New Engl J Med* (2004) 350(19):1945–52. doi: 10.1056/NEJMoa032641

7. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a Large-scale small-molecule sensitivity dataset. *Cancer Discov* (2015) 5(11):1210–23. doi: 10.1158/2159-8290.CD-15-0235

8. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* (2016) 12(2):109–16. doi: 10.1038/nchembio.1986

9. Guan YF, Li GR, Wang RJ, Yi YT, Yang L, Jiang D, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer* (2012) 31(10):463–70. doi: 10.5732/cjc.012.10216

10. Lee YT, Tan YJ, Oon CE. Molecular targeted therapy: Treating cancer with specificity. *Eur J Pharmacol* (2018) 834:188–96. doi: 10.1016/j.ejphar.2018.07.034

11. Thorn CF, Oshiro C, Marsh S, Hernandez-Boussard T, McLeod H, Klein TE, et al. Doxorubicin pathways: pharmacodynamics and adverse effects. *Pharmacogenet Genomics* (2011) 21(7):440–6. doi: 10.1097/FPC.0b013e32833ffb56

12. Sakagami H, Okudaira N, Masuda Y, Amano O, Yokose S, Kanda Y, et al. Induction of apoptosis in human oral keratinocyte by doxorubicin. *Anticancer Res* (2017) 37(3):1023–9. doi: 10.21873/anticanres.11412

13. Ambatipudi S, Gerstung M, Pandey M, Samant T, Patil A, Kane S, et al. Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Genes Chromosomes Cancer* (2012) 51(2):161–73. doi: 10.1002/gcc.20940

14. Hazar-Rethinam M, de Long LM, Gannon OM, Topkas E, Boros S, Vargas AC, et al. A novel E2F/sphingosine kinase 1 axis regulates anthracycline response in squamous cell carcinoma. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2015) 21 (2):417–27. doi: 10.1158/1078-0432.CCR-14-1962

15. Wu G, Haw R. Functional interaction network construction and analysis for disease discovery. *Methods Mol Biol (Clifton NJ)* (2017) 1558:235–53. doi: 10.1007/978-1-4939-6783-4_11

16. Sidiropoulos K, Viteri G, Sevilla C, Jupe S, Webber M, Orlic-Milacic M, et al. Reactome enhanced pathway visualization. *Bioinf (Oxford England)* (2017) 33 (21):3461–7. doi: 10.1093/bioinformatics/btx441

17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Gene Ontol Consort Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

18. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res* (2021) 49 (D1):D325–D34. doi: 10.1093/nar/gkaa1113

19. Toth R, Schiffmann H, Hube-Magg C, Buscheck F, Hoflmayer D, Weidemann S, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin Epigenetics* (2019) 11(1):148. doi: 10.1186/s13148-019-0736-8

20. Lind AP, Anderson PC. Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PloS One* (2019) 14(7):e0219774. doi: 10.1371/journal.pone.0219774

21. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* (2003) 4:2. doi: 10.1186/1471-2105-4-2

22. 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(W1):W556–W60. doi: 10.1093/nar/gkz430

23. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095

24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signaling* (2013) 6(269):11. doi: 10.1126/scisignal.2004088

25. Li W, Song P, Zhao M, Gao L, Xie J, You C. BOP1 used as a novel prognostic marker and correlated with tumor microenvironment in pan-cancer. *J Oncol* (2021) 2021:3603030. doi: 10.1155/2021/3603030

26. Chen G, Sun J, Xie M, Yu S, Tang Q, Chen L. PLAU promotes cell proliferation and epithelial-mesenchymal transition in head and neck squamous cell carcinoma. *Front Genet* (2021) 12. doi: 10.3389/fgene.2021.651882

27. Ju H-Q, Lin J-F, Tian T, Xie D, Xu R-H. NADPH homeostasis in cancer: functions, mechanisms and therapeutic implications. *Signal Transd Target Ther* (2020) 5(1):231. doi: 10.1038/s41392-020-00326-0

28. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ* (2009) 338:b375. doi: 10.1136/bmj.b375

29. Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Glazko G, Emmert-Streib F. Prognostic gene expression signatures of breast cancer are lacking a sensible biological meaning. *Sci Rep* (2021) 11(1):156. doi: 10.1038/s41598-020-79375-y

30. Inoue S, Yoshida T, Nishino T, Goto M, Aoyama M, Kawakita N, et al. Biomarkers predicting the response to chemotherapy and the prognosis in patients with esophageal squamous cell carcinoma. *Gen Thorac Cardiovasc Surg* (2021) 69 (3):525–33. doi: 10.1007/s11748-021-01586-5

31. Wei X, Zhou X, Zhao Y, He Y, Weng Z, Xu C. A 14-gene gemcitabine resistance gene signature is significantly associated with the prognosis of pancreatic cancer patients. *Sci Rep* (2021) 11(1):6087. doi: 10.1038/s41598-021-85680-x

32. Yoshida T, Miyoshi T, Seike J-I, Yamai H, Takechi H, Yuasa Y, et al. Gene expression changes in a chemoresistant model with human esophageal cancer xenografts using cDNA microarray. *Anticancer Res* (2009) 29(4):1163–8.

33. Li B, Li J, Wen Xu W, Guan XY, Qin YR, Zhang LY, et al. Suppression of esophageal tumor growth and chemoresistance by directly targeting the PI3K/AKT pathway. *Oncotarget* (2014) 5(22):11576–87. doi: 10.18632/oncotarget.2596

34. Dhanda J, Triantafyllou A, Liloglou T, Kalirai H, Lloyd B, Hanlon R, et al. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. *Br J Cancer* (2014) 111(11):2114–21. doi: 10.1038/bjc.2014.500

35. Yang K, Zhang S, Zhang D, Tao Q, Zhang T, Liu G, et al. Identification of SERPINE1, PLAU and ACTA1 as biomarkers of head and neck squamous cell carcinoma based on integrated bioinformatics analysis. *Int J Clin Oncol* (2019) 24 (9):1030–41. doi: 10.1007/s10147-019-01435-9

36. Liu B, Su Q, Ma J, Chen C, Wang L, Che F, et al. Prognostic value of eight-gene signature in head and neck squamous carcinoma. *Front Oncol* (2021) 11:657002. doi: 10.3389/fonc.2021.657002

37. Monteiro LS, Diniz-Freitas M, Warnakulasuriya S, Garcia-Caballero T, Forteza-Vila J, Fraga M. Prognostic significance of cyclins A2, B1, D1, and E1 and CCND1 numerical aberrations in oral squamous cell carcinomas. *Anal Cell Pathol (Amst)* (2018) 2018:7253510. doi: 10.1155/2018/7253510

38. Li Z, Chen C, Wang J, Wei M, Liu G, Qin Y, et al. Overexpressed PLAU and its potential prognostic value in head and neck squamous cell carcinoma. *PeerJ* (2021) 9:e10746. doi: 10.7717/peerj.10746

39. Seker F, Cingoz A, Sur-Erdem İ, Erguder N, Erkent A, Uyulur F, et al. Identification of SERPINE1 as a regulator of glioblastoma cell dispersal with transcriptome profiling. *Cancers (Basel)* (2019) 11(11):1651. doi: 10.3390/cancers11111651

40. Pan J-X, Qu F, Wang F-F, Xu J, Mu L-S, Ye L-Y, et al. Aberrant SERPINE1 DNA methylation is involved in carboplatin induced epithelial-mesenchymal transition in epithelial ovarian cancer. *Arch Gynecol Obstet* (2017) 296(6):1145–52. doi: 10.1007/s00404-017-4547-x

41. Xu JH, Wang Y, Xu D. CKS2 promotes tumor progression and metastasis and is an independent predictor of poor prognosis in epithelial ovarian cancer. *Eur Rev Med Pharmacol Sci* (2019) 23(8):3225–34. doi: 10.26355/eurrev_201904_17681

42. Li L, Zhu Z, Zhao Y, Zhang Q, Wu X, Miao B, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci Rep* (2019) 9(1):7827. doi: 10.1038/s41598-019-43924-x

43. Ai C, Zhang J, Lian S, Ma J, Győrffy B, Qian Z, et al. FOXM1 functions collaboratively with PLAU to promote gastric cancer progression. *J Cancer* (2020) 11(4):788–94. doi: 10.7150/jca.37323

44. Zhang Q, Lei L, Jing D. Knockdown of SERPINE1 reverses resistance of triple−negative breast cancer to paclitaxel *via* suppression of VEGFA. *Oncol Rep* (2020) 44(5):1875–84. doi: 10.3892/or.2020.7770

45. Urban P, Vuaroqueaux V, Labuhn M, Delorenzi M, Wirapati P, Wight E, et al. Increased expression of urokinase-type plasminogen activator mRNA determines adverse prognosis in ErbB2-positive primary breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2006) 24(26):4245–53. doi: 10.1200/JCO.2005.05.1912

46. Chen Z, Chen X, Xie R, Huang M, Dong W, Han J, et al. DANCR promotes metastasis and proliferation in bladder cancer cells by enhancing IL-11-STAT3 signaling and CCND1 expression. *Mol Ther J Am Soc Gene Ther* (2019) 27(2):326–41. doi: 10.1016/j.ymthe.2018.12.015

47. Chen R, Feng C, Xu Y. Cyclin-dependent kinase-associated protein Cks2 is associated with bladder cancer progression. *J Int Med Res* (2011) 39(2):533–40. doi: 10.1177/147323001103900222

48. Fang L, Che Y, Zhang C, Huang J, Lei Y, Lu Z, et al. PLAU directs conversion of fibroblasts to inflammatory cancer-associated fibroblasts, promoting esophageal squamous cell carcinoma progression *via* uPAR/Akt/NF-κB/IL8 pathway. *Cell Death Discov* (2021) 7(1):32. doi: 10.1038/s41420-021-00410-6

49. Qi J, Yu Y, Akilli Öztürk Ö, Holland JD, Besser D, Fritzmann J, et al. New wnt/β-catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals. *Gut* (2016) 65(10):1690–701. doi: 10.1136/gutjnl-2014-307900

50. Chen X, Zhao Y. Block of proliferation 1 promotes cell migration and invasion in human colorectal cancer cells *via* the JNK pathway. *J Clin Lab Anal* (2020) 34(7):e23283. doi: 10.1002/jcla.23283

51. Chung KY, Cheng IK, Ching AK, Chu JH, Lai PB, Wong N. Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. *Hepatol (Baltimore Md)* (2011) 54(1):307–18. doi: 10.1002/hep.24372

52. Gupta R, Bugide S, Wang B, Green MR, Johnson DB, Wajapeyee N. Loss of BOP1 confers resistance to BRAF kinase inhibitors in melanoma by activating MAP kinase pathway. *Proc Natl Acad Sci USA* (2019) 116(10):4583–91. doi: 10.1073/pnas.1821889116

53. Ruan JS, Zhou H, Yang L, Wang L, Jiang ZS, Wang SM. CCNA2 facilitates epithelial-to-mesenchymal transition *via* the integrin αvβ3 signaling in NSCLC. *Int J Clin Exp Pathol* (2017) 10(8):8324–33.

54. Leik CE, Su EJ, Nambi P, Crandall DL, Lawrence DA. Effect of pharmacologic plasminogen activator inhibitor-1 inhibition on cell motility and tumor angiogenesis. *J Thromb Haemost JTH* (2006) 4(12):2710–5. doi: 10.1111/j.1538-7836.2006.02244.x

55. Xu B, Bai Z, Yin J, Zhang Z. Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. *PeerJ* (2019) 7:e7091. doi: 10.7717/peerj.7091

56. Kong HJ, Kwon EJ, Kwon OS, Lee H, Choi JY, Kim YJ, et al. Crosstalk between YAP and TGFβ regulates SERPINE1 expression in mesenchymal lung cancer cells. *Int J Oncol* (2021) 58(1):111–21. doi: 10.3892/ijo.2020.5153

57. Danø K, Behrendt N, Høyer-Hansen G, Johnsen M, Lund LR, Ploug M, et al. Plasminogen activation and cancer. *Thromb Haemost* (2005) 93(4):676–81. doi: 10.1160/TH05-01-0054

58. Pestov DG, Strezoska Z, Lau LF. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol* (2001) 21(13):4246–55. doi: 10.1128/MCB.21.13.4246-4255.2001

59. Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadéa G, et al. A novel function for cyclin A2: control of cell invasion *via* RhoA signaling. *J Cell Biol* (2012) 196(1):147–62. doi: 10.1083/jcb.201102085

60. Sternlicht MD, Dunning AM, Moore DH, Pharoah PD, Ginzinger DG, Chin K, et al. Prognostic value of PAI1 in invasive breast cancer: Evidence that tumor-specific factors are more important than genetic variation in regulating PAI1 expression. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored by Am Soc Prev Oncol* (2006) 15(11):2107–14. doi: 10.1158/1055-9965.EPI-06-0351

61. Meng S, Tripathy D, Shete S, Ashfaq R, Saboorian H, Haley B, et al. uPAR and HER-2 gene status in individual breast cancer cells from blood and tissues. *Proc Natl Acad Sci* (2006) 103(46):17361–5. doi: 10.1073/pnas.0608113103

62. Helenius MA, Saramäki OR, Linja MJ, Tammela TLJ, Visakorpi T. Amplification of urokinase gene in prostate cancer. *Cancer Res* (2001) 61(14):5340–4.

63. Lips EH, van Eijk R, de Graaf EJR, Oosting J, de Miranda NFCC, Karsten T, et al. Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis. *BMC Cancer* (2008) 8(1):314. doi: 10.1186/1471-2407-8-314

64. Micke P, Edlund K, Holmberg L, Kultima HG, Mansouri L, Ekman S, et al. Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2011) 6(11):1833–40. doi: 10.1097/JTO.0b013e3182295917

65. Jin DH, Park SE, Lee J, Kim KM, Kim S, Kim DH, et al. Copy number gains at 8q24 and 20q11-q13 in gastric cancer are more common in intestinal-type than diffuse-type. *PloS One* (2015) 10(9):e0137657. doi: 10.1371/journal.pone.0137657

66. Gao F, Li C, Zhao X, Xie J, Fang G, Li Y. CKS2 modulates cell-cycle progression of tongue squamous cell carcinoma cells partly *via* modulating the cellular distribution of DUTPase. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol* (2021) 50(2):175–82. doi: 10.1111/jop.13116

67. Zhang J, Stevens MF, Bradshaw TD. Temozolomide: mechanisms of action, repair and resistance. *Curr Mol Pharmacol* (2012) 5(1):102–14. doi: 10.2174/1874467211205010102

68. Shriwas O, Arya R, Mohanty S, Mohapatra P, Kumar S, Rath R, et al. RRBP1 rewires cisplatin resistance in oral squamous cell carcinoma by regulating hippo pathway. *Br J Cancer* (2021) 124:2004–16. doi: 10.1101/2020.03.18.998070

69. Zhao C, Liu Z. MicroRNA 617 targeting SERPINE1 inhibited the progression of oral squamous cell carcinoma. *Mol Cell Biol* (2021) 41(6): e0056520. doi: 10.1128/MCB.00565-20

70. Jevrić M, Matić IZ, Krivokuća A, Đorđić Crnogorac M, Besu I, Damjanović A, et al. Association of uPA and PAI-1 tumor levels and 4G/5G variants of PAI-1 gene with disease outcome in luminal HER2-negative node-negative breast cancer patients treated with adjuvant endocrine therapy. *BMC Cancer* (2019) 19(1):71. doi: 10.1186/s12885-018-5255-z

71. Li S, Wu H, Huang X, Jian Y, Kong L, Xu H, et al. BOP1 confers chemoresistance of triple-negative breast cancer by promoting CBP-mediated β-catenin acetylation. *J Pathol* (2021) 254:265–78. doi: 10.1002/path.5676

72. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. *PloS One* (2014) 9(3):e91771. doi: 10.1371/journal.pone.0091771

73. Jonsson M, Fjeldbo CS, Holm R, Stokke T, Kristensen GB, Lyng H. Mitochondrial function of CKS2 oncoprotein links oxidative phosphorylation with cell division in chemoradioresistant cervical cancer. *Neoplasia (New York NY)* (2019) 21(4):353–62. doi: 10.1016/j.neo.2019.01.002

74. Bandiera E, Todeschini P, Romani C, Zanotti L, Erba E, Colmegna B, et al. The HIV-protease inhibitor saquinavir reduces proliferation, invasion and clonogenicity in cervical cancer cell lines. *Oncol Lett* (2016) 12(4):2493–500. doi: 10.3892/ol.2016.5008

75. Liu R, Zhang L, Yang J, Zhang X, Mikkelsen R, Song S, et al. HIV Protease inhibitors sensitize human head and neck squamous carcinoma cells to radiation by activating endoplasmic reticulum stress. *PloS One* (2015) 10(5):e0125928. doi: 10.1371/journal.pone.0125928

76. Koltai T. Nelfinavir and other protease inhibitors in cancer: mechanisms involved in anticancer activity. *F1000Research* (2015) 4:9. doi: 10.12688/f1000research.5827.2

77. Trezza A, Cicaloni V, Pettini F, Spiga O. Chapter 2 - potential roles of protease inhibitors in anticancer therapy. In: SP Gupta, editor. *Cancer-leading proteases*. Cambridge, Massachusetts: Academic Press (2020). p. 13–49.

78. Yang P, Li ZY, Li HQ. Potential roles of protease inhibitors in cancer progression. *Asian Pac J Cancer Prev APJCP* (2015) 16(18):8047–52. doi: 10.7314/APJCP.2015.16.18.8047

79. Sapio RT, Nezdyur AN, Krevetski M, Anikin L, Manna VJ, Minkovsky N, et al. Inhibition of post-transcriptional steps in ribosome biogenesis confers cytoprotection against chemotherapeutic agents in a p53-dependent manner. *Sci Rep* (2017) 7(1):9041. doi: 10.1038/s41598-017-09002-w

80. Pecoraro A, Pagano M, Russo G, Russo A. Ribosome biogenesis and cancer: Overview on ribosomal proteins. *Int J Mol Sci* (2021) 22(11):5496. doi: 10.3390/ijms22115496

81. Mahmood N, Mihalcioiu C, Rabbani SA. Multifaceted role of the urokinase-type plasminogen activator (uPA) and its receptor (uPAR): Diagnostic, prognostic, and therapeutic applications. *Front Oncol* (2018) 8(24). doi: 10.3389/fonc.2018.00024

82. Asanuma K, Wakabayashi H, Okamoto T, Asanuma Y, Akita N, Yoshikawa T, et al. The thrombin inhibitor, argatroban, inhibits breast cancer metastasis to bone. *Breast Cancer* (2013) 20(3):241–6. doi: 10.1007/s12282-012-0334-5

83. Asanuma K, Wakabayashi H, Hayashi T, Okuyama N, Seto M, Matsumine A, et al. Thrombin inhibitor, argatroban, prevents tumor cell migration and bone metastasis. *Oncology* (2004) 67(2):166–73. doi: 10.1159/000081004

84. Hua Y, Tang L, Keep RF, Schallert T, Fewel ME, Muraszko KM, et al. The role of thrombin in gliomas. *J Thromb Haemost JTH* (2005) 3(9):1917–23. doi: 10.1111/j.1538-7836.2005.01446.x

85. Kuriakose S, Muleme H, Onyilagha C, Okeke E, Uzonna JE. Diminazene aceturate (Berenil) modulates LPS induced pro-inflammatory cytokine production by inhibiting phosphorylation of MAPKs and STAT proteins. *Innate Immun* (2014) 20(7):760–73. doi: 10.1177/1753425913507488

86. Smith J, Stewart BJ, Glaysher S, Peregrin K, Knight LA, Weber DJ, et al. The effect of pentamidine on melanoma *ex vivo*. *Anti-cancer Drugs* (2010) 21(2):181–5. doi: 10.1097/CAD.0b013e3283340cee

87. Jung HJ, Suh SI, Suh MH, Baek WK, Park JW. Pentamidine reduces expression of hypoxia-inducible factor-1alpha in DU145 and MDA-MB-231 cancer cells. *Cancer Lett* (2011) 303(1):39–46. doi: 10.1016/j.canlet.2011.01.008

88. Chow TY, Alaoui-Jamali MA, Yeh C, Yuen L, Griller D. The DNA double-stranded break repair protein endo-exonuclease as a therapeutic target for cancer. *Mol Cancer Ther* (2004) 3(8):911–9. doi: 10.1158/1535-7163.911.3.8

89. Pathak MK, Dhawan D, Lindner DJ, Borden EC, Farver C, Yi T. Pentamidine is an inhibitor of PRL phosphatases with anticancer activity. *Mol Cancer Ther* (2002) 1(14):1255–64.

90. Hart J, Spencer B, McDermott CM, Chess-Williams R, Sellers D, Christie D, et al. A pilot retrospective analysis of alpha-blockers on recurrence in men with localised prostate cancer treated with radiotherapy. *Sci Rep* (2020) 10(1):8191. doi: 10.1038/s41598-020-65238-z

91. Oh MS, Guzner A, Wainwright DA, Mohindra NA, Chae YK, Behdad A, et al. The impact of beta blockers on survival outcomes in patients with non–small-cell lung cancer treated with immune checkpoint inhibitors. *Clin Lung Cancer* (2021) 22(1):e57–62. doi: 10.1016/j.cllc.2020.07.016

92. Madel M-B, Elefteriou F. Mechanisms supporting the use of beta-blockers for the management of breast cancer bone metastasis. *Cancers* (2021) 13(12):2887. doi: 10.3390/cancers13122887

93. del Rincón SV, Widschwendter M, Sun D, Ekholm-Reed S, Tat J, Teixeira LK, et al. Cks overexpression enhances chemotherapeutic efficacy by overriding DNA damage checkpoints. *Oncogene* (2015) 34(15):1961–7. doi: 10.1038/onc.2014.137

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fonc.2022.952640

OPEN ACCESS

EDITED BY
Giovana Tardin Torrezan,
A. C. Camargo Cancer Center, Brazil

REVIEWED BY
Janusz Blasiak,
University of Łódź, Poland
Gianni Binotto,
University of Padua, Italy
Mohamed A. Yassin,
Hamad Medical Corporation, Qatar

*CORRESPONDENCE
Krzysztof Jamroziak
k.m.jamroziak@gmail.com

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

RECEIVED 25 May 2022
ACCEPTED 11 August 2022
PUBLISHED 23 September 2022

CITATION
Madejczyk AM, Canzian F, Góra-Tybor J, Campa D, Sacha T, Link-Lenczowska D, Florek I, Prejzner W, Całbecka M, Rymko M, Dudziński M, Orzechowska MJ and Jamroziak K (2022) Impact of genetic polymorphisms of drug transporters *ABCB1* and *ABCG2* and regulators of xenobiotic transport and metabolism *PXR* and *CAR* on clinical efficacy of dasatinib in chronic myeloid leukemia. *Front. Oncol.* 12:952640. doi: 10.3389/fonc.2022.952640

Impact of genetic polymorphisms of drug transporters *ABCB1* and *ABCG2* and regulators of xenobiotic transport and metabolism *PXR* and *CAR* on clinical efficacy of dasatinib in chronic myeloid leukemia

Anna Marta Madejczyk[1], Federico Canzian[2], Joanna Góra-Tybor[1], Daniele Campa[3], Tomasz Sacha[4], Dorota Link-Lenczowska[4], Izabela Florek[4], Witold Prejzner[5], M. Całbecka[6], M. Rymko[6], M. Dudziński[7], Magdalena Julita Orzechowska[8] and Krzysztof Jamroziak[9]*

[1]Department of Hematology, Medical University of Łódź, Łódź, Poland, [2]Genomic Epidemiology Group, German Cancer Research Center Deutsche Krebsforschungszentrum (DKFZ), Heidelberg, Germany, [3]Department of Biology, University of Pisa, Pisa, Italy, [4]Department of Hematology, Jagiellonian University Medical College, Kraków, Poland, [5]Department of Hematology, Medical University of Gdańsk, Gdańsk, Poland, [6]Department of Hematology, Copernicus Specialist Municipal Hospital, Toruń, Poland, [7]Department of Hematology, Teaching Hospital No 1, Rzeszów, Poland, [8]Department of Molecular Carcinogenesis, Chair of Molecular Medicine and Biotechnology, Faculty of Medicine, Medical University of Łódź, Łódź, Poland, [9]Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, Warsaw, Poland

Introduction: Functional single-nucleotide polymorphisms (SNPs) in genes regulating cellular uptake, elimination, and metabolism of xenobiotics may potentially influence the outcome of chronic myeloid leukemia (CML) patients treated with BCR-ABL1 tyrosine kinase inhibitors (TKI). Dasatinib, a second-generation TKI, is a substrate of the ABC-superfamily xenobiotic transporters ABCB1 (MDR1, Pg-P) and ABCG2 (BCRP). Pregnane X receptor (PXR, NR1I2) and constitutive androstane receptor (CAR, NR1I3) are involved in the control of expression of *ABCB1* and *ABCG2*.

Aim of the study: In this study, we assessed the impact of inherited variants in *ABCB1*, *ABCG2*, *PXR*, and *CAR* genes on dasatinib efficacy and toxicity in CML.

Materials and methods: Sixty-one tagging SNPs in *ABCB1*, *ABCG2*, *PXR*, and *CAR* genes were analyzed by real-time quantitative PCR with specific probes in 86 CML patients who failed imatinib therapy.

Results: We found the associations between SNPs rs7787082 (*ABCB1*, OR = 0.2; 95% CI = 0.06-0.66, p = 0.008), rs12505410 (*ABCG2*, OR = 3.82; 95% CI = 1.38-10.55; p = 0.010), and rs3114018 (*ABCG2*, OR = 0.24; 95% CI = 0.08-0.71; p = 0.010) and the probability of achieving CCyR. Furthermore, progression-free survival (PFS) was significantly influenced by SNPs rs3732357 (HR = 0.2, 95% CI = 0.26-0.70; p = 0.001), rs3732360 (HR = 0.59; 95% CI = 0.38-0.93; p = 0.020), rs11917714 (HR = 0.58; 95% CI = 0.36-0.92; p = 0.020), and rs3732359 (HR = 0.57; 95% CI = 0.36-0.91; p = 0.024) in *PXR*; rs2307418 (HR = 2.02; 95% CI = 1.19-3.43; p = 0.048) in *CAR*; and rs2235023 (HR = 2.49; 95% CI = 1.13-5.50; p = 0.011) and rs22114102 (HR = 1.90; 95% CI = 1.00-3.63; p = 0.028) in *ABCB1*. Moreover, overall survival (OS) was impacted by rs3842 (HR = 1.84; 95% CI = 1.01-3.33; p = 0.012) and rs2235023 (HR = 2.28; 95% CI = 1.03 = 5.02; p = 0.027) in *ABCB1*, rs11265571 (HR = 1.59; 95% CI = 0.82-3.08; p = 0.037) and rs2307418 (HR = 73.68; 95% CI = 4.47-1215.31; p = 0.003) in *CAR*, and rs3732360 (HR = 0.64; 95% CI = 0.40 = 1.04; p = 0.049) in *PXR*. Taking into account the influence of the tested SNPs on treatment toxicity, we found a significant relationship between allele G of polymorphism in the *ABCB1* rs7787082 (OR = 4.46; 95% CI = 1.38-14.39 p = 0.012) and hematological complications assuming the codominant gene inheritance model as well as a significant correlation between the presence of minor allele (G) of SNP rs2725256 in the *ABCG2* gene (OR = 4.71; 95% CI = 1.20-18.47; p = 0.026) and the occurrence of non-hematological complications assuming a recessive gene inheritance model.

Conclusion: Our data suggest that inherited variants in the genes encoding for proteins involved in the transport of xenobiotics may modify the toxicity and efficacy of dasatinib therapy in CML patients.

KEYWORDS

Chronic myeloid leukemia, dasatinib, ABCB1, ABCG2, PXR, CAR, single nucleotide polymorphisms

Introduction

Chronic myelogenous leukemia (CML) is a rare hematological malignancy with an yearly incidence of one per 100,000 individuals (1). The molecular background of the disease has been elucidated since the discovery of the Philadelphia (Ph) chromosome, a consequence of the reciprocal translocation between chromosomes 9 and 22 and the resulting *BCR-ABL1* fusion gene that encodes for constitutively activated tyrosine kinase (2, 3). Introduction of imatinib, the first-generation BCR-ABL1 tyrosine kinase inhibitor (TKI), has been a great proof of concept of targeted therapy and has dramatically improved CML prognosis (4). However, approximately one-third of patients develop primary or secondary resistance to imatinib (5). Second-generation TKIs such as dasatinib, and later nilotinib and bosutinib, were introduced to treat imatinib-resistant or intolerant CML patients and subsequently also become first-line treatment options (6, 7).

Dasatinib has a different chemical structure than imatinib and higher potential to inhibit BCR-ABL1 kinase (8). The drug targets a broader spectrum of kinases including SRC kinases, and this contributes to its efficacy as well as specific side effects in CML (9). The pharmacokinetic properties of dasatinib are similar to other TKIs. After oral administration, the molecule binds plasma proteins and is metabolized by liver P450 cytochrome family enzymes (mainly CYP3A4) as well as FMO-3 and UDT-glucuronylotransferase. The process of dasatinib excretion from cells is mediated by ATP-binding cassette (ABC) transporters ABCB1 and ABCG2 (10). ABC transporters are a conserved family of membrane proteins responsible for cell protection against xenobiotics (11).

The expression of genes involved in metabolism and excretion of xenobiotics including the ABC genes is regulated by specific xenosensors—genes that are activated as a response to higher concentrations of xenobiotics. Among them, *PXR* (*NR1I2*) and *CAR* (*NR1I3*) possess specific DNA-binding

domains built up from zinc fingers which enable them to recognize DNA elements, characteristic for enzymes taking part in metabolism of xenobiotic substances (12, 13). PXR and CAR stimulate the expression not only of ABCB1 and ABCG2 but also of cytochrome P450 enzymes and many other genes (14, 15). It was shown that single-nucleotide polymorphisms (SNPs) in the aforementioned genes impact on the metabolism of various drugs (16, 17).

The present study aimed to define genetic markers influencing the outcome of dasatinib therapy in patients with imatinib-resistant or intolerant CML. To that end, we analyzed 61 tagging SNPs in *ABCB1, ABCG2, PXR*, and *CAR* genes and studied their effects regarding different parameters of response depth and duration as well as toxicity in CML patients.

Materials and methods

Patients

The study included 86 Polish Caucasian CML patients treated in five tertiary hematological centers in Poland (Department of Hematology, Medical University of Łódź; Department of Hematology, Jagiellonian University Medical College, Kraków; Department of Hematology, Medical University of Gdańsk; Department of Hematology, Copernicus Specialist Municipal Hospital, Toruń; Department of Hematology, Teaching Hospital No 1, Rzeszów). The group included 43 women and 43 men with a median age of 48 years at CML diagnosis (range 18-100). All patients received imatinib at 400 mg/day as a first-line treatment with TKI. The initial dose of dasatinib administered in the second (after imatinib failure) or third (after imatinib and nilotinib failure) line of therapy was 100 mg/day. Complete clinical and laboratory data concerning the course of dasatinib therapy were collected for the analyses of the association with the tested SNPs. Cytogenetic responses were evaluated by classical chromosome banding technique or FISH. The *BCR-ABL1* gene expression was assessed by quantitative real-time PCR according to standard protocols described elsewhere (18).

Clinical endpoints

For the purpose of this analysis, we used definitions of treatment endpoints consistent with the European Leukemia Net recommendations (19). Cytogenetic response (CyR) was defined as complete (0% of Ph chromosome, CCyR), partial (1%–34% of Ph chromosome, PCyR), minimal (35%–65% Ph chromosome, mCyR), and no response (>65% of Ph chromosome). Optimal cytogenetic response was categorized as CCyR achievement within 12 months from treatment start. Molecular responses were defined as complete (CMR) when the *BCR-ABL1* gene transcript level was below 0.01%, major (MMR) when the *BCR-ABL1* gene transcript level was between 0.01% and 0.1%, and no response when the *BCR-ABL1* gene transcript level was >0.1%. Optimal molecular response was defined as achievement of at least MMR within 18 months from treatment start. Progression-free survival (PFS) was defined as the interval between start of dasatinib treatment and CML progression or death from any cause, or last follow-up without progression. Overall survival (OS) was counted as the time between dasatinib treatment start and date of death, or last follow-up when the patient was still alive. Analyzed adverse events of the treatment with dasatinib included hematological toxicities of grade 3 or 4 and any non-hematological complication of grades 2–4.

Polymorphism selection

Tagging SNPs for *ABCB1*, *ABCG2*, *PXR*, and *CAR* were selected using the Tagger algorithm, available through Haploview (20), using pairwise SNP selection with a minimum r^2 threshold of 0.8.

The set of common genetic variants (sequences including 5 kb upstream of the first exon and 5 kb downstream of the last exon of each gene), with minor allele frequency (MAF) ≥5% in Caucasians from the International HapMap Project (21), was included for *ABCB1*, *ABCG2*, *PXR*, and *CAR*. This process resulted in a selection of 26 tagging SNPs for *ABCB1* (average r^2 of tagging SNPs with the SNPs they tag = 0.958), 17 SNPs for *ABCG2* (average r^2 = 0.965), 11 SNPs for *PXR* (average r^2 = 0.975), and seven SNPs for *CAR* (average r^2 = 1.000). This selection therefore captures a very high degree of the known common variability in these genes (22).

Genotyping

DNA was isolated from peripheral blood samples using DNA Blood Mini Kit (Qiagen) and genotyped. Duplicates of 8% of the samples were interspersed throughout the plate for ensuring the internal quality controls. Both TaqMan (ABI, Applied Biosystems, Foster City, CA, USA) and KASP (KBioscence, Hoddesdon, UK) technologies were used for genotyping according to the manufacturer's protocol. PCR plates for TaqMan as well as KASP assays were read on a Viia7 Real-Time PCR platform (Applied Biosystems). The Viia7 RUO Software (Applied Biosystems) was used to determine the genotypes. In our analysis, all individuals with a call rate <80% were excluded from further investigation.

Statistical analysis

Statistical significance of the associations between selected parameters of response to dasatinib therapy and genetic variants was evaluated using R v2.11 software. Logistic regression was used to assess the association between the genetic variability of the SNPs and treatment efficacy, defined by the following endpoints: CCyR at 12 months, MMR at 18 months. Treatment toxicity was investigated with logistic regression as well, using the following endpoints: appearance of any non-hematological toxicities of grades 2–4 including fluid retention, hematological toxicity of grade 3 or 4. These endpoints were used as dichotomous variables. To study the associations between SNPs and PFS and OS, Cox proportional hazard regression was used. SNPs were analyzed according to the following inheritance models: "co-dominant," where the homozygous major allele (reference category) was compared separately with two different genotypes that include the minor allele (heterozygotes and homozygotes for the variant allele), and "recessive," whereby the comparison groups were minor homozygous genotypes against the rest (combining heterozygotes and homozygotes for the major allele). The most significant test between the codominant and the recessive genetic models was used to determine the statistical significance of each association of each SNP. All analyses were adjusted by age at diagnosis, sex, CML phase, Sokal score, and use of dasatinib in the second or third line. The results of logistic regression analyses were expressed as odds ratios (OR), and the results of Cox regression analyses were expressed as hazard ratios (HR), with 95% confidence intervals (CI). The results were adjusted for multiple comparisons by Bonferroni correction which was calculated as $p = 0.05/(61 \text{ SNPs} \times 2 \text{ models}) = 0.00041$.

Functional characterization of the SNPs

For the SNPs that achieved statistical significance, several online databases were used for analysis of the influence of SNPs on cellular processes such as transcription or association with protein activity.

RegulomeDB is a database that assigns SNPs to predicted and known regulatory sites in the human genome regions that affect transcriptional processes, DNAase hypersensitivity regions, transcription factor binding sites, and promoter regions.

HaploReg is a tool designed to search for significant effects of SNPs in haplotype blocks, including SNPs in disease-related loci, on gene expression levels and protein activity. Using HaploReg, it is possible to assess whether specific SNPs are located in the eQTL (expression quantitative trait loci) loci and to check the effect of SNPs on gene regulatory regions. GTEx is a database that allows to search for relationships between genetic variation (including SNPs) and the expression of genes and proteins in individual tissues.

Results

Clinical data

Eighty-six Polish Caucasian CML patients treated in five tertiary hematological centers in Poland were genotyped within the present study. The characteristics of the study population are described in Table 1. The majority of patients was diagnosed in the chronic phase (85%). Only 6% of the patients carried additional cytogenetic aberrations; the remainder were characterized only by the presence of the Ph chromosome. Regarding Sokal score, 43% of patients were diagnosed as low risk, and 31% and 26% were characterized as intermediate and high risk, respectively.

Eighteen patients (21%) discontinued imatinib due to intolerance to the treatment, 68 patients (79%) due to disease progression. The median duration of imatinib treatment in these patients was 637 days (1.75 years), while the median duration of dasatinib treatment, calculated as the time between the initiation of therapy and the last medical check-up, was 1,492 days (4.09 years). The overall survival and PFS of the cohort is showed in Figure 1.

All patients received imatinib as first-line therapy of CML in the standard dose of 400 mg once daily and were later treated

TABLE 1 Summary of the dasatinib treatment results.

Dasatinib treatment	***Number of patients (%)***
All	86 (100%)
In the 2nd line	78 (90.7%)
In the 3rd line	8 (9.3%)
Response to treatment[a]	
Optimal CyR[b]	52 (60.5%)
CCyR	49 (57%)
PCyR	9 (11%)
mCyR	21 (24%)
noCyR	7 (8%)
Optimal MR[c]	8 (9.3%)
CMR	24 (30%)
MMR	19 (24%)
No MR	37(46%)
Progression	14 (16.3%)
Fluid retention	15 (17.4%)
Incidence of any hematological complications	33 (38.4%)

[a]Response to treatment: in the database were included results of the last medical visit for each patient.
[b]Optimal CyR: achievement of the complete cytogenetic response within 12 months from treatment start.
[c]Optimal MR: achievement of at least the major molecular response within 18 months from treatment start. Only eight patients (9.3%) achieved optimal response, i.e., at least MR3 within 18 months from the start of treatment; other patients achieved CMR and MMR but after 18 months of treatment, therefore they were not included as optimal MR. Data on the cytogenetic and molecular response as well as the occurrence of treatment side effects are included.

with dasatinib at 100 mg/day due to the ineffectiveness or intolerance of imatinib. Patients treated with dasatinib right after imatinib treatment (in the second line) accounted for 91% (78 patients) of the group, while eight patients (9%) received dasatinib in the third line, after nilotinib was given for imatinib failure or toxicity.

Sixty-five patients (76%) had no changes in dasatinib dosage while in five cases (6%) the dose was subsequently increased to 140 mg/day at the physician's discretion and 16 (18%) patients had the dose decreased to 80 mg/day due to toxicity. The main complications that occurred during the follow-up in the present analysis were fluid retention (15 patients, 17%) and hematological complications including thrombocytopenia, cytopenia, neutropenia, lymphopenia, and agranulocytosis (33 patients, 38%). Other undesirable effects included pulmonary hypertension, abdominal pain, arrhythmias, and increased creatine kinase level.

Genotyping quality control

Genotyping of 61 preselected SNPs for *ABCB1*, *ABCG2*, *CAR*, and *PXR* was successful in all included patients. The average call rate was 98.42% ranging from 85.25% to 100%. Genotype distributions were in accordance with the Hardy–Weinberg equilibrium for all tested loci.

Influence of SNPs on cytogenetic and molecular response to dasatinib

The analysis using a logistic regression model identified the significant impact of tested SNPs on the probability of achieving CCyR following dasatinib treatment. Assuming the codominant model of inheritance, noteworthy associations ($p < 0.05$) were found between the achievement of CCyR after 12 months of therapy and the following SNPs: *ABCB1* gene: rs7787082, *PXR* rs2461818 and two SNPs in the *ABCG2* gene: rs12505410 and rs3109823 (Table 2). Assuming a recessive inheritance model, statistical significance was observed for two SNPs in the *ABCG2* gene: rs2622621 and rs3114018 (Table 2). No significant results were found regarding the potential influence of SNPs on the probability of the achievement of MMR after 18 months of dasatinib treatment.

Influence of SNPs on dasatinib therapy side effects

There was no significant association between tested clinical factors and the occurrence of analyzed side effects including grade 2–4 non-hematological complications or grade 3 and 4 neutropenia or thrombocytopenia. In contrast, a noteworthy correlation was observed between rs7787082, in the *ABCB1*, and the occurrence of grade 3 or 4 hematological complications assuming the codominant inheritance model. Furthermore, a significant association was found between rs2725256 in *ABCG2* and occurrence of non-hematological complications assuming a recessive model of inheritance (Table 2).

Influence of SNPs on OS and PFS on dasatinib therapy

We found a statistically significant influence of Sokal score on OS (HR = 1.34, 95% CI = 1.21-2.09, p = 0.001). No other clinical pretreatment parameter was related to patients' survival functions. Interestingly, assuming the codominant inheritance model, there was a significant correlation between four SNPs (*ABCB1* rs3842, *ABCB1* rs2235023, *CAR* rs11265571, *PXR*

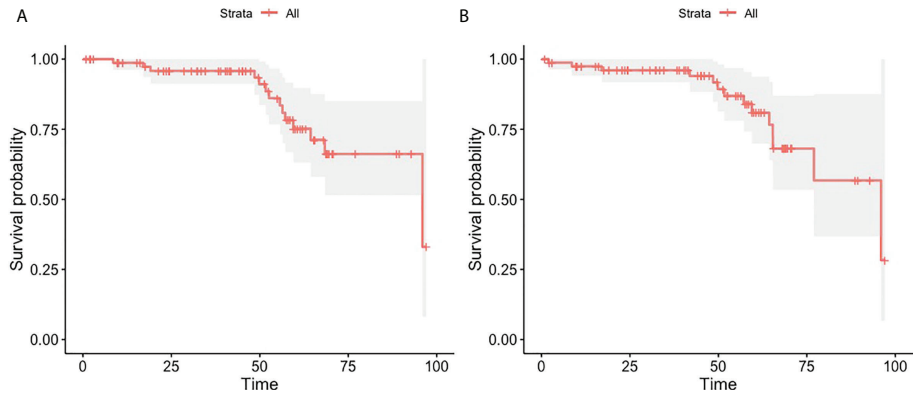


FIGURE 1
The Kaplan-Meier plots showing overall survival **(A)** and progression- free survival **(B)** in the whole cohort of 86 CML patients treated with dasatinib.

TABLE 2 Associations between selected SNPs and cytogenetic response after 12 months of treatment as well as incidence of hematological and non-hematological complications.

SNP	Gene	Alleles M/m[a]	Endpoint[b]	Model[c]	OR	95% CI	p-value
rs7787082	*ABCB1*	G/A	CyR-12	CD	0.20	0.06-0.66	0.008
rs2461818	*PXR*	C/T	CyR-12	CD	0.16	0.03-0.89	0.036
rs12505410	*ABCG2*	T/G	CyR-12	CD	3.82	1.38-10.55	0.010
rs3109823	*ABCG2*	T/C	CyR-12	CD	2.87	1.11-7.40	0.029
rs2622621	*ABCG2*	C/G	CyR-12	R	0.21	0.05-0.92	0.038
rs3114018	*ABCG2*	A/C	CyR-12	R	0.24	0.08-0.71	0.010
rs7787082	*ABCB1*	G/A	HC	CD	4.46	1.38-14.39	0.012
rs2725256	*ABCG2*	A/G	NHC	R	4.71	1.20-18.47	0.026

[a]M, major allele; m, minor allele.
[b]CyR-12, achievement of cytogenetic response within 12 months from treatment start; HC, hematological complications; NHC, non-hematological complications.
[c]CD, codominant; R, recessive.

rs3732360) and OS in the patients treated with dasatinib. Furthermore, rs2307418 in the *CAR* gene impacted OS assuming a recessive inheritance model.

Moreover, seven tested SNPs significantly influenced the probability of PFS. Assuming the codominant model of inheritance, these were the following SNPs: *CAR* rs2307418 *PXR* rs3732357, *ABCB1* rs2235023 *ABCB1* rs22114102 *PXR* rs3732360, and *PXR* rs11917714 *PXR* rs3732359 (Table 3 and Figure 2).

Functional SNP annotation

Table 4 shows a summary of the potential functional impact of the tested SNPs based on HaploReg, RegulomeDB, and GTEx databases. The table includes SNPs that showed statistical importance in this analysis.

The RegulomeDB portal achieving scores 1–3 indicates the probability of the analyzed SNP belonging to the sequence affecting the binding process of transcription factors. All tested SNPs scored 4, 5, and 6, which means that the binding of transcription factors is very unlikely.

Analysis with HaploReg showed the probable influence of SNPs on the regulation of genes. Noteworthy is the influence of some studied SNPs on the family of FOX transcription factors, which affect a number of cellular processes. All details are included in Table 4.

In the present work using the GTEx portal, a significant link with the gene expression level was demonstrated for most of the candidate SNPs tested. However, only for SNPs rs12505410, rs2622621, rs3109823 in *ABCG2* were the associations in blood cells, while the rest were in tissues not related to CML pathogenesis.

TABLE 3 Associations between selected SNPs, overall survival, and progression-free survival during dasatinib treatment.

SNP	Gene	Alleles M/m[a]	Endpoint[b]	Model[c]	HR	95%CI	p value
rs3842	*ABCB1*	T/C	OS	CD	1.84	1.01-3.33	0.012
rs2235023	*ABCB1*	C/T	OS	CD	2.28	1.03-5.02	0.027
rs11265571	*CAR*	A/T	OS	CD	1.59	0.82-3.08	0.037
rs3732360	*PXR*	T/C	OS	CD	0.64	0.40-1.04	0.049
rs2307418	*CAR*	T/G	OS	R	73.68	4.47-1215.31	0.003
rs2307418	*CAR*	T/G	PFS	CD	2.02	1.19-3.43	0.048
rs3732357	*PXR*	A/G	PFS	CD	0.42	0.26-0.70	0.001
rs2235023	*ABCB1*	C/T	PFS	CD	2.49	1.13-5.50	0.011
rs22114102	*ABCB1*	C/T	PFS	CD	1.90	1.00-3.63	0.028
rs3732360	*PXR*	T/C	PFS	CD	0.59	0.38-0.93	0.020
rs11917714	*PXR*	C/T	PFS	CD	0.58	0.36-0.92	0.020
rs3732359	*PXR*	A/G	PFS	CD	0.57	0.36-0.91	0.024

[a]M, major allele; m, minor allele.
[b]OS, overall survival; PFS, progression-free survival.
[c]CD, codominant; R, recessive.

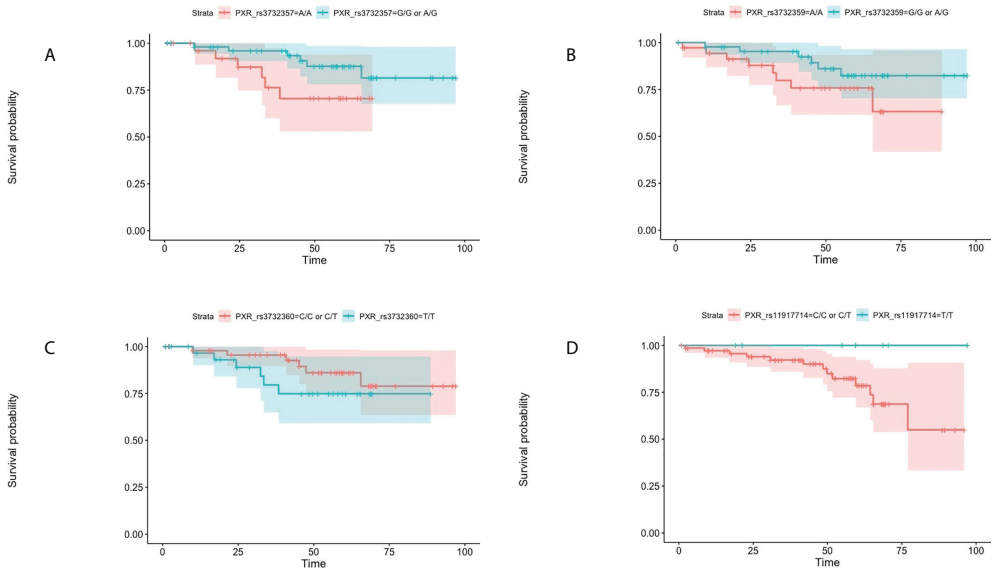


FIGURE 2
Kaplan-Meier plots for progression- free survival in regard to the presence of rs3732357 **(A)**, rs3732359 **(B)**, rs3732360 **(C)**, and rs11917714 **(D)** in the PXR gene in the co-dominant inheritance model.

Discussion

We used a candidate gene approach to evaluate the impact of inherited genetic differences on the outcome of CML treatment with dasatinib. The *ABCB1* and *ABCG2* genes, encoding for known transporters of dasatinib, and *PXR* and *CAR* xenosensor genes, with the role of transcription factors for many genes which take part in the pharmacokinetic processes, were chosen for analysis.

A tagging approach was applied to capture the common genetic variability of *ABCB1*, *ABCG2*, *PXR*, and *CAR* and resulted in a selection of a total of 61 tagging SNPs that were subsequently genotyped in 86 CML patients treated with dasatinib in the second and third lines of treatment.

Significant amount of data has been reported regarding the influence of SNPs in drug transport and metabolism genes on imatinib, another TKI. Kim et al. investigated the influence of SNPs in genes potentially involved in metabolism of imatinib. They found that the rs2231137 GG homozygotes (*ABCG2*), rs776746 AA (*CYP3A5*) homozygotes, and advanced stage strongly correlated with poor response to treatment with imatinib; however, the *SLC22A1*-rs683369 GG homozygotes and advanced stage were correlated with therapy failure (23). Seong et al. analyzed the impact of SNPs in cytochrome P450 enzymes and drug transporters on the imatinib concentration in plasma and clinical response in CML patients. They concluded that rs2231142 (421C>A) situated in *ABCG2* is highly associated with MMR achieved by CML patients (24). In contrast, Takahashi et al. showed that homozygotes AA in the same SNP had a higher imatinib concentration than CC (25). In our study, rs2231142 (421C>A) in *ABCG2* showed no statistically significant association on dasatinib treatment endpoints.

However, to the best of our knowledge there are little data available regarding influence on inherited background in ABC transporters on dasatinib therapy in CML. Skoglund et al. found that wild-type ABCG2 had a protective effect against the cytotoxicity of all investigated tyrosine kinase inhibitors in exception of bosutinib. Skoglund et al.'s finding of SNPs *ABCG2* 421C>A, 623T>C, 886G>C, and 1574T>G showed a reduction in *ABCG2* cell membrane expression and the protective effect of *ABCG2* against imatinib, CGP74588, dasatinib, and nilotinib cytotoxicity (26). In our study, rs2231142 (421C>A) showed no significance regarding dasatinib treatment endpoints.

To our knowledge, there are no publications available on the effect of single-nucleotide polymorphisms in the *PXR* and *CAR* nuclear receptor genes on the outcomes of dasatinib treatment in CML patients. However, based on the results of our work, it is assumed that such an influence exists. Analysis of SNPs showed that rs2461818 in *PXR* has an influence on achieving clinical outcome expressed through CCyR after 12 months (OR = 0.16; 95% CI = 0.03-0.89; p = 0.036). The CAR and PXR proteins belonging to the same family of nuclear receptors are known as transcription factors for genes involved in the metabolism of exogenous substances and their removal from the body (13). As CAR stimulates the expression of proteins related to imatinib metabolism ABCB1, ABCG2, ABCC2, hOCT1, and CYP3A4 (13, 27, 28), genetically influenced changes in the activity of this

TABLE 4 Summary of bioinformatic SNP annotations.

SNP ID	Gene	Alleles M/m[a]	MAF[b]	Rank[c]	HaploReg[d]	eQTL[e]	GTEx[f]
rs7787082	*ABCB1*	G/A	0.18	6	*CRX FOXD3 GFI1*	–	Testes, skin
rs3842	*ABCB1*	T/C	0.14	n/a	*PLZF*	*ABCB4*	Brain, nerves
rs2235023	*ABCB1*	C/T	0.09	6	*CEBPB ISL2 POU1F1 RHOX11*	*ABCB1 ABCB4*	Testes, muscles
rs2214102	*ABCB1*	C/T	0.09	4	*GR*	–	Heart, colon
rs2725256	*ABCG2*	A/G	0.33	6	*FOXA HNF4 HBP1 POU1F1 RXRA STAT*	*SPP1*	Adipocytes
rs12505410	*ABCG2*	T/G	0.40	4	*PBX3*	–	Testes, blood
rs2622621	*ABCG2*	C/G	0.30	5	-	–	Blood
rs3114018	*ABCG2*	A/C	0.49	5	*LHX3 MEF2 NANOG POU2F2*	–	Esophagus, heart
rs3109823	*ABCG2*	T/C	0.26	6	*CTCF NRF-2 YY1*	–	Blood
rs2307418	*CAR*	T/G	0.15	5	*ERalpha-A GCMGR*	*TOMM40L USF1*	Brain, muscles, skin
rs11265571	*CAR*	A/T	0.17	4	*ERalpha-A PAX-5 RHOX11*		Skin, testes, colon
rs2461818	*PXR*	C/T	0.08	6	*ARID5B FOXO2 FOXO3 FOXP1 HDAC2 IK2 PLZF SOX6*	*PLA1A GSK3B*	Thyroid
rs11917714	*PXR*	C/T	0.17	6	*HIC1*	*GSK3B*	Nerves, small intestine, esophagus, testes
rs3732357	*PXR*	A/G	0.35	5	*GLI GLIS2 ZIC*	GSK3B	Adipocytes, colon, esophagus, stomach, arteries, thyroid, brain, muscles, lungs
rs3732360	*PXR*	T/C	0.23	6	*E2A LMO2 MYF TATA*	*GSK3B GPR156*	Adipocytes, nerves, esophagus, brain, colon, arteries
rs3732359	*PXR*	A/G	0.20	6	*BCL NRSF PLAG1 SIN3AK20 TAL1 YY1*	*GSK3B GPR156*	Adipocytes, nerves, esophagus, brain, colon, arteries, thyroid

The table contains associations for all SNPs, while those that occur only in blood cells are bold.
[a]M: major allele; m: minor allele.
[b]MAF: minor allele frequency in the 1000 Genomes European population.
[c]Rank from RegulomeDB: 1 is given to SNPs showing the strongest evidence of a role in regulating the transcription process by binding transcription factors, while 6 to SNPs with a low probability of influencing to transcription.
[d]HaploReg: the tested SNP probably influences the expression of mentioned genes.
[e]eQTL: the tested SNP is located in the eQTL (expression quantitative trait loci) of mentioned genes.
[f]GTEx: the relationship between the tested SNP and the tissue in which the gene is expressed.

gene product may have an indirect effect on the plasma concentration of imatinib and thus potentially also the effects of TKI BCR-ABL1 treatment. Although dasatinib belongs to the same group of drugs, its molecular structure and metabolism differ from that of imatinib, which may result in a different response to treatment in the context of the same polymorphic changes, as evidenced by the obtained results.

Loscocco et al., in their study, examined SNPs in genes from the ABC family (*ABCB1*, *ABCG2*, *ABCC1*, *ABCC2*) to determine their effectiveness in treatment with another tyrosine kinase inhibitor—nilotinib—in a group of 90 CML patients. They found that CC and CT genotypes in *ABCC2* rs3740066 as well as the TT genotypes in *ABCB1* rs1045642 correlated with a higher probability of achieving MR3 in a shorter time ($p = 0.02$, $p = 0.004$, and $p = 0.01$), where the GG genotype of *ABCG2* rs2231137 was associated with a lower probability of MR3 achievement ($p = 0.005$). Moreover, the *ABCC2* rs3740066 CC genotype and the *ABCB1* rs1045642 CC and TT genotypes were positively correlated with MR4 achievement ($p = 0.02$, $p = 0.007$, and $p = 0.003$) (29). Our study does not reveal any correlations between examined SNPs and molecular response in patients treated with dasatinib.

Our results suggest that the naturally occurring germline variation in tested genes has influence on such important endpoints of dasatinib therapy in CML as well as probability of cytogenetic and molecular responses, PFS and OS.

To the best of our knowledge, this is the first study on the impact of *ABCB1*, *ABCG2*, *PXR*, and *CAR* gene polymorphisms on the outcomes of dasatinib treatment in CML. The work clearly shows the influence of SNPs present in genes related to dasatinib metabolism (*ABCB1* and *ABCG2*) and genes encoding transcription factors (*PXR*, *CAR*) on treatment outcomes, susceptibility to side effects, and overall survival and progression-free time. Furthermore, the analysis we present here has a number of strengths. First of all, cases were collected in a relatively small number of hematological centers with high

medical reference. Information about multiple clinical endpoint variables of patients has been thoroughly compiled in a single database. For each patient, detailed clinical history and treatment history were checked. The collected data showed that the whole group was homogenous, especially in terms of lack of a good response to imatinib. To the best of our knowledge, this is the first study on the impact of *ABCB1*, *ABCG2*, *PXR*, and *CAR* gene polymorphisms on the outcomes of dasatinib treatment in CML.

The main weakness of this study is the reduced sample size, which limits the statistical power. Taking into account the large number of SNPs included in the study and the different analysis models, the Bonferroni-corrected threshold for statistical significance was rather stringent ($p = 0.00041$). None of the associations reported here were statistically significant if this threshold is used. However, it has to be kept in mind that CML is not a common disease and is also often treated with imatinib in the first line. Approximately one-third of cases are switched to the next-generation inhibitors (dasatinib, nilotinib, or others) due to lack of response or intolerance to this treatment regimen. We would like to emphasize that the population that participated in this study were patients treated in the first line with imatinib, and in the second (or third) line with dasatinib. Therefore, the above-presented results cannot be directly translated into the population of patients treated in the first line with dasatinib, because the studied population consisted of people who showed intolerance or progression during treatment, which may be of primary origin, regardless of the treatment used.

Another factor that may be a weakness is the fact that only a group of patients of Polish origin was taken into account in the analysis. Therefore, the results cannot be used in comparison to other ethnic groups (e.g., from Asia), whose genetic variability of the tested SNPs and frequency of their occurrence may differ significantly. To confirm this, additional analysis on different populations should be done.

Taking all into account, there is still little known about the impact of inherited changes in genes involved in pharmacokinetic processes in the case of TKI treatment of CML. Our work is one of the first to describe the influence of *ABCB1*, *ABCG2*, *PXR*, and *CAR* polymorphisms on the effects of dasatinib therapy (expressed through CCyR, PFS, OS) and toxicity of the therapy (expressed by association with hematological and non-hematological complications). Further analyses are needed to confirm these initial results.

Data availability statement

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

Ethics statement

This study was reviewed and approved by Ethical Committee of Medical University of Lodz (RNN/168/13/KE; 18 June 2013) Łódź, gen. Józefa Hallera Square 1B bioetyka@umed.lodz.pl. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KJ conceived and designed the study. AM performed the lab work and prepared the database. FC performed the statistical analysis. AM analyzed the results and drafted the manuscript. FC and KJ reviewed and edited the manuscript. All other authors provided samples and data. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the grants of Young Hematologists Club of Polish Society of Hematology and Transfusiology Medicine.

Acknowledgments

The authors would like to express their deep gratitude to Dr. Alessandro Martino for his involvement in the project of evaluation polymorphisms in chronic myeloid leukemia. The authors would also like to extend their thanks to Angelika Stein, technician at the DKFZ Genomic Epidemiology Group, for her patience and help.

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. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, et al. SEER cancer statistics review, 1975-2009 (Vintage 2009 populations). *Natl Cancer Inst* (2012).

2. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* (1960) 25:85–109.

3. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature* (1973) 243(5405):290–3. doi: 10.1038/243290a0

4. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, et al. Effects of a selective inhibitor of the abl tyrosine kinase on the growth of bcr-abl positive cells. *Nat Med* (1996) 2(5):561–6. doi: 10.1038/nm0596-561

5. Hochhaus A, Hughes T. Clinical resistance to imatinib: mechanisms and implications. *Hematol Oncol Clin North Am* (2004) 18:641–56. doi: 10.1016/j.hoc.2004.03.001

6. Lombardo LJ, Lee FY, Chen P, Norris D, Barrish JC, Behnia K, et al. Discovery of n-(2-chloro-6-methyl- phenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino) thiazole-5 carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. *J Med Chem* (2004) 47(27):6658–61. doi: 10.1021/jm049486a

7. Deremer DL, Ustun C, Natarajan K. Nilotinib: a second-generation tyrosine kinase inhibitor for the treatment of chronic myelogenous leukemia. *Clin Ther* (2008) 30(11):1956–75. doi: 10.1016/j.clinthera.2008.11.014

8. Muller MC, Cortes J, Kim D-W, Druker BJ, Erben P, Pasquini R, et al. Dasatinib efficacy in patients with chronic myeloid leukemia in chronic phase and preexisting BCR-ABL mutations (abstract 449). *Blood* (2008) 112(11):171.

9. Manley PW, Cowan-Jacob SW, Mestan J. Advances in the structural biology, design and clinical development of bcr-abl kinase inhibitors for the treatment of chronic myeloid leukaemia. *Biochim Biophys Acta* (2005) 1754:3–13. doi: 10.1016/j.bbapap.2005.07.040

10. Kamath AV, Wang J, Lee FY, Marathe PH. Preclinical pharmacokinetics and *in vitro* metabolism of dasatinib (BMS-354825): A potent oral multi-targeted kinase inhibitor against SRC and BCR-ABL. *Cancer Chemother Pharmacol* (2008) 61(3):365–76. doi: 10.1007/s00280-007-0478-8

11. Sharom FJ. ABC Multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* (2008) 9(1):105–27. doi: 10.2217/14622416.9.1.105

12. Liddle C, Goodwin B. Regulation of hepatic drug metabolism: role of the nuclear receptors PXR and CAR. *Semin Liver Dis* (2002) 22(2):115–22. doi: 10.1055/s-2002-30098

13. Maglich JM, Stoltz CM, Goodwin B, Hawkins-Brown D, Moore JT, Kliewer SA, et al. Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Mol Pharmacol* (2002) 62(3):638–46. doi: 10.1124/mol.62.3.638

14. Geick A, Eichelbaum M, Burk O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* (2001) 276 (18):14581–7. doi: 10.1074/jbc.M010173200

15. Synold TW, Dussault I, Forman BM. The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat Med* (2001) 7(5):584–90. doi: 10.1038/87912

16. Zazuli Z, Barliana MI, Mulyani UA, Perwitasari DA, Ng H, Abdulah R, et al. Polymorphism of PXR gene associated with the increased risk of drug-induced liver injury in Indonesian pulmonary tuberculosis patients. *J Clin Pharm Ther* (2015) 40 (6):680–4. doi: 10.1111/jcpt.12325

17. Moon JY, Chang BC, Lee KE, Bang JS, Gwak HS. Effects of pregnane X receptor genetic polymorphisms on stable warfarin doses. *J Cardiovasc Pharmacol Ther* (2015) 20(6):532–8. doi: 10.1177/1074248415578906

18. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe against cancer program. *Leukemia* (2003) 17(12):2318–57. doi: 10.1038/sj.leu.2403135

19. Hochhaus A, Baccarani M, Silver RT, Schiffer C, Apperley JF, Cervantes F, et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia* (2020) 34(4):966–84. doi: 10.1038/s41375-020-0776-2

20. Available at: http://www.broad.mit.edu/mpg/tagger/http://www.broad.mit.edu/mpg/haploview/.

21. Available at: https://pubmed.ncbi.nlm.nih.gov/14685227/.

22. Campa D, Butterbach K, Slager SL, Skibola CF, de Sanjosé S, Benavente Y, et al. A comprehensive study of polymorphisms in ABCB1, ABCC2 and ABCG2 and lung cancer chemotherapy response and prognosis. *Int J Cancer* (2012) 131 (12):2920–8. doi: 10.1002/ijc.27567

23. Kim DH, Sriharsha L, Xu W, Kamel-Reid S, Liu X, Siminovitch K, et al. Clinical relevance of a pharmacogenetic approach using multiple candidate genes to predict response and resistance to imatinib therapy in chronic myeloid leukemia. *Clin Cancer Res* (2009) 15(14):4750–8. doi: 10.1158/1078-0432.CCR-09-0145

24. Seong SJ, Lim M, Sohn SK, Moon JH, Oh SJ, Kim BS, et al. Influence of enzyme and transporter polymorphisms on trough imatinib concentration and clinical response in chronic myeloid leukemia patients. *Ann Oncol* (2013) 24 (3):756–60. doi: 10.1093/annonc/mds532

25. Takahashi N, Miura M, Scott SA, Kagaya H, Kameoka Y, Tagawa H, et al. Influence of CYP3A5 and drug transporter polymorphisms on imatinib trough concentration and clinical response among patients with chronic phase chronic myeloid leukemia. *J Hum Genet* (2010) 55(11):731–7. doi: 10.1038/jhg.2010.98

26. Skoglund K, Boiso Moreno S, Jönsson JI, Vikingsson S, Carlsson B, Gréen H, et al. Single-nucleotide polymorphisms of ABCG2 increase the efficacy of tyrosine kinase inhibitors in the K562 chronic myeloid leukemia cell line. *Pharmacogenet Genomics* (2014) 24(1):52–61. doi: 10.1097/FPC.0000000000000022

27. Kast HR, Goodwin B, Tarr PT, Jones SA, Anisfeld AM, Stoltz CM, et al. Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J Biol Chem* (2002) 277(4):2908–15. doi: 10.1074/jbc.M109326200

28. Sueyoshi T, Negishi M. Phenobarbital response elements of cytochrome P450 genes and nuclear receptors. *Annu Rev Pharmacol Toxicol* (2001) 41:123–43. doi: 10.1146/annurev.pharmtox.41.1.123

29. Loscocco F, Visani G, Ruzzo A, Bagaloni I, Fuligni F, Galimberti S, et al. Clinical relevance of ABCB1, ABCG2, and ABCC2 gene polymorphisms in chronic myeloid leukemia patients treated with nilotinib. *Front Oncol* (2021) 11:672287. doi: 10.3389/fonc.2021.672287

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fonc.2022.969238

OPEN ACCESS

EDITED BY
Alexandre How-Kit,
Fondation Jean Dausset Centre d'Etude du Polymorphisme Humain, France

REVIEWED BY
Beifang Niu,
Computer Network Information Center (CAS), China
Richard Gallon,
Newcastle University, United Kingdom

*CORRESPONDENCE
Zhenyu Xu
zxu@sophiagenetics.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 14 June 2022
ACCEPTED 18 October 2022
PUBLISHED 17 November 2022

CITATION
Marques AC, Ferraro-Peyret C, Michaud F, Song L, Smith E, Fabre G, Willig A, Wong MML, Xing X, Chong C, Brayer M, Fenouil T, Hervieu V, Bancel B, Devouassoux M, Balme B, Meyronet D, Menu P, Lopez J and Xu Z (2022) Improved NGS-based detection of microsatellite instability using tumor-only data.
Front. Oncol. 12:969238.
doi: 10.3389/fonc.2022.969238

Improved NGS-based detection of microsatellite instability using tumor-only data

Ana Claudia Marques[1†], Carole Ferraro-Peyret[2,3†], Frederic Michaud[1†], Lin Song[1†], Ewan Smith[1], Guillaume Fabre[1], Adrian Willig[1], Melissa M. L. Wong[1], Xiaobin Xing[1], Chloe Chong[1], Marion Brayer[1], Tanguy Fenouil[3], Valérie Hervieu[3], Brigitte Bancel[3], Mojgan Devouassoux[4], Brigitte Balme[4], David Meyronet[3], Philippe Menu[1], Jonathan Lopez[2,5] and Zhenyu Xu[1*]

[1]SOPHiA GENETICS, Saint-Sulpice, Switzerland, [2]Cancer Research Centre of Lyon, INSERM 1052, Centre National de la Recherche Scientifique (CNRS) 5286, University of Lyon, Lyon, France, [3]Hospices Civils de Lyon, Biopathology of Tumours, GH Est (GHE) Hospital, Bron, France, [4]Hospices Civils de Lyon, Department of Anatomopathology, Lyon-Sud Hospital, Lyon, France, [5]Hospices Civils de Lyon, Biochemistry and Molecular Biology Department, Lyon-Sud Hospital, Lyon, France

Microsatellite instability (MSI) is a molecular signature of mismatch repair deficiency (dMMR), a predictive marker of immune checkpoint inhibitor therapy response. Despite its recognized pan-cancer value, most methods only support detection of this signature in colorectal cancer. In addition to the tissue-specific differences that impact the sensitivity of MSI detection in other tissues, the performance of most methods is also affected by patient ethnicity, tumor content, and other sample-specific properties. These limitations are particularly important when only tumor samples are available and restrict the performance and adoption of MSI testing. Here we introduce MSIdetect, a novel solution for NGS-based MSI detection. MSIdetect models the impact of indel burden and tumor content on read coverage at a set of homopolymer regions that we found are minimally impacted by sample-specific factors. We validated MSIdetect in 139 Formalin-Fixed Paraffin-Embedded (FFPE) clinical samples from colorectal and endometrial cancer as well as other more challenging tumor types, such as glioma or sebaceous adenoma or carcinoma. Based on analysis of these samples, MSIdetect displays 100% specificity and 96.3% sensitivity. Limit of detection analysis supports that MSIdetect is sensitive even in samples with relatively low tumor content and limited microsatellite instability. Finally, the results obtained using MSIdetect in tumor-only data correlate well (R=0.988) with what is obtained using tumor-normal matched pairs, demonstrating that the solution addresses the challenges posed by MSI detection from tumor-only data. The accuracy of

10.3389/fonc.2022.969238

MSI detection by MSIdetect in different cancer types coupled with the flexibility afforded by NGS-based testing will support the adoption of MSI testing in the clinical setting and increase the number of patients identified that are likely to benefit from immune checkpoint inhibitor therapy.

KEYWORDS

microsatellite, next-generating sequencing, tumor-only sequencing, pan-cancer, MSI, Mismatch Repair deficiency, Microsatellite instability

Introduction

The DNA mismatch repair (MMR) pathway safeguards the genome from base substitution and insertion-deletion (indels) during DNA replication (1). Genetic or epigenetic loss of one or more of the involved proteins results in MMR deficiency (dMMR), leading to increased mutation rates (2).

dMMR is a predictive pan-cancer marker of response to immune checkpoint inhibitor therapy (3, 4) (5). The current standard of dMMR testing is evaluating the expression of the four MMR proteins by immunohistochemistry (IHC) (6). However, IHC tests cannot be combined with other molecular diagnostics, limiting its adoption in cancer types where this molecular phenotype is rare, and false-positive and negative immunostaining results impact their accuracy. Detection of microsatellite instability (MSI), a well-established signature of dMMR (2), is a suitable alternative to IHC (6). Microsatellites (1-6 nucleotide tandem repeat motifs) are informative for dMMR status since their contraction or expansion, resulting from DNA replication errors, are normally repaired by the MMR pathway (7).

In the clinical setting, the most used method to evaluate MSI status analysis of allelic size variation in a panel of five mononucleotide repeats (homopolymers) (6, 8) is using polymerase chain reaction (PCR) followed by capillary electrophoresis. However and despite its widespread use, the analytical performance of this solution in cancers other than colorectal cancer, for which the solution was designed for (8), is relatively low (9, 10). The relatively small number of loci that can be simultaneously analyzed by PCR-based methods limits the opportunities to account for tissue of origin and other sample-specific factors. In addition, common population polymorphisms within homopolymers can reduce the sensitivity of PCR-based MSI detection methods, especially when matched normal samples are unavailable (11, 12).

Next-Generation Sequencing (NGS) based MSI detection allows the simultaneous analysis of a larger number of microsatellite regions, thus limiting the impact of sample-specific factors, including tissue of origin or population-specific variation in microsatellite length (13). In addition, NGS-based MSI analysis can be combined with other cancer-related molecular signatures and genetic lesions, facilitating the adoption of MSI clinical testing and increasing the number of patients considered for immunotherapy (14). Indeed NGS-based methods that rely on analysis of paired tumor-normal samples support accurate MSI detection across multiple tumor types (15). However, this data type is not commonly available in the clinic. Whereas NGS-based methods that leverage information from tumor-only data would circumvent this challenge, inter- and intra-tumor specific differences in the frequency and position of MSI diagnostic events (16) (17, 18) still impact their accuracy (15). For example, many MSI events are private to one sample, and frequently occurring events can be tumor-type specific (16). Additionally, microsatellite regions are often polymorphic in healthy individuals, and their sequence differs across the human population (7, 19). All these factors limit the analytical performance of methods that rely on a baseline reference distribution to determine MSI status.

To address these limitations, we developed MSIdetect, a new MSI detection method. MSIdetect uses a curve fitting algorithm, thus accounting for the impact of tumor content and indel burden on homopolymer instability. To minimize the effect of intra- and inter- tumor-specific factors, we additionally restrict our analysis to a set of ~100 homopolymer regions that we found are minimally variable between tissues and individuals. Using a large cohort of clinical samples, we demonstrate that MSIdetect can sensitively detect MSI signatures from tumor-only data in various cancer types, even in samples with limited tumor content.

Results and discussion

NGS-based detection of MSI using Whole Exome Sequencing data

MMR deficiency (dMMR) results in microsatellite contraction and expansion. To optimize detection of this signature using NGS from tumor-only data, MSI detection

10.3389/fonc.2022.969238

solutions must account for the factors that can limit their sensitivity and specificity (Figure 1A). In NGS workflows, microsatellite instability is reflected by a difference, relative to a normal reference, in the distribution of read counts supporting different microsatellite lengths. MSIdetect relies on a curve-fitting algorithm (described in Materials and Methods section) that accounts for the impact of tumor heterogeneity and the indel burden on microsatellite length distribution (Figure 1B).

We used publicly available The Cancer Genome Atlas (TCGA) Whole Exome Sequencing (WES) data from 363 Colorectal Adenocarcinoma, 428 Stomach Adenocarcinoma and 492 Uterine Corpus Endometrial Carcinoma samples, with known MSI status (20), to investigate how different limiting factors (Figure 1A) might contribute to miscalls in our analytical workflows. Homopolymer length impact MSI detection by NGS in two ways. First, homopolymer length negatively correlates with the fraction of reads that span the

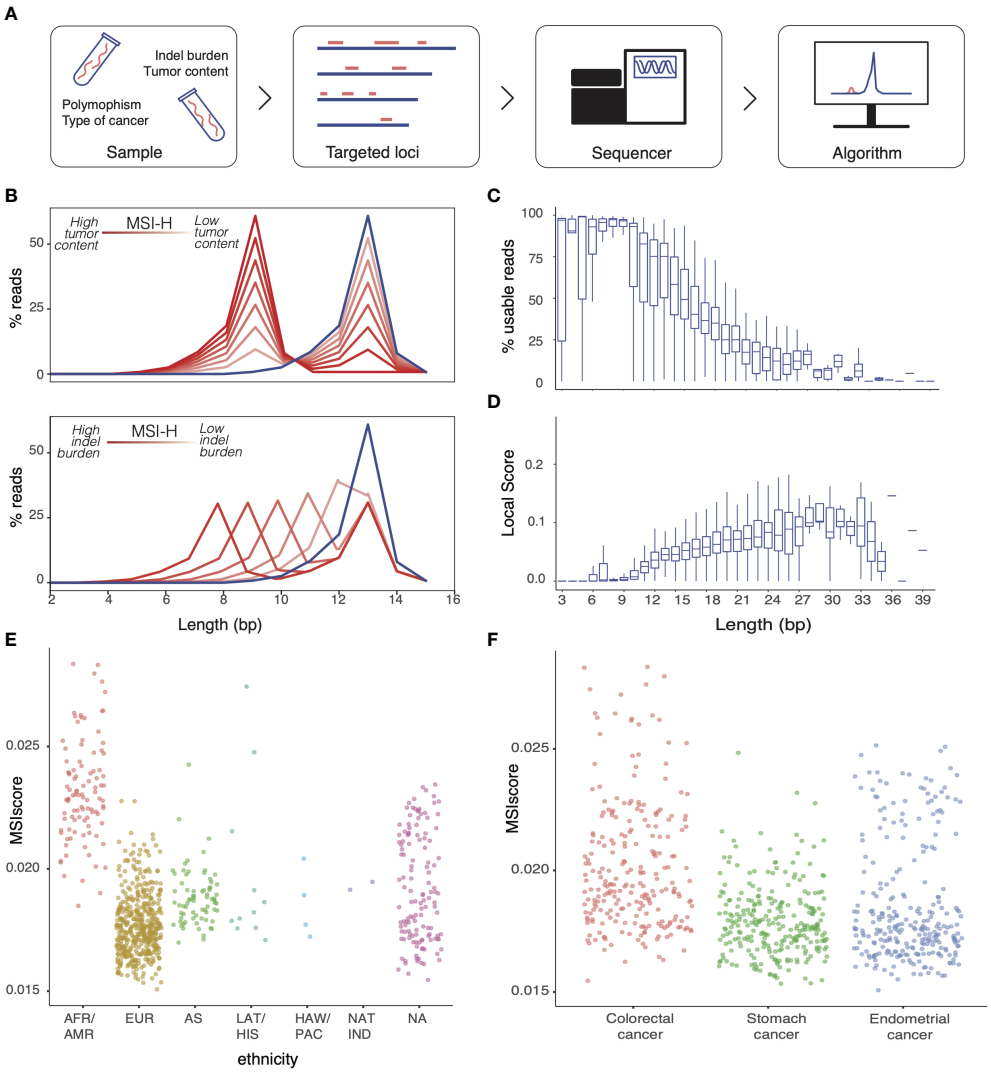


FIGURE 1
Factors limiting MSI detection in NGS workflow **(A)** Factors impacting detection of MSI in tumor-only NGS workflows **(B)** Schematic representation of the impact of increased indel burden (bottom panel) and tumor content (top panel) on the homopolymer length distribution measured by NGS at an illustrative homopolymer (MSI-H, red). Line color darkness correlates with decreased tumor content (top panel) or indel burden (bottom panel). Reference homopolymer length distribution for microsatellite stable is depicted in blue. Distribution of **(C)** Fraction of usable reads per total number of reads mapping to the homopolymer and **(D)** homopolymer score for homopolymers of the same length. MSI score obtained with MSIdetect using WES homopolymers for microsatellite stable (MSS) samples derived from **(E)** individuals of different ethnic origin and for samples from **(F)** different tumor types.

entirety of the region, and that can be used by the algorithm to infer the region's length stability (Figure 1C). In addition, the length distribution of relatively short homopolymers is very stable even in MSI-H samples, limiting their value to measure local instability (Figure 1D). These two factors are likely to define an optimal range of homopolymer length for MSI detection by NGS-based approaches.

In addition to indel burden and tumor content that is accounted for by the algorithm, other samples characteristics can also impact results. Specifically, homopolymers replication is error-prone (21), with MMR independent factors such as ethnicity (Figure 1E) or tissue origin (Figure 1F) impacting homopolymer length, as reflected by changes in MSIscore, in MSS samples.

Identification of homopolymers for optimal NGS-based detection of MSI

We computed the MSI score based on all homopolymers captured in the WES datasets (3602 loci (22), Supplementary Figure 1A) and assessed the concordance between MSIdetect results and pre-determined MSI status. We plotted the true-positive rate as a function of the false-negative rate obtained for the different tissues (Figure 2A). We found that MSIdetect results were highly concordant with MSI status (AUC>0.9926). When all homopolymers captured by the WES data set are considered, we observed tissue-specific differences in accuracy, with results being less accurate in Uterine Corpus Endometrial Carcinoma (AUC=0.9926), followed by

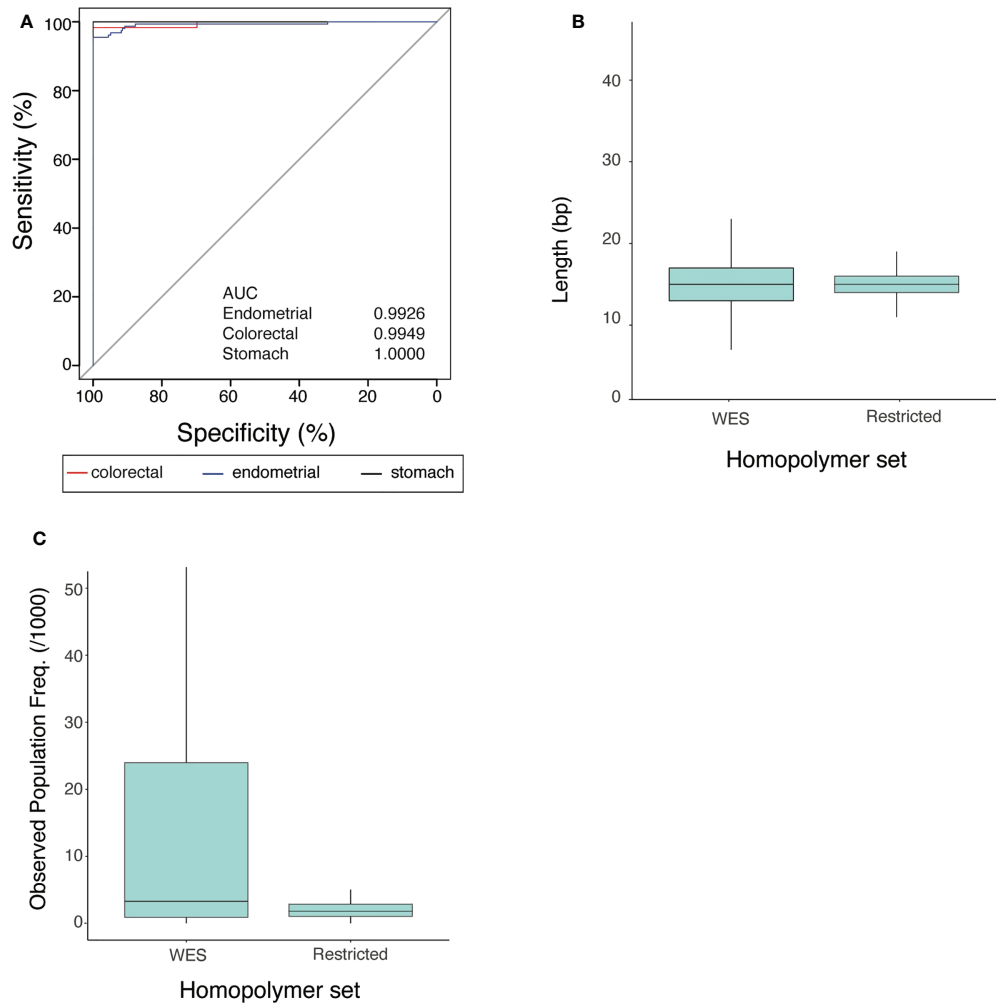


FIGURE 2
Properties of the MSIdetect restricted homopolymer set **(A)** Receiver Operating curves and corresponding Area Under the Curve (AUC) values (in the inset) for endometrial (blue), colorectal (red) and stomach (black) cancers for MSI classification by MSIdetect using WES homopolymers given the MSI status reported by TCGA. **(B)** Homopolymer length distribution in WES and in the restricted set. **(C)** Distribution of average variant population frequency observed in gnomAD for homopolymers in WES and in the restricted set with lengths ranging from 11-25 bp.

Colorectal Adenocarcinoma (AUC=0.9976) and Stomach Adenocarcinoma (AUC=1.000).

We compared the analytical performance of MSIdetect with that of two other widely used NGS-based MSI detection algorithms mSINGS (22) and MANTIS (15). These algorithms were chosen because, like MSIdetect, they rely on comparing microsatellite allele length distributions. Similar to MSIdetect, mSINGS (22) is compatible with tumor-only data, whereas MANTIS (15) relies on comparing the results obtained for a tumor sample with its matched normal sample. Like MSIdetect, the other algorithms are less accurate in endometrial cancer, followed by colorectal and stomach cancer (Supplementary Figures 1B-D). In all cancer types considered, MANTIS showed the highest overall performance with >97.4% sensitivity at 95% specificity (Table 1). We attribute the higher analytical performance of this algorithm to the limited impact of sample-specific factors (Figure 1A) on the results of approaches such as MANTIS (13, 15) that rely on comparison to matched normal samples. Between the two algorithms that rely on comparison to a set of baseline samples, MSIdetect had a slightly higher performance with >96.1% compared to >94.9% sensitivity for mSINGS at a 95% specificity.

We hypothesized that homopolymer selection could account, at least in part, for some of the limitations of MSI detection solutions that rely on NGS-based approaches, particularly those that leverage information from tumor-only data. This hypothesis is supported by evidence that the size and composition of the set of homopolymers considered impacts analytical performance (15).

To identify a set of homopolymers that would optimize MSI detection by NGS, we considered half of the samples in the pan-cancer dataset, hereafter referred to as training set, to identify homopolymers that would maximize the differences between MSI-H and MSS samples across multiple tumor types. To do so, we estimated the score at all homopolymers using MSIdetect. We defined groups of homopolymers based on whether the score in samples classified as MSI-H was higher than a fixed percentile (between 25-95%) of the maximal score observed for that homopolymer in samples classified as MSS from the same cancer type (Supplementary Table S1). Based on the MSIscore we computed for samples in the training set using the different homopolymers combinations (Supplementary Figures 2A-C) we determined the analytical performance and MSIscore variability associated with the different homopolymer sets. Based on these results (Supplementary Table S2) we concluded that the 136 homopolymers with a score in MSI-H samples higher than MSS in samples in than 80% of samples, offers optimal MSI detection relative to the other tested homopolymer sets. We hereafter refer to this homopolymer set as restricted homopolymer set.

We investigated what distinguished homopolymers in the restricted set from the remaining homopolymers captured by the WES solution. Relative to all considered homopolymers, those in the restricted set tend to be of intermediate length (median 15 bp, 11-25 bp, Figure 2B). This intermediate length is likely to facilitate read mapping and render homopolymers sensitive to dMMR dependent expansion and contraction.

In addition, we found that homopolymers in the restricted set have ~1.8x lower average population frequency amongst humans, based on gnomAD (two-tailed, Mann-Whitney test p-value<0.0002, Figure 2C) than other homopolymers of the same length (11-25 nt) which is likely to minimize the impact of population polymorphism in MSI score.

To assess the impact of implementing analysis of the restricted set on MSIdetect's analytical performance we considered the remaining samples of the pan-cancer data set, hereafter referred to as the test set. Restricting MSIdetect analysis to the restricted set of homopolymers improves performance relative to when all homopolymers in WES are considered. Specifically, restricting the analysis to the restricted homopolymer is associated with 100% sensitivity at 95% specificity (Table 2) and an increase in AUC (>0.995) in all tested tissues (Supplementary Table S3). This difference is also reflected in a slight increase in AUC (0.9995 for restricted homopolymer set compared to 0.9926 for all homopolymers). Like MSIdetect, the performance of the other algorithms tested (Table 2; Supplementary Table S3) also improved when only the restricted homopolymer set was considered. In line with previous work (15), this observation supports the use of specific microsatellite marker, including the set identified here, can improve the analytical performance of NGS-based methods of MSI detection.

In conclusion, the increase in analytical performance associated with the combination of algorithm and restricted set homopolymer regions limits the impact of biological and technical factors on the ability to detect by NGS the differences

TABLE 1 Sensitivity at 95% specificity for different algorithms in endometrial, colorectal and stomach cancer using WES homopolymers.

	Endometrial	Colorectal	Stomach
MSIdetect*	96.1% [98.6-91.8]	98.3% [99.9-90.9]	100.0% [100-95.7]
mSINGS*	94.9% [97.8-90.1]	98.3% [99.9-90.9]	100.0% [100-95.7]
MANTIS**	97.4% [99.3-93.6]	98.1% [99.9-90.1]	100.0% [100-95.7]

Asterisks indicate that * algorithm relies on comparison of tumor sample with a set of baseline samples or ** matching normal sample. Values inside square brackets indicate the 95% Confidence Interval for all estimates.

TABLE 2 Sensitivity at 95% specificity for different algorithms when considering restricted homopolymer set in endometrial, colorectal and stomach cancer.

	Endometrial	Colorectal	Stomach
MSIdetect*	100% [100-95.8]	100% [100-87.6]	100% [100-91-9]
mSINGS*	98.8% [100-93.7]	100% [100-87.6]	100% [100-91.9]
MANTIS**	100% [100-95.8]	100% [100-86.3]	100% [100-91.9]

Asterisks indicate that * algorithm relies on comparison of tumor sample with a set of baseline samples or ** matching normal sample. Values inside square brackets indicate the 95% Confidence Interval for all estimates.

in homopolymer length distribution caused by loss of MMR gene function, using tumor-only data.

MSIdetect is sensitive and specific in colorectal and endometrial cancer

Next, we sought to assess the analytical performance of MSIdetect in combination with the restricted homopolymer set in Formalin-Fixed Paraffin-Embedded (FFPE) clinical samples.

We first considered colorectal and endometrial cancer samples (44 and 30 samples, respectively) with MMR and MSI status defined using immunohistochemistry (IHC) and PCR (MSI-PCR) methods, respectively. The MSI and MMR status for these samples were concordant (Supplementary Table S4). We generated NGS data for homopolymers in the restricted set for these samples. We observe no overlap between the distribution of score obtained using MSIdetect for these samples dMMR/MSI-H from pMMR/MSS samples indicating the method allows distinction of. he two classes with 100% sensitivity and specificity (Figure 3A). To define the MSIscore threshold, we considered the standard deviation and the median score estimated for MSS samples (0.001 and 0.0028). We defined the thresholds for sample classification as follows: MSS samples have an MSIscore smaller than 0.005; MSI low confidence (MSI-LC) an MSIscore between 0.005 and 0.01; and MSI High confidence (MSI-H) an MSIscore higher than 0.010. These thresholds were chosen to maximize MSIdetect analytical performance. Change in the number or composition of homopolymer set considered should entail reevaluation of these thresholds (Supplementary Note 1).

To investigate the impact of tumor content on MSI detection performance, we diluted (1-90%), in replicate, one MSI-H tumor DNA in MSS tumor DNA from samples with relatively high tumor content samples. As expected, the MSIscore decreased with decreasing amounts of MSI-H tumor DNA (Figure 3B). The impact on sample classification of this decrease is similar to what was seen for MSI-PCR (Figure 3B). MSIscore is highly correlated between replicates ($R>0.99$, p-value$<2X10^{-8}$, Figure 3C), supporting the robustness of the approach. MSIdetect classified dilutions with limited MSI tumor DNA content (<2%) as MSI-LC indicating that MSIscore is sensitive to relatively low levels of homopolymer instability.

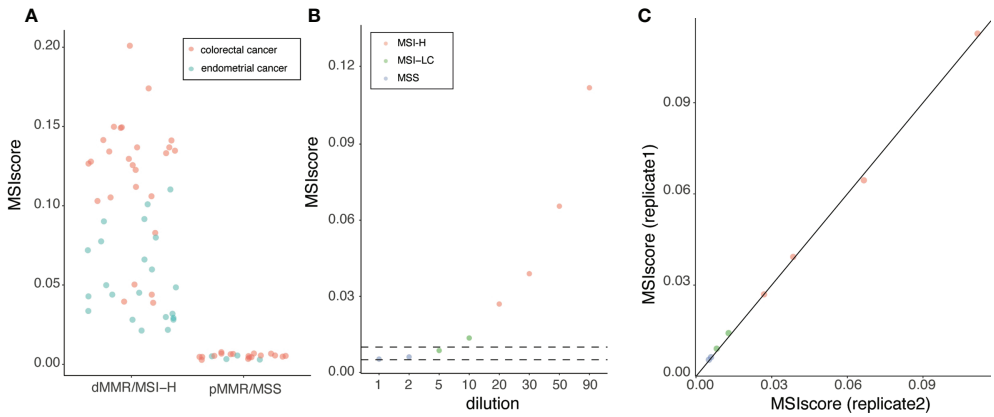


FIGURE 3
MSI detection in colorectal and endometrial FFPE clinical samples **(A)** MSIscore obtained for colorectal or endometrial cancer samples. Samples were grouped based on their respective MSI-PCR and IHC result. Each point corresponds to one sample colored by tissue of origin (refer to legend in figure) **(B)** MSIscore for a dilution series containing between 1 and 90% (x-axis) of DNA extracted from one MSI-H tumor DNA diluted in MSS tumor DNA in duplicates. Each point corresponds to one sample. Samples are colored according to results of MSI-PCR test (refer to legend in figure). **(C)** MSIscore obtained for replicate 1 and 2 for dilution series of MSI-H DNA in MSS DNA.

10.3389/fonc.2022.969238

MSIdetect detects MMR deficiency in various cancers, including glioma and sebaceous adenomas and carcinomas

Next, we considered samples from tumor types where MSI detection is more challenging, including glioma. When we considered the MMR status based on IHC, the method of preference for classification of these samples, we found that MSIdetect is 100% specific and 91% sensitive (Figure 4A) when only challenging samples are included. For 2 out of the 3 dMMR samples missed by MSIdetect (Figure 4A), MSI-PCR results were also available (Supplementary Table S4). In both cases, the number of loci found to be unstable (2/5) was low and below the recommended test's threshold for MSI classification. The remaining sample was from glioma, where MSI-PCR is not routinely performed due to the lack of sensitivity of MSI-H status detection in this tumor type.

In addition to glioma, MSI detection is also challenging in other tumors such as cholangiocarcinoma, urothelial or adrenal carcinoma and sebaceous adenoma or carcinoma (23–26). When we considered these 3 cancers, we found that 94% of the 18 dMMR samples from these cancer types were classified as MSI by MSI detect. This includes 2 samples classified by MSI-PCR as MSS, 1 sebaceoma and 1 cholangiocarcinoma (Figure 4B).

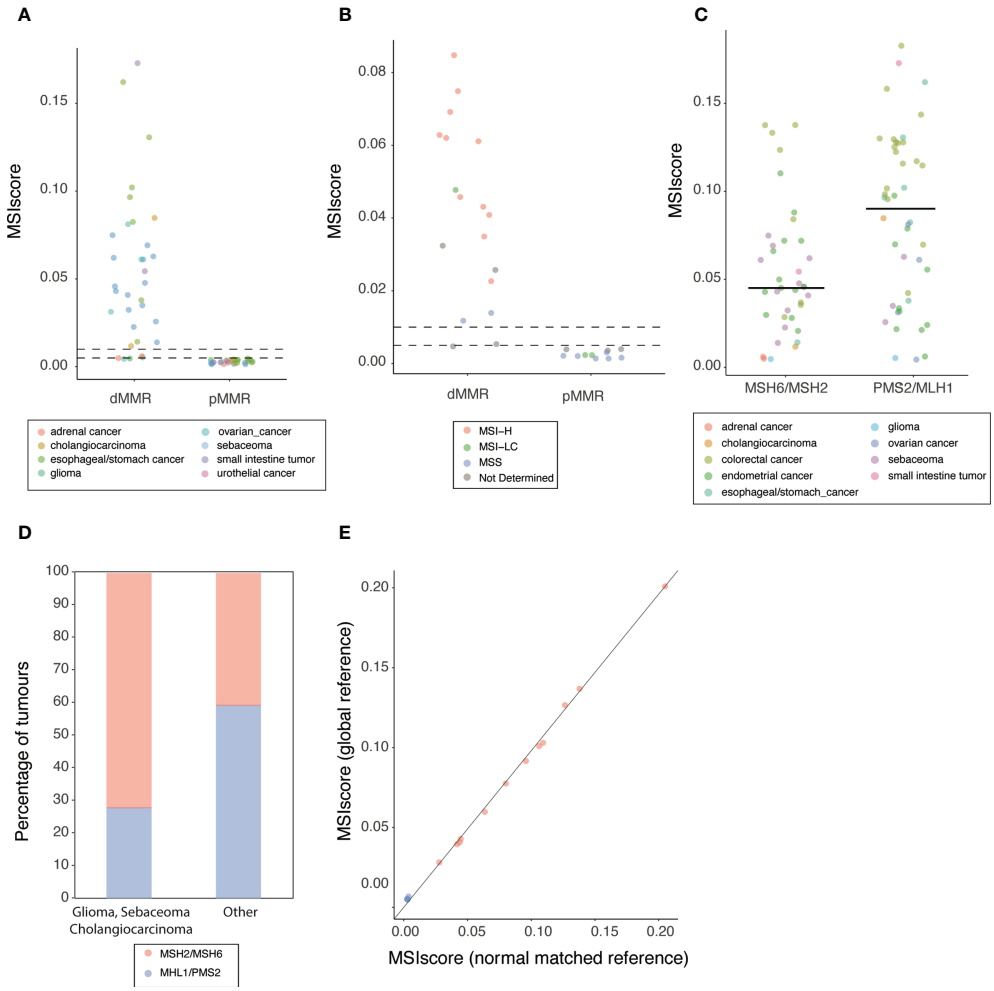


FIGURE 4
MSI detection in FFPE clinical samples **(A)** MSIscore obtained for dMMR or pMMR FFPE samples. Each point corresponds to one sample colored by tissue of origin (refer to legend in figure). Horizontal lines top to bottom indicates MSI-HC and MSI-LC threshold respectively **(B)** MSIscore obtained for glioma, sebaceoma and cholangiocarcinoma FFPE samples classified by IHC as dMMR or pMMR. Each point corresponds to one sample colored by MSI-PCR status. Horizontal lines top to bottom indicates MSI-H and MSI-LC threshold respectively. **(C)** MSIscore obtained for dMMR FFPE samples grouped by pairs of protein lost (x-axis) Each point corresponds to one sample colored by tissue of origin (refer to legend in figure). Horizontal lines indicate the median score for the group. **(D)** Histogram of the percentage of genes with detected loss of MSH2/MSH6 or MHL1/PSM2 grouped by cancer type **(E)** MSIscore obtained using either a global reference (y-axis) or a reference build using a matched-normal samples. Each point corresponds to one sample colored by MSI-PCR result.

Differences in mutational patterns between tumor types have been proposed to account for decreased MSI detection sensitivity (27, 28). Given the relatively small number of samples were expression of only one protein in the functional heterodimer pairs MLH1/PMS2 or MSH2/MSH6 is loss (Supplementary Table S4) we grouped samples according to heterodimer loss of function.

Interestingly, dMMR samples where MLH1 or PMS2 (median MSIscore=0.090) were lost have significantly higher levels of microsatellite instability (two-tailed Mann-Whitney test p-value<0.005) than dMMR samples with loss of function in MSH2 or MSH6 (median MSIscore=0.045) (Figure 4C). Loss of MSH6 function is known to result in lower levels of microsatellite instability (29). However the relatively low number of samples where only MSH2 or MSH6 appears to be lost by IHC, that we attribute to protein regulation by dimer stabilization (30), limits our ability to assess the impact of loss of function either gene to the MSIscore observed of MHS2/MHS6 deficient tumors.

This difference in MSIscore observed between MLH1/PMS2 and MHS2/MHS6 deficient tumors explains, at least in part, the low levels of instability observed in glioma and sebaceous adenoma or carcinoma. Indeed, in these tumors, MSH2/MSH6 mutations are significantly (two-tailed Fisher's exact test p-value<0.005) more frequent (13/18 cases) than in the rest of the cohort where MSH2/MSH6 mutations are less common (26/64 samples) (Figure 4D).

For a subset, in addition to tumor samples, non-tumor matched normal material was also available (16 samples). These samples allowed us to assess the impact of the results obtained when microsatellite instability is measured relative to a panel of normal samples (global reference) or a matched normal sample (Methods). We observed a strong correlation (R= 0.996, correlation test p-value<$2.2X10^{-16}$, Figure 4E) between the MSIscore obtained using the global and match reference. The observation that the score is similar when using global or normal matched supports that the combination of algorithm and restricted homopolymer set allows overcoming some of the challenges of tumor only analysis of microsatellite regions.

Conclusion

Mismatch repair deficiency (dMMR) confers sensitivity to immune checkpoint inhibition therapy across different cancer types (3–5). However, and despite its pan-cancer value, clinical detection of this molecular signatures is often restricted to colorectal and endometrial cancer where this molecular phenotype is most common (20). This is in part because dedicated assays, analysis of protein loss of function by immunohistochemistry or of MSI by PCR, are still preferred to next generation sequencing (NGS) based methods (6) but require tumors to be matched to paired normal samples for analysis of non-colorectal cancers sample. The main advantage of NGS based methods is that they allow integration of MSI detection as part of comprehensive molecular profiling assays, supporting adoption of dMMR testing and increasing the number of patients considered for immunotherapy (19).

Here we describe MSIdetect, a NGS based solution developed to support accurate detection of MSI from tumor-only data. We identified the sample-specific and analytical factors that limit performance MSI detection by NGS. We found that when considering tumor only data, accounting for homopolymer properties, indel burden and tumor content increases sensitivity. However, algorithm improvements alone cannot account for the impact of tissue of origin and patient ethnicity when only tumor samples are available. To address this limitation, we used publicly available data to identify a set of loci that is minimally impacted by sample specific factors. Integration of these insights limits the impact of the identified confounders on the results from tumor-only data and supports performances comparable to what can be obtained when normal matched samples are available.

We investigated the accuracy of MSIdetect in a diverse cohort of clinical samples using results of IHC as ground truth. As highlighted by a recent meta-analysis the evidence supporting the value of MSI-NGS solutions in non-colorectal cancers is low, demonstrating the need for development and validation of NGS based methods that can accurately detect MSI in other cancer types (31).

We show that MSIdetect is 100% accurate in colorectal and endometrial cancer. This is despite the MSIscore being lower in endometrial relative to colorectal cancer, consistent with the previously reported (17, 18) differences in size and frequency of indels at microsatellites in these two cancer types. Tissue specific differences on the impact of loss of MMR on microsatellite instability have also been reported in other cancer types, including glioma or sebaceous adenoma or carcinoma, where MSI detection is known to be challenging (23, 24, 26, 27). When MSIdetect was used to analyze samples from these cancer types we observed a slightly lower overall accuracy (accuracy 97.8%). For 2 out of the 3 false negative samples, MSI status based on a commonly used PCR based method was also available. Both these samples were also classified by the PCR based method as MSS indicating that the impact of loss of MMR function on expansion and contraction of homopolymer is low and generally hard to detect in these cases. Interestingly, we found that dMMR in these samples is caused by loss of MSH6 alone or together with MSH2 which is associated with loss of sensitivity to detect MSI (27, 28).

In summary, we show that MSIdetect supports accurate detection of MSI signatures in different cancer types. Its adoption alone or as part of molecular profiling solutions can increase the number of patients identified that are likely to benefit from immune checkpoint inhibitor therapy, particularly in cancers where PCR based MSI detection methods were found

to have limited sensitivity and in samples with low tumor content.

Materials and methods

MSI analysis of public data

Tumor-normal whole-exome sequencing data for 78, 85 and 156 MSI-H and 245, 265 and 274 MSS colorectal adenocarcinoma, stomach adenocarcinoma, and uterine endometrial carcinoma, respectively, was obtained from The Cancer Genome Atlas (TCGA). Aligned BAM files (to hg38) and associated clinical information for all the samples was downloaded from Genomic Data Commons.

We considered the coverage by sufficient quality reads and excluded from our analysis homopolymers with insufficient coverage.

Description and MSI calling using MSIdetect

MSIdetect score for sample *j* is calculated as the median homopolymer score, *HPscore*, for all homopolymers considered in the analysis. The *HPscore* for homopolymer *i* in sample *j* is defined as the product between the values of the parameters, *p1* and *p2*, that maximize the fit between the read length distribution obtained for homopolymer *i* in sample *j* (D_j^i) with the read length distribution of homopolymer *i* in reference MSS sample(s), hereafter referred to as reference (D_{ref}^i), using the multiparametric function defined by equation 1.

equation 1

$$(p_1, \quad p_2, \quad p_3) = \arg\min\left(\int_0^{l_{max}} |D_j^i(l) - T(D_{ref}^i(l), p_1, p_2, p_3)| dl \right)$$

Where $l_{max}$is the maximum homopolymer length observed in D_{ref}^i, l is the homopolymer length and T is the function which transform D_{ref}^i (l) according to the transformation described below (equation 2):

equation 2

$$T(D_{ref}^i(l), p_1, p_2, p_3) = (1 - p_1) \cdot p_3 \cdot D_{ref}^i\left(\left(\frac{l - l_{ref}}{p_3} + l_{ref}\right) - p_2 \cdot l_{ref}\right) + p_1 \cdot D_{ref}^i(l)$$

where l_{ref} is the reference length at this locus.

For a given homopolymer *i*, *p1* is the difference between the measured height of the read distribution peak in sample *j* and in the reference distribution; *p2* is the maximum difference observed in homopolymer length between sample *j* and the reference and reflects the difference in peak position in sample *j* relative to the reference distribution; and p3 is the width of the length distribution for homopolymer *i* in sample *j*. As depicted in Figure 1B, p1 and p2 are expected to change as function of tumor content and indel burden, respectively. The parameter p3 captures changes in homopolymer lengths distribution width between the sample and the reference distribution.

Because in MSS samples the value of either *p1* or *p2* will be close to 0, meaning that value taken by any of the other parameters on score, we chosen to consider only *p1* and *p2* in the estimation of the homopolymer score.

Reference length distribution is pre-computed from aligned sequence data for MSS or matched normal samples. Unless stated otherwise analysis of TGCA and clinical samples were done based on the comparison to a reference length distribution computed using aligned sequencing data for 10 MSS samples selected randomly from either the cancer genome atlas (TCGA) or clinical samples, respectively. As documented, in Supplementary Note 2 the set of MSS samples chosen to build the reference distribution minimally impacts *MSIscore*.

Only reads that are perfectly matched to the homopolymer region excluding the homopolymer region plus or minus 3 nucleotides were considered. Reads mapping to the forward and reverse strand are considered separately and$HP \quad score_j^i$ is the average of the score in both directions.

MSI calling using mSINGS

We considered 25 MSS samples from colorectal adenocarcinoma, stomach adenocarcinoma and uterine endometrial carcinoma to build the reference distribution using default parameters. Loci with no variance were excluded as recommended by the developers. MSI score was computed as described by developer's version v.4.0.

MSI calling using MANTIS

MSI score was computed using MANTIS (version v1.0.5) and the parameters recommended in (15), (mrq = 20, mlq = 25, mlc = 20,mrr = 1) for tumor and normal matched paired samples.

Analysis of human polymorphism

We extracted variants reported in from gnomAD v2.1.1 that impact homopolymer length distribution and computed their frequency using their allele count across populations.

Characterization of clinical samples

Tissue samples from patients diagnosed for their MSI and MMR status between 2016 and 2020 in the pathology department of the *Hospices Civils de Lyon* (HCL, France). The properties of the clinical samples are listed in Supplementary Table S4. Non-CRC carcinomas were classified according to the World Health Organization (WHO) histopathological classifications and were reviewed independently by two pathologists for tumor classification and cellularity. MSI status was done using multiplex PCR and capillary electrophoresis-based assay PCR– based MSI test used in our laboratory was done accordingly to the instructions provided by the manufacturers (Promega Corporation, Madison, WI, USA). Two µL of DNA which concentration was adjusted to 10 ng/µL was used to co-amplify by multiplex PCR 5 mononucleotide repeat markers: BAT-25, BAT-26, NR-21, NR-24 and MONO-27, and 2 pentanucleotide repeat markers (Penta C and Penta D). The PCR products are separated by capillary electrophoresis using an Applied Biosystems® 3130 Genetic Analyzer. The output data were analyzed with GeneMapper® software (Applied Biosystems) to determine MSI status of test samples.

To investigate the mismatch repair protein (MMR) expression standard 4-µm thick FFPE tumor sections were subjected to immunohistochemistry staining (IHC) analysis using MLH1 antibody (Ab) (clone G168-728, Ventana Ab, 1/25), MSH2 Ab (clone 25D12 DBS Clinisciences, 1/25), MSH6 Ab (clone 44 BD Biosciences, 1/500) and PMS2 Ab (A16-4, Pharmingen, 1/200) on a Ventana automated staining platform (BenchMark ULTRA, Tucson, AZ, USA). Internal positive control was included in the tissue section. Loss of MMR expression was considered in case of total absence of nuclear expression by tumor cells while normal cells express the protein (32–34). All samples were from the tumor bank "Tissu-tumorotheque Est" and "Tissu-tumorotheque Sud" of the Biological Resource Centre (Centre de Ressource Biologique, CRB) of the HCL (Lyon, France).

Clinical sample preparation and sequencing

The regions corresponding to the restricted homopolymer set (136 loci) plus their neighboring genomic regions in hg19 were downloaded and DNA repeat content analyzed. After exclusion of homopolymers within repetitive regions probes of 117 homopolymers were designed and ordered.

Targeted libraries were created using capture-based enrichment technology. First, 50 ng of input FFPE extracted genomic DNA was enzymatically fragmented, end-repaired and A-tailed, followed by ligation to custom short y-shaped adapters. The ligation products were purified with AMPure beads (Beckman Coulter) and then amplified by PCR for 10 to 14 cycles (depending on the amount of input DNA) using Illumina-compatible primers with dual-indices. Amplified libraries were cleaned-up with AMPure beads (Beckman Coulter) and libraries pooled to give a total of 1.8 µg. The pools were mixed with human Cot-1 DNA (Life Technologies) and xGen Universal Blockers-TS Mix oligos (Integrated DNA Technologies) and lyophilized. Pellets were resuspended in a hybridization mixture, denatured for 10 min at 95°C and incubated for 4-16 h at 65°C in the presence of biotinylated probes (xGEN Lockdown IDT®). Probe-hybridized library fragments were captured with Dynabeads M270 Streptavidin (Invitrogen) and then washed. The captured libraries were amplified by PCR for 15 cycles and cleaned-up using AMPure beads (Beckman Coulter).

Paired end (150 base pair) reads libraries were sequenced on the Illumina Miseq or NextSeq platform (Illumina Inc., San Diego, CA, USA). Sequencing data was processed using the SOPHiA GENETICS proprietary pipelines accessible through SOPHiA GENETICS DDM platform. All samples were sequenced to approximately 1000 x coverage which is more than the estimated minimal depth required to ensure accurate distinction between MSI and MSS samples (Supplementary Note 3).

Statistical analysis

Statistical analysis and graphics were done using R.

Data availability statement

MSIdetect algorithm is a SOPHiA GENETICS proprietary algorithm and is available as part of SOPHiA GENETICSDDM platform. 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 Biological Resource Center of the Hospices Civils de Lyon. 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

AM, FM, LS, XX, AW, and ZX conceived and planned the study. FM and LS developed MSIdetect pipeline. ES and GF generated NGS data under AW supervision, FM analyzed NGS dataset with support from MW and XX. CF-P, TF, VH, BBal, MD, BBan, DM, and JL were responsible for sample collection. CC, MB, and PM coordinated the study, AM and FM did the statistical analysis and prepared figures and tables. CF-P, JL, and ZX provided intellectual input for data interpretation. AM wrote the first draft of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgments

We thank Corinne Perrin and Elisabeth Blasco from the tumour bank "Tissu-tumorotheque Est" and "Tissu-tumorotheque Sud" of the Hospices Civils de Lyon's Biological Resource Centre for the collection of patient's consents. We thank Aurélie Gauthier (GHE pathology department) for its technical assistance to handle and send the MMR samples, Dr PP Bringuier and M Barritault who facilitated the sending of the samples.

Conflict of interest

AM, FM, LS, ES, GF, AW, MW, XX, CC, MB, PM, and ZX are SOPHiA GENETICS employees. CF-P reports sponsorship for meeting attendance from Roche and personal fees for advisory board work from Novartis, outside the submitted work. JL reports consulting for SOPHiA GENETICS and Decibio and personal fees for advisory board work and attendance to scientific meeting by Roche, Astra-Zeneca, BMS, Lilly and Nanostring.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.969238/full#supplementary-material

SUPPLEMENTARY TABLE 4
Properties of the clinical samples used in the study.

SUPPLEMENTARY FIGURE 1
(A) Score obtained with MSIdetect, using WES homopolymers in colorectal, stomach and endometrial cancer. Each point corresponds to one sample colored by reported MSI status (refer to legend in the figure). Receiver Operating curves and corresponding Area Under the Curve (AUC) values (in the inset) for endometrial **(B)**, colorectal **(C)** and stomach **(D)** cancers for MSI classification by MSIdetect, mSIGNS and MANTIS using WES homopolymers given the MSI status reported by TCGA.

SUPPLEMENTARY FIGURE 2
Distribution of the MSIscore for microsatellite instability high (MSI-H, red) and stable (MSS, blue) samples in the training set using the different homopolymers combinations in endometrial **(A)**, colorectal **(B)** and Stomach **(C)** cancer.

References

1. Kunkel TA. Evolving views of DNA replication (in)fidelity. *Cold Spring Harb Symp Quant Biol* (2009) 74:91–101. doi: 10.1101/sqb.2009.74.027

2. Hsieh P, Yamane K. DNA Mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev* (2008) 129:391–407. doi: 10.1016/j.mad.2008.02.012

3. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596

4. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* (2017) 357:409–13. doi: 10.1126/science.aan6733

5. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA Approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors. *Clin Cancer Res* (2019) 25:3753–8. doi: 10.1158/1078-0432.CCR-18-4070

6. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol* (2019) 30:1232–43. doi: 10.1093/annonc/mdz116

7. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* (2004) 5:435–45. doi: 10.1038/nrg1348

8. Goel A, Nagasaka T, Hamelin R, Boland CR. An optimized pentaplex PCR for detecting DNA mismatch repair-deficient colorectal cancers. *PloS One* (2010) 5: e9393. doi: 10.1371/journal.pone.0009393

9. Stelloo E, Jansen AML, Osse EM, Nout RA, Creutzberg CL, Ruano D, et al. Practical guidance for mismatch repair-deficiency testing in endometrial cancer. *Ann Oncol* (2017) 28:96–102. doi: 10.1093/annonc/mdw542

10. Siemanowski J, Schömig-Markiefka B, Buhl T, Haak A, Siebolts U, Dietmaier W, et al. Managing difficulties of microsatellite instability testing in endometrial cancer-limitations and advantages of four different PCR-based approaches. *Cancers (Basel)* (2021) 13:1268. doi: 10.3390/cancers13061268

11. Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, Flejou J-F, et al. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to

determine the microsatellite instability status of human tumors. *J Clin Oncol* (2006) 24:241–51. doi: 10.1200/JCO.2005.02.7227

12. Campanella NC, Berardinelli GN, Scapulatempo-Neto C, Viana D, Palmero EI, Pereira R, et al. Optimization of a pentaplex panel for MSI analysis without control DNA in a Brazilian population: correlation with ancestry markers. *Eur J Hum Genet* (2014) 22:875–80. doi: 10.1038/ejhg.2013.256

13. Baudrin LG, Deleuze J-F, How-Kit A. Molecular and computational methods for the detection of microsatellite instability in cancer. *Front Oncol* (2018) 8:621. doi: 10.3389/fonc.2018.00621

14. Albayrak A, Garrido-Castro AC, Giannakis M, Umeton R, Manam MD, Stover EH, et al. Clinical pan-cancer assessment of mismatch repair deficiency using tumor-only, targeted next-generation sequencing. *JCO Precis Oncol 1084–1097* (2020) 1084–97. doi: 10.1200/PO.20.00185

15. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* (2017) 8:7452–63. doi: 10.18632/oncotarget.13918

16. Cortes-Ciriano I, Lee S, Park W-Y, Kim T-M, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* (2017) 8:15180.

17. Wang Y, Shi C, Eisenberg R, Vnencak-Jones CL. Differences in microsatellite instability profiles between endometrioid and colorectal cancers: A potential cause for false-negative results? *J Mol Diagn* (2017) 19:57–64. doi: 10.1016/j.jmoldx.2016.07.008

18. Wu X, Snir O, Rottmann D, Wong S, Buza N, Hui P, et al. Minimal microsatellite shift in microsatellite instability high endometrial cancer: a significant pitfall in diagnostic interpretation. *Mod Pathol* (2019) 32:650–8. doi: 10.1038/s41379-018-0179-3

19. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* (2001) 409:860–921. doi: 10.1038/35057062

20. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med* (2016) 22:1342–50. doi: 10.1038/nm.4191

21. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res* (2003) 544:65–85. doi: 10.1016/s1383-5742(03)00032-2

22. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem* (2014) 60:1192–9. doi: 10.1373/clinchem.2014.223677

23. Cerretelli G, Ager A, Arends MJ, Frayling IM. Molecular pathology of lynch syndrome. *J Pathol* (2020) 250:518–31. doi: 10.1002/path.5422

24. Eckert A, Kloor M, Giersch A, Ahmadi R, Herold-Mende C, Hampl JA, et al. Microsatellite instability in pediatric and adult high-grade gliomas. *Brain Pathol* (2007) 17:146–50. doi: 10.1111/j.1750-3639.2007.00049.x

25. Limpaiboon T. Prognostic significance of microsatellite alterations at 1p36 in cholangiocarcinoma. *WJG* (2006) 12:4377. doi: 10.3748/wjg.v12.i27.4377

26. Goeppert B, Roessler S, Renner M, Singer S, Mehrabi A, Vogel MN, et al. Mismatch repair deficiency is a rare but putative therapeutically relevant finding in non-liver fluke associated cholangiocarcinoma. *Br J Cancer* (2019) 120:109–14. doi: 10.1038/s41416-018-0199-2

27. Goodfellow PJ, Billingsley CC, Lankes HA, Ali S, Cohn DE, Broaddus RJ, et al. Combined microsatellite instability, MLH1 methylation analysis, and immunohistochemistry for lynch syndrome screening in endometrial cancers from GOG210: An NRG oncology and gynecologic oncology group study. *J Clin Oncol* (2015) 33:4301–8. doi: 10.1200/JCO.2015.63.9518

28. Wang A, McCracken J, Li Y, Xu L. The practice of universal screening for lynch syndrome in newly diagnosed endometrial carcinoma. *Health Sci Rep* (2018) 1:e43. doi: 10.1002/hsr2.43

29. Verma L, Kane MF, Brassett C, Schmeits J, Evans DG, Kolodner RD, et al. Mononucleotide microsatellite instability and germline MSH6 mutation analysis in early onset colorectal cancer. *J Med Genet* (1999) 36:678–82.

30. Arlow T, Kim J, Haye-Bertolozzi JE, Martínez CB, Fay C, Zorensky E, et al. MutSα mismatch repair protein stability is governed by subunit interaction, acetylation, and ubiquitination. *G3 Genes|Genomes|Genetics* (2021) 11:jkaa065. doi: 10.1093/g3journal/jkaa065

31. Bartley AN, Mills AM, Konnick E, Overman M, Ventura CB, Souter L, et al. Mismatch repair and microsatellite instability testing for immune checkpoint inhibitor therapy: Guideline from the college of American pathologists in collaboration with the association for molecular pathology and fight colorectal cancer. *Arch Pathol Lab Med* (2022) 146:1194–210. doi: 10.5858/arpa.2021-0632-CP

32. Shia J, Black D, Hummer AJ, Boyd J, Soslow RA. Routinely assessed morphological features correlate with microsatellite instability status in endometrial cancer. *Hum Pathol* (2008) 39:116–25. doi: 10.1016/j.humpath.2007.05.022

33. Shia J. The diversity of tumours with microsatellite instability: molecular mechanisms and impact upon microsatellite instability testing and mismatch repair protein immunohistochemistry. *Histopathology* (2021) 78:485–97. doi: 10.1111/his.14271

34. Bartley AN, Hamilton SR, Alsabeh R, Ambinder EP, Berman M, Collins E, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the colon and rectum. *Arch Pathol Lab Med* (2014) 138:166–70. doi: 10.5858/arpa.2013-0231-CPc

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