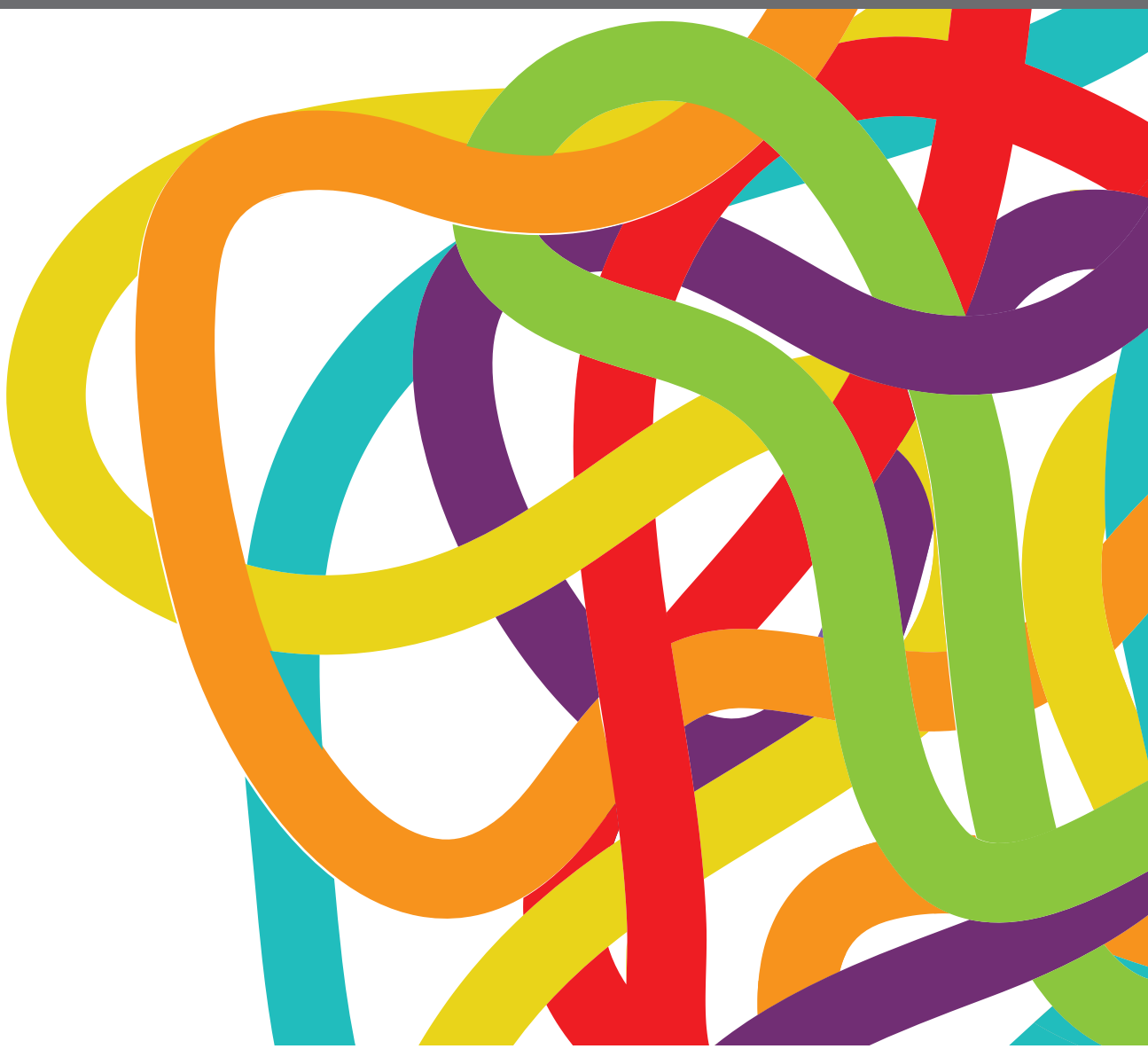
INSIGHTS INTO CANCER OF UNKNOWN PRIMARY (CUP) RESEARCH: NOVEL DIAGNOSTIC AND THERAPEUTIC STRATEGIES

EDITED BY: Xichun Hu, Yi Ba, Xiaoyan Zhou, Natalie Cook and Qinghua Xu
PUBLISHED IN: Frontiers in Oncology, Frontiers in Genetics and
Frontiers in Cell and Developmental Biology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-275-4
DOI 10.3389/978-2-83250-275-4

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

INSIGHTS INTO CANCER OF UNKNOWN PRIMARY (CUP) RESEARCH: NOVEL DIAGNOSTIC AND THERAPEUTIC STRATEGIES

Topic Editors:
Xichun Hu, Fudan University Shanghai Cancer Center, China
Yi Ba, Tianjin Medical University, China
Xiaoyan Zhou, Fudan University, China
Natalie Cook, The University of Manchester, United Kingdom
Qinghua Xu, The Canhelp Genomics Research Center, China

Citation: Hu, X., Ba, Y., Zhou, X., Cook, N., Xu, Q., eds. (2022). Insights Into Cancer of Unknown Primary (CUP) Research: Novel Diagnostic and Therapeutic Strategies. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-275-4

Table of Contents

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fonc.2021.634563

The Diminishing Importance of Primary Site Identification in Cancer of Unknown Primary: A Canadian Single-Center Experience

Boaz Wong[1,2], *Michael M. Vickers*[2,3] *and Paul Wheatley-Price*[2,3]*

[1] Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada, [2] Cancer Therapeutics Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada, [3] Division of Medical Oncology, The Ottawa Hospital Cancer Centre, Ottawa, ON, Canada

Background: Cancer of unknown primary (CUP) describes patients with metastatic disease without an identified primary tumor site. Successful diagnosis and treatment of these patients remains difficult. Published guidelines on CUP have highlighted "favorable" subtype groups. We investigated a series of CUP patients to review adherence to guidelines, and identification of primary cancers or "favorable" subtypes.

Methods: Patients with histologically confirmed CUP at an academic institution from 2012 to 2018 were identified. Patient demographics, tumor presentation, diagnostic work-up and treatment information were retrospectively collected from electronic data records for descriptive analysis and compared to published clinical guidelines. The primary endpoint was the proportion of patients where the primary site was identified. Multivariable logistic regression models were used to identify factors associated with primary site identification. Kaplan-Meier survival curves were used to determine factors associated with poorer OS.

Results: Three hundred and five patients were included with a median follow-up time of 4.3 months. Primary tumor sites were identified in 109 patients (37.5%), which was most commonly lung cancer (33%). Statistical analyses did not identify any demographic or initial presentation factors associated with identifying the primary or not. More diagnostic tests did not increase the likelihood of primary site identification (*P=0.44*). Patients with an identified primary did not have longer OS than other patients (median 5.2 months vs. 4.7 months, *P=0.47*). 57 patients (18.7%) who had a defined "favorable" subtype experienced superior OS (36.6 months vs. 3.8 months; *P<0.0001*). Further, patients with good prognostic status who followed published treatment guidelines had longer OS (17.6 months vs. 13.2 months; *P=0.04*).

Conclusions: CUP remains a difficult cancer to diagnose and treat. These results suggest identifying the primary has less impact than anticipated, but particular efforts to identify patients with "favorable" subtypes of CUP is important prognostically.

Keywords: cancer of unknown primary, cancer epidemiology, patient prognosis, cancer diagnostics, favorable subtype, retrospective analysis

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Genshu Wang, Sun Yat-Sen University, China
Yan Du, Fudan University, China

***Correspondence:**
Paul Wheatley-Price
pwheatleyprice@toh.ca

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

Received: *28 November 2020*
Accepted: *27 January 2021*
Published: *03 March 2021*

Citation:
Wong B, Vickers MM and Wheatley-Price P (2021) The Diminishing Importance of Primary Site Identification in Cancer of Unknown Primary: A Canadian Single-Center Experience. Front. Oncol. 11:634563. doi: 10.3389/fonc.2021.634563

INTRODUCTION

Cancer of unknown primary site (CUP) describes the diagnosis of a metastatic cancer where the location of the primary tumor is unable to be identified following thorough medical investigation (1). Despite the steady decline in CUP incidence from 5% since the 1980s to around 2% of all new invasive cancer diagnoses, the prognosis for patients remains poor and is the fourth most common cause of cancer death (2–4). From 2000 to 2005, 3,564 new cases of CUP were diagnosed in Ontario, Canada (5).

A thorough diagnosis and work-up of the primary tumor site is paramount for directing treatment options, especially given the emergence of targeted therapies. At baseline, published guidelines by the National Comprehensive Cancer Network (NCCN), European Society of Medical Oncology (ESMO) and Spanish Society of Medical Oncology (SEOM) require a thorough medical history, physical examination, basic blood and biochemistry analyses, imaging, immunohistochemical analysis of biopsies and other specific tests where necessary (6–8). The ultimate aim when investigating CUP remains to try and identify the primary tumor site in order to optimize treatment plans according to other published guidelines. Median overall survival (OS) times for a patient with a diagnosed primary tumor site in Ontario are significantly longer than CUP patients (median, 11.9 months vs. 1.9 months) (9). Additionally, not only can proper workup suggest primary tumor sites, but also identify patients who may fall into a "favorable" subtype with a known treatment regimen (**Table 1**). Examples of favorable subtypes include neuroendocrine tumors, isolated axillary nodal metastases in females or non-supraclavicular cervical squamous cell carcinomas. These CUP patients make up 20% of all cases and have well-defined treatment regimens towards dramatically improved survival outcomes (10, 11).

The objective of this study was to describe how successful medical oncologists at our cancer center were in identifying the primary site, and secondly, whether or not primary site identification and adherence to current published CUP guidelines improves survival outcomes.

METHODS

Study Design and Patient Selection

A retrospective chart review of patients with CUP, seen as a new consult by The Ottawa Hospital Cancer Centre (TOHCC) medical oncologists between January 1st, 2012 and September 30th, 2018. Inclusion criteria for study were patients referred to medical oncologist as CUP, have histological confirmed metastatic CUP, and aged 18 or above at the time of the consultation. Patients who have had another primary cancer within 5 years prior to diagnosis were excluded. The data collection protocol was approved by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) with informed consent requirements waived given the retrospective nature of the study. The primary endpoint was the proportion of patients in whom an origin for the CUP was identified.

Data Collection

Patient demographic factors were gathered from electronic medical records including age, sex, and Eastern Cooperative Oncology Group (ECOG) performance status and date of diagnosis. Initial presentation characteristics including biopsy technique, location, number of metastatic sites, and histological subtype were described. To determine which diagnosis and treatment parameters were collected, the aforementioned published ESMO and SEOM guidelines were consulted (6, 8).

Diagnostic work-up was assessed by collecting variables including complete blood count and biochemistry at presentation (hemoglobin, platelets, white blood cell count) and appropriate serum tumor markers including alpha fetoprotein (AFP), beta human chorionic gonadotropin (hCG), chromogranin A, and prostate-specific antigen (PSA) where appropriate. Immunohistochemistry (IHC) analysis of the pathology specimen including cytokeratin 7 (CK7) cytokeratin 20 (CK20), and all other tested biomarkers were recorded. Each patient was then classified for compliance as: non-compliant, primary only (completion of CK7 and CK20), partial adherence (CK7 and CK20 with at least one additional recommended IHC biomarker), and complete adherence (CK7 and CK20 with all subsequent recommended IHC biomarkers) according to the published ESMO guidelines. Imaging tests performed (*e.g.* computed tomography/CT, positron emission tomography/PET scan), genetic screening and other diagnostic procedures (e.g. endoscopy, mammography) were also recorded. Abnormal bloodwork and biochemical thresholds were set according to guidelines by the Medical Council of Canada. A primary site was considered identified only if the physician explicitly makes the diagnosis in an initial consultation or subsequent progress note.

TABLE 1 | Table outlining common "favorable" clinical subtypes and recommended treatment compiled from the European Society of Medical Oncology (ESMO) and Spanish Society for Medical Oncology (SEOM) (6, 8).

Clinical subtype	Recommended treatment
Females with isolated axillary adenopathy	As per stages II–III breast cancer
Females with peritoneal adenocarcinomatosis	As per stage III ovarian cancer
Poorly-differentiated neuroendocrine carcinoma	Platinum + Etoposide
Well-differentiated neuroendocrine carcinoma	Somatostatin analogues, 5-fluorouracil
Squamous cell carcinoma with cervical adenopathy	Neck dissection with possible chemotherapy and radiation therapy
Squamous cell carcinoma with inguinal adenopathy	Inguinal node dissection with possible chemotherapy and radiation therapy
Tumor with CK7-/CK20+/CDX2+ molecular profile	As per stage IV colorectal cancer
Single metastatic site	Local resection with possible chemotherapy and radiation therapy
Males with bone metastases and PSA expression	Androgen deprivation therapy with possible radiation therapy

Treatment regimens were recorded including number of treatment lines, type of treatment, treatment details (chemotherapy regimen, radiation dose and site), number of cycles, and time to first treatment from date of diagnosis. First-line treatment plans were then compared to the published ESMO and SEOM guidelines for compliance to the recommended treatment algorithm outlined as follows. Specific treatment plans were defined if a patient fell within a "favorable" subtype (**Table 1**). Patients were considered to have good prognostic status according to published guidelines if they had an ECOG 0 or 1 and normal lactose dehydrogenase (LDH). Recommended treatment for these patients included a published list of defined two-drug chemotherapy regimens (6, 8). Other patients with a poorer prognosis were recommended to receive palliative radiation, single-line chemotherapy or best supportive care (BSC). Survival parameters were collected including time to death or last follow-up from date of diagnosis, vital status and cause of death where applicable.

Statistics

Clinicopathological factors were classified into dichotomous, categorical or continuous variables based upon clinically relevant thresholds expressed as a percentage of the cohort. Proportional differences in demographics, initial tumor presentation, date of diagnosis and diagnostic tests between the identified vs. unidentified primary experimental groups were determined using Fisher's exact test, Chi Square test or Student's t-test where appropriate. A univariable (UVA) Cox logistic regression analysis was used to evaluate any association between diagnostic factors and number of diagnostic tests with primary site identification. Kaplan-Meier survival curves were plotted for OS and the log-rank test used to compare differences between experimental groups. For the entire cohort, survival curves were tested and plotted for overall survival, identified vs. unidentified primary site, ECOG status 0–1 vs. 2+, and favorable subtype vs. other. In patients that did not fall within a favorable subtype but still had a favorable prognosis, survival curves were tested and plotted for those receiving treatment according to published guidelines vs. patients that did not. For all statistical analyses, the Statistical Package for Social Sciences (SPSS) version 13.0 (SPSS, Chicago, IL) and Prism 8 (GraphPad, San Diego, CA) were used. A P-value less than 0.05 was considered statistically significant.

RESULTS

Patient Demographics

Patient demographics and initial tumor presentation characteristics are summarized in **Table 2**. Three hundred and five patients were identified and retrospectively reviewed after applying the outlined inclusion and exclusion criteria. The mean age was 67.8 years with males accounting for 51% of the entire cohort. One hundred and thirteen patients (37%) had ECOG PS 0-1, 62 patients (20%) had ECOG PS 2, and 130 patients (43%) had ECOG PS 3-4. The median number of metastatic sites was 2 with the liver being the most common metastatic site (48%). Median smoking pack years was 30 years. Core biopsies were obtained for majority (53%) of the cohort. Histologically, adenocarcinoma accounted for the greatest proportion of patients (*N*=163, 53%).

In our cohort, 109/305 (36%) patients had a primary identified. The distribution of the primary sites identified can be found in **Figure 1**. The most commonly predicted and identified tumor site was the lung at 35/109 (33%), followed by cholangiocarcinoma (*N=13*, 17%) and duodenal cancer (*N=8*, 7%), respectively. The majority of patients with an identified lung primary were cigarette smokers (68%). No baseline demographic or presentation factors were associated with a significantly increased proportion of primary site identification (**Tables 2**, **3**) The only significant finding was that patients diagnosed in the second half of the study (2015-2018) were

TABLE 2 | Demographics and initial tumor presentation of entire CUP cohort.

Patient Characteristics	Total patients (%), *N=305*	Unidentified primary (%), *N=196*	Identified primary (%), *N=109*	P-value
Age				
Average	67.8	68.6	66.4	0.146†
< 39	5 (1.6)	1 (0.5)	4 (3.7)	
40–49	16 (5.2)	10 (5.1)	6 (5.5)	
50–59	54 (17.7)	36 (18.4)	18 (16.5)	
60–69	90 (29.5)	55 (28.1)	35 (32.1)	
70–79	84 (27.5)	52 (26.5)	32 (28.4)	
80+	56 (18.4)	42 (21.4)	14 (12.8)	
Sex				
Male	156 (51.1)	104 (53.1)	52 (47.7)	0.404
Female	149 (48.9)	92 (46.9)	57 (52.3)	
ECOG				
0	36 (11.8)	21 (10.7)	15 (13.8)	0.902
1	77 (25.2)	51 (26.0)	26 (23.9)	
2	62 (20.3)	36 (18.4)	26 (23.8)	
3	102 (33.4)	68 (34.7)	34 (31.2)	
4	28 (9.2)	20 (10.2)	8 (7.3)	
Smoking Status				
Current	57 (18.7)	34 (17.4)	23 (21.1)	0.257
Pack years (average)	37.4	34.6	42.1	
Ex	113 (37.0)	74 (37.8)	39 (35.8)	
Pack years (average)	32.3	32.0	32.8	
Never	120 (39.3)	75 (38.3)	45 (41.3)	
Unknown	15 (4.9)	13 (6.6)	2 (1.8)	
Histological Subtype				
Adenocarcinoma	163 (53.4)	107 (54.6)	56 (51.4)	0.214
Squamous Cell Carcinoma	38 (12.5)	25 (12.8)	13 (11.9)	
Neuroendocrine Carcinoma	47 (15.4)	35 (17.9)	12 (11.0)	
Poorly differentiated Carcinoma	34 (11.2)	17 (8.7)	17 (15.6)	
Other	20 (6.6)	10 (5.0)	10 (9.2)	
Unknown	3 (1.0)	2 (1.0)	1 (0.9)	
Number of Metastatic Sites				
1	104 (34.1)	71 (36.2)	33 (30.3)	0.315
2	90 (29.5)	54 (27.6)	36(33.0)	
3	64 (21.0)	45 (23.0)	19 (17.4)	
4	30 (9.8)	17 (8.7)	13 (11.9)	
5+	17 (5.6)	9 (4.6)	8 (7.3)	
Date of Diagnosis				
First half (2012–2015)	153 (50.2)	108 (55.1)	45 (41.3)	0.023*
Second half (2015–2018)	152 (49.8)	88 (44.9)	64 (58.7)	

†*by Student's t-test, all other by Fisher's exact test or Chi-square test, * denotes statistical significance.*

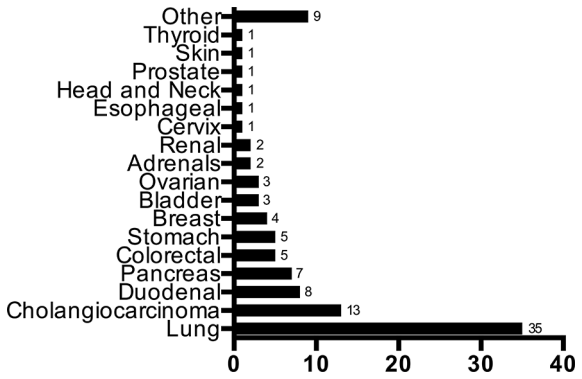


FIGURE 1 | Distribution of identified primary tumour sites in 109 patients.

more likely to have an identified primary tumor site compared to patients from the first half (42% vs. 29%; *P=0.02*).

Diagnostic Work-Up and Identification of Unknown Primary Tumor Site

Diagnostic work-up as outlined by published guidelines can be divided into three broad categories: bloodwork, immunohistochemistry (IHC) and imaging/diagnostic tests. A summary of diagnostic work-up are outlined in **Table 3**. Almost all patients received a complete history (99.7%) and physical examination (98%). Bloodwork and biochemical compliance to guidelines was high with 98% patients testing for complete blood counts, 53% for LDH, 100% for creatinine, 98% for electrolytes and 87% for calcium. Amongst serum markers for specific patient populations, the completion rates were variable from 59% for PSA (amongst males with bone metastases) to 14% for hCG (amongst patients with midline metastases).

All published guidelines pinpoint IHC as the most important diagnostic approach for CUP. As an initial screen, testing for cytokeratin (CK) 7 and 20 can broadly triage tumors into four potential categories. CK7 and CK20 were performed for 229 (75%) and 226 (74%) patients of our entire cohort respectively while 89 patients (29%) were not evaluated for either baseline biomarker. Frequency of additional IHC markers tested that are mentioned by the published guidelines can be found in **Supplementary Table S1**. In univariate analysis, no pattern defined category was significantly associated with identification of a primary site: CK7/20 only (OR 1.1[0.6–1.9]; *P=0.64*), CK7/20 with partial adherence to additional markers (OR 1.5[0.9–2.4]; *P=0.09*), and CK7/20 with complete adherence to all additional markers (OR 2.8[0.5–16.9]; *P=0.27*).

CT thorax (96%), abdomen and pelvis (95%) were commonly done as part of the minimal basic work-up. CT scans of the head

TABLE 3 | Diagnostic work-up summary of CUP cohort.

Patient Characteristics	Total patients (%), *N=305*	Unidentified primary (%), *N=196*	Identified primary (%), *N=109*	P-value
Documented Patient History	304 (99.7)	195 (99.5)	109 (100)	>0.999
Documented Physical Examination	298 (97.7)	191 (97.4)	107 (98.2)	>0.999
Bloodwork				
Complete Blood Count	303 (98.0)	194 (99.0)	109 (100)	0.593
Lactase Dehydrogenase (LD/LDH)	161 (52.8)	100 (51.0)	61 (56.0)	0.473
Creatinine	305 (100)	196 (100)	109 (100)	0.178
Electrolytes	299 (98.0)	191 (97.4)	108 (99.1)	0.282
Calcium	265 (86.9)	169 (86.2)	96 (88.1)	0.784
Alpha Fetoprotein (midline metastatic sites only, n=241)	58 (24.0)	35 (22.2)	23 (27.8)	0.346
Beta human chorionic gonadotropin (midline metastatic sites only, n=244)	33 (13.5)	19 (11.8)	14 (16.9)	0.324
Plasma Chromogranin A (neuroendocrine tumors only, n=51)	21 (41.2)	18 (46.2)	3 (25.0)	0.315
Prostate-Specific Antigen (male with bone metastases only, n=41)	24 (58.5)	16 (61.5)	8 (53.3)	0.719
Immunohistochemistry (IHC)				
CK7	229 (75.1)	144 (73.5)	85 (78.0)	
CK20	226 (74.1)	142 (72.4)	84 (77.1)	
Adherence to published IHC diagnostic guidelines				
No	89 (29.2)	59 (30.1)	30 (27.5)	
CK7 and CK20 only	65 (21.3)	47 (24.0)	18 (16.5)	
Partial adherence to additional markers	146 (47.9)	88 (44.9)	58 (53.2)	
Complete adherence to additional markers	5 (1.6)	2 (1.0)	3 (2.8)	
Imaging/Diagnostic Tests				
CT Head	181 (59.3)	113 (57.7)	68 (62.4)	0.467
CT Thorax	293 (96.1)	187 (95.4)	106 (97.3)	0.548
CT Abdomen/Pelvis	291 (95.4)	188 (95.9)	103 (94.5)	0.578
Octreoscan (neuroendocrine tumors only, n=49)	19 (38.8)	14 (38.9)	5 (38.5)	>0.999
Mammography (female only, n=148)	50 (33.8)	29 (31.5)	21 (37.5)	0.478
Endoscopy	121 (39.7)	80 (40.8)	41 (37.6)	0.716
Positron Emission Transmission (PET) Scan (cervical or single-site tumors only, n=117)	24 (20.5)	14 (17.9)	10 (25.6)	0.238
Next Generation Sequencing (NGS)	5 (1.6)	3 (1.5)	2 (1.8)	>0.999

were also common amongst the entire cohort (59%) despite only being recommended for cervical metastases. Guideline compliance rates for specialized diagnostic imaging and tests where indicated are as follows: octreotide scans (39%), mammography (34%), endoscopy (40%), and PET scans (21%). Rates for next generation sequencing (NGS) completion was only 2%: only 5 patients had documented molecular testing, 4 of which were Foundation Medicine and the last patient receiving a targeted FusionPlex sequencing assay.

The average number of diagnostic tests completed was 6.30 in the identified primary cohort compared to 6.16 in the unidentified primary cohort (*P=0.44*). No threshold number of diagnostic tests was found to be significantly associated with successful identification of the primary site. Finally, in univariate analysis, no single diagnostic test was significantly associated with identification of a primary tumor site (**Table 4**).

Treatment Regimen and Survival Analyses

Treatment regimens for the entire cohort are outlined in **Table 5**. Overall, 130 patients (43%) received systemic therapy and 118 patients (39%) received radiation therapy. As the first line of therapy, CUP patients received either BSC (*N=107*, 35%), chemotherapy (*N=98*, 32%), radiation therapy (*N=61*, 20%), concurrent chemoradiotherapy (*N=11*, 4%) or surgery (*N=28*, 9%). Of the 130 patients that received systemic therapy, 53 patients (41%) received only one line of treatment.

Median OS of the entire cohort was 4.3 months [range, 0.1–86.5] with a 90-day mortality of 34% at the time this study was performed (**Figure 2A**). Cancer-related causes were responsible for 98% of patient deaths in our cohort. Survival by ECOG status only can be found in **Supplementary Figure S1** . Patients with an identified primary site did not experience a significantly longer median OS when compared to patients without an identified primary site (median, 5.2 months, identified primary vs. 4.7 months, unidentified; *P=0.47*) (**Figure 2B**). The 57 patients (19%) with a favorable subtype (**Table 1**) experienced significantly longer median OS compared to those without favorable subtype (36.6 months vs. 3.8 months, *P<0.0001*) (**Figure 2C**). Of these 57 patients, 39 (68%) were prescribed a treatment regimen consistent with the recommended published guidelines. For patients without a favorable subtype but nevertheless have a good prognostic status (ECOG 0/1 and normal LDH), those who were prescribed a two-drug chemotherapy regimen according to treatment guidelines experienced longer median OS at 17.6 months compared to their counterparts at 13.2 months (*P=0.04*) (**Figure 2D**).

TABLE 4 | Univariable logistic regression analysis between diagnostic tests performed and primary site identification.

Diagnostic Test	Odd Ratio (95% CI)	P-value
History and Physical Examination	1.4 (0.3–0.9)	0.690
Alpha fetoprotein (AFP)	1.3 (0.7–2.5)	0.338
Human chorionic gonadotropin (hCG)	1.5 (0.7–3.2)	0.275
Chromogranin A	0.4 (0.1–1.5)	0.202
Prostate specific antigen (PSA)	0.9 (0.4–1.8)	0.689
CK7/CK20 immunohistochemistry	1.1 (0.7–1.9)	0.635
Computed tomography (CT) Head	1.2 (0.7–2.0)	0.420
CT Thorax	1.7 (0.5–7.8)	0.433
CT Abdomen/Pelvis	0.7 (0.2–2.3)	0.571
Octreoscan	1.0 (0.2–3.6)	0.978
Mammography	1.3 (0.6–2.6)	0.456
Endoscopy	0.9 (0.5–1.4)	0.584
Positron Emission Tomography (PET) Scan	1.5 (0.8–2.9)	0.189
Next Generation Sequencing (NGS)	1.2 (0.2–7.4)	0.841

TABLE 5 | Treatment summary of CUP cohort.

Patient Characteristics	Total patients (%), *N=305*	Unidentified primary (%), *N=196*	Identified primary (%), *N=109*	P-value
First Treatment Type				
Best Supportive Care	107 (35.1)	76 (38.8)	31 (28.4)	0.347
Chemotherapy	98 (32.1)	63 (32.1)	35 (32.1)	
Radiation Therapy	61 (20.0)	35 (17.9)	26 (23.9)	
Chemoradiotherapy	11 (3.6)	6 (3.1)	5 (4.6)	
Surgery	28 (9.2)	16 (8.2)	12 (11.0)	
Days to treatment (average)	39.7	38.9	41.0	0.649†
Treatment Regimen				
Systemic Therapy	130 (42.6)	84 (42.9)	46 (42.2)	>0.999
Radiation Therapy	118 (38.7)	69 (35.2)	49 (45.0)	0.111
Lines of Therapy				
1	196 (64.3)	129 (65.8)	67 (61.5)	0.850
2	62 (20.3)	37 (18.9)	25 (22.9)	
3	27 (8.9)	17 (8.7)	10 (9.2)	
4+	20 (6.6)	13 (6.6)	7 (6.4)	
Patients receiving 2+ lines of chemotherapy	43 (14.1)	28 (14.3)	13 (11.9)	0.604

†*by Student's t-test, all other by Fisher's exact test or Chi-square test.*

DISCUSSION

Over the past decade, research has made significant advancements in cancer care, yet CUP remains one of the most difficult cancer types to diagnose and treat. It remains associated with a very poor prognoses as reflected by a dismal median survival of 4.3 months in our cohort (12). Clinical guidelines by several medical organizations are available to reduce the uncertainty in the patient care process by suggesting a list of diagnostic tests; however, whether Canadian oncologists comply with these guidelines is unknown. We collected diagnostic and treatment details on 305 patients referred for consultation for CUP at the Ottawa Hospital from 2012 to 2018 and compared them to clinical guidelines set out by the ESMO and SEOM. While our physicians have a high compliance rate to most diagnostic tests, only 109 (35%) of patients successfully have a primary site identified. The rationale behind wanting to identify the primary site is for several reasons. Firstly, knowing the primary site aids the clinician in choosing a more selective treatment regimen specific to that cancer type as opposed to general chemotherapy. Secondly, the primary site may inform better patient prognoses. And thirdly, knowing the

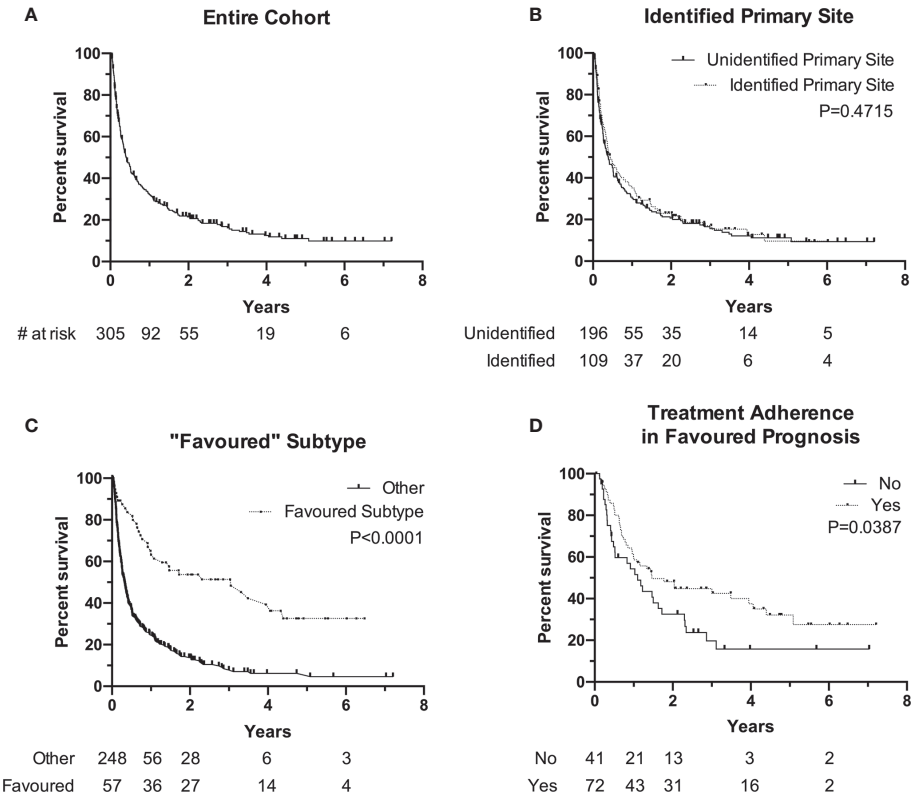


FIGURE 2 | Kaplan-Meier Survival curve of overall survival (OS0 of different subgroups. **(A)** Overall cohort median OS is 4.3 months (N + 305). **(B)** OS of patients with identified primary (median, 5.2 months, N = 10 vs. unidentified primary (median, 4.2 months, N = 196). **(C)** OS of patients with "favorable" subtype (median, 36.6 months, N = 57) vs. other (median, 3.8 months, N = 248). **(D)** OS of patients with favorable prognosis (ECOG 0/1 and normal LDH) and followed traetment guidelines (median, 17.6 months, N = 72) vs. did not follow guidelines (median, 13.2 months, N = 41). Notches denote censored events.

cancer type can reduce the anxiety associated with uncertainty that patients with CUP face.

The proportions of demographic and initial tumor presentation characteristics between the identified primary cohort compared to the unidentified were insignificant, and no single diagnostic test was associated with predicting the primary tumor site. Both cohorts performed a similar average number of diagnostic tests. Together, we were unable to identify a specific pattern or quantity of diagnostics that are associated with increased primary site identification, rather initial tests are more essential in providing context to the histological diagnosis by the pathologist. Immunohistochemical (IHC) pathological analyses are touted by multiple guidelines to be the ultimate diagnostic test for identifying the primary site, as well as importantly excluding potentially curable tumors (lymphomas, germ-cell tumors). Cytokeratins (CK7 and CK20) are the most important initial markers to classify carcinomas and adenocarcinomas, which was followed in 71% of our patients. Additional tumor-specific biomarkers can then be performed to provide clues into the cell type and origin (6, 8). Some patterns of IHC, combined with the appropriate clinical picture, are immensely helpful in suggesting a primary site. In addition to being the most common cancer in Canada, this may explain why the lung was the most commonly identified site, accounting for 33% of all identified primary sites, consistent with descriptive analyses of CUP patients at other academic centres (13, 14). The specific IHC pattern of CK7+/CK20- with a positive thyroid transcription factor-1 (TTF1) stain was commonly encountered and was confidently diagnosed as lung cancer in almost all cases (15). Outside of a few specific IHC combinations, a systematic approach by the pathologist remains the gold-standard in reaching a specific diagnosis (16).

Of minor note, a greater proportion of patients in the second half of our study (2015–2018) had a primary site identified (42% vs. 29% from 2012 to 2015). This finding may be explained given the publication of the ESMO and SEOM CUP guidelines in 2015 and 2018 respectively, giving oncologists and pathologists a more definitive approach to finding the primary site, but assumes that those clinicians were aware of or followed those guidelines. This included an increased awareness for the standardized approach to IHC, 32% of patients did not receive CK7 and CK20 markers in the first half which modestly decreased to 26.3% in the second half ($P=0.69$).

Being on the lower spectrum of compliance, we were also particularly interested in the contribution of PET scans (performed in 21% cases) and NGS (2%) on primary site

identification. Our results demonstrate that patients receiving PET scans were slightly more likely, albeit not significant, to have an identified site (18% vs. 13%; *P=0.24*). This finding is in accordance to multiple reviews highlighting that PET scans are instrumental, especially in head and neck cancers, in improving detection of primary sites missed by conventional imaging and previously undiagnosed metastases (17, 18). However, the use of PET scans is limited given the restrictions on availability and accessibility in Canada. Not all patients will be covered for PET under the Canadian provincial based healthcare system. Future studies investigating the efficacy of PET to justify its high cost remain warranted. Progress is also being made on identifying patterns in the genetic signatures of CUP in hopes of identifying patients with more responsive subtypes (19, 20). As only five patients (2%) in our study cohort underwent NGS, we are unable to make any conclusions on its efficacy in identifying the primary site or on improving OS. However, recent advances in other molecular techniques have shown promising results with a retrospective study by Moran et al. demonstrating an 87% success rate in identifying the primary tumor site by microarray DNA methylation signatures (21). It is clear that advancements in NGS is the direction that CUP diagnosis and treatment regimens should be following. Whether primary site identification by this method improves patient prognosis remains to be seen.

Our study shows that although oncologists in our academic center are consistent with CUP diagnostic guidelines in the majority of cases; however, no clear trend exists between the types of diagnostic tests performed and being able to successfully identify the primary site. Consequently, the question becomes whether a complete evaluation to identify the primary site is truly necessary. Our results would suggest against this notion as we demonstrate that despite having an identified primary site, these patients do not experience significantly improved OS compared to their unidentified counterparts (median, 5.2 months, identified vs. 4.2 months, unidentified; *P=0.47*). Rather, the main purpose of a diagnostic work-up at this point of time is to identify patients in favorable subgroups with defined treatment regimens who clearly demonstrate a superior OS (median, 36.6 months vs. 3.8 months; P<0.0001) (10). For example, the ESMO guidelines highlights a retrospective study by Hainsworth et al. demonstrating that CUP patients with an immunohistochemical profile similar to colorectal cancers responded well to colorectal-specific therapy (e.g. FOLFOX, FOLFIRI) (22).

Patients who were defined to have a favorable prognosis (not to be confused with "favorable" subtype) are recommended to undergo the recommended two-drug chemotherapy regimens to improve their prognosis (median, 17.6 months, followed guidelines vs. 13.2 months, did not follow guidelines; *P=0.04*). This finding has previously been discovered in a similar patient population (9). Patients in poorer-risk categories however, continue to be lacking in effective treatment options, with several clinical studies demonstrating no added benefit between different types of chemotherapy (23, 24). In accordance with our findings that 47% of patients without a favorable prognosis are managed with BSC, current focus should be placed on symptom management and preservation of quality of life. However, there remains hope for poor prognosis CUP patients given the positive results of the recent NivoCUP trial (UMIN000030649) (25), or our upcoming phase 2 Pembrolizumab study (NCT03391973).

The rise of immunotherapy in the recent decade will open up an entirely new discussion on its efficacy in CUP. However, given the targeted nature of immunotherapy and the ambiguous nature of CUP tumors, there needs to be a minimum level of investigation and resemblance to a particular tumor subtype in order for the chosen therapy to be effective. Immunoprofiling with biomarkers, at its current stage, has typically been unsuccessful in identifying CUP candidates for immunotherapy (26, 27). Clinical trials are ongoing investigating immunotherapy in CUP patients. As an encompassing primary site identification strategy is not yet available, future clinical trials should continue to aim in identifying CUP "favorable" subtypes that may respond well to novel therapeutic strategies. A recent breakthrough example of this by Verver et al. demonstrates that melanoma of unknown primary (MUP) are responsive to immune checkpoint inhibitors and targeted therapies, significantly improving their OS to 11 months from 4 months with these novel treatments (P<0.001) (28).

Taken together, our study did not demonstrate any particular diagnostic pattern, even suggested by published guidelines, to be superior in identifying the primary site. Rather, diagnostics should focus on identifying patients with "favorable" subtypes and prescribing chemotherapy according to published treatment guidelines. It is important to note that the conclusions drawn by our study are limited by its retrospective, single-center design. While we can identify trends of diagnosis and treatment by the oncologists at our center, the same trends may not be present in centers with different standards of care or those that have implemented different guidelines. Patients that were included in this study were first referred to medical oncology exclusively for CUP; therefore, it is possible that we were excluded patients who initially presented with CUP but had an identified primary site prior to their first consultation. In addition, our definition of an identified primary site was solely extracted from clinical notes and subject to interpretation. One oncologist might be more explicit in confirming the primary site with the same diagnostic results whereas another might not document their suspicions. Finally, the retrospective nature of this study only allows us to establish associations, not causative relationships, between our measured factors, the identification of the primary tumor site and patient outcomes. Next steps will include investigation of CUP guideline compliance at other oncology centers across the country and with this larger series, we can further evaluate whether primary site identification and guidelines adherence is truly effective for improving primary site identification and patient prognoses.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ottawa Health Science Network Research Ethics Board (OHSN-REB). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BW is the primary author and was involved in protocol writing, ethics approval, study design, generation of data collection tools, collection of data, statistical analysis, interpretation of data, and writing of the manuscript. MV was involved in interpretation of study data and editing of the manuscript. PW-P is the corresponding author and was involved in protocol writing, ethics approval, study design, data collection, interpretation of data, and editing of the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.634563/full#supplementary-material

REFERENCES

1. van de Wouw AJ, Jansen RLH, Speel EJM, Hillen HFP. The unknown biology of the unknown primary tumour: a literature review. *Ann Oncol* (2003) 14 (2):191–6. doi: 10.1093/annonc/mdg068
2. Mnatsakanyan E, Tung W-C, Caine B, Smith-Gagen J. Cancer of unknown primary: time trends in incidence, United States. *Cancer Causes Control* (2014) 25(6):747–57. doi: 10.1007/s10552-014-0378-2
3. Urban D, Rao A, Bressel M, Lawrence YR, Mileshkin L. Cancer of unknown primary: a population-based analysis of temporal change and socioeconomic disparities. *Br J Cancer* (2013) 109(5):1318–24. doi: 10.1038/bjc.2013.386
4. Rassy E, Pavlidis N. The currently declining incidence of cancer of unknown primary. *Cancer Epidemiol* (2019) 61:139–41. doi: 10.1016/j.canep.2019.06.006
5. Kim CS, Hannouf MB, Sarma S, Rodrigues GB, Rogan PK, Mahmud SM, et al. Identification and survival outcomes of a cohort of patients with cancer of unknown primary in Ontario, Canada. *Acta Oncol* (2015) 54(10):1781–7. doi: 10.3109/0284186X.2015.1020965
6. Fizazi K, Greco FA, Pavlidis N, Daugaard G, Oien K, Pentheroudakis G. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2015) 26:v133–8. doi: 10.1093/annonc/mdv305
7. Ettinger DS, Agulnik M, Cates JMM, Cristea M, Denlinger CS, Eaton KD, et al. Occult Primary: Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2011) 9(12):38. doi: 10.6004/jnccn.2011.0117
8. Losa F, Soler G, Casado A, Estival A, Fernández I, Giménez S, et al. SEOM clinical guideline on unknown primary cancer (2017). *Clin Transl Oncol* (2018) 20(1):89–96. doi: 10.1007/s12094-017-1807-y
9. Kim CS, Hannouf MB, Sarma S, Rodrigues GB, Rogan PK, Mahmud SM, et al. Survival outcome differences based on treatments used and knowledge of the primary tumour site for patients with cancer of unknown and known primary in Ontario. *Curr Oncol* (2018) 25(5):307-16. doi: 10.3747/co.25.4003
10. Greco FA, Oien K, Erlander M, Osborne R, Varadhachary G, Bridgewater J, et al. Cancer of unknown primary: progress in the search for improved and rapid diagnosis leading toward superior patient outcomes. *Ann Oncol* (2012) 23(2):298–304. doi: 10.1093/annonc/mdr306
11. Pavlidis N, Petrakis D, Golfinopoulos V, Pentheroudakis G. Long-term survivors among patients with cancer of unknown primary. *Crit Rev Oncol Hematol* (2012) 84(1):85–92. doi: 10.1016/j.critrevonc.2012.02.002
12. Amela EY, Lauridant-Philippin G, Cousin S, Ryckewaert T, Adenis A, Penel N. Management of "unfavourable" carcinoma of unknown primary site: Synthesis of recent literature. *Crit Rev Oncol Hematol* (2012) 84(2):213–23. doi: 10.1016/j.critrevonc.2012.03.003
13. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, et al. Projected estimates of cancer in Canada in 2020. *Can Med Assoc J* (2020) 192 (9):E199–205. doi: 10.1503/cmaj.191292
14. Uzunoglu S, Erdogan B, Kodaz H, Cinkaya A, Turkmen E, Hacibekiroglu I, et al. Unknown primary adenocarcinomas: a single-center experience. *Bosn J Basic Med Sci* (2016) 16(4):292–7. doi: 10.17305/bjbms.2016.1495
15. Rubin BP, Skarin AT, Pisick E, Rizk M, Salgia R. Use of cytokeratins 7 and 20 in determining the origin of metastatic carcinoma of unknown primary, with special emphasis on lung cancer. *Eur J Cancer Prev* (2001) 10(1):77–82. doi: 10.1097/00008469-200102000-00009
16. Selves J, Long-Mira E, Mathieu M-C, Rochaix P, Ilié M. Immunohistochemistry for Diagnosis of Metastatic Carcinomas of Unknown Primary Site. *Cancers* (2018) 10(4):108. doi: 10.3390/cancers10040108
17. Yoo J, Henderson S, Walker-Dilks C. Evidence-based Guideline Recommendations on the Use of Positron Emission Tomography Imaging in Head and Neck Cancer. *Clin Oncol* (2013) 25(4):e33–66. doi: 10.1016/j.clon.2012.08.007
18. Eskander A, Ghanem T, Agrawal A. The Education Committee of American Head and Neck Society (AHNS). AHNS Series: Do you know your guidelines? Guideline recommendations for head and neck cancer of unknown primary site. *Head Neck* (2018) 40(3):614–21. doi: 10.1002/hed.25026
19. Subbiah IM, Tsimberidou A, Subbiah V, Janku F, Roy-Chowdhuri S, Hong DS. Next generation sequencing of carcinoma of unknown primary reveals novel combinatorial strategies in a heterogeneous mutational landscape. *Oncoscience* (2017) 4:47–56. doi: 10.18632/oncoscience.352
20. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, et al. NGSCUP: Phase II trial of site-specific treatment based on gene expression and mutation profiling by next generation sequencing (NGS) for patients (pts) with cancer of unknown primary site (CUP). *J Clin Oncol* (2020) 38 (15_suppl):e15577–7. doi: 10.1200/JCO.2020.38.15_suppl.e15577
21. Moran S, Martínez-Cardús A, Sayols S, Musulén E, Balañá C, Estival-Gonzalez A, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol* (2016) 17 (10):1386–95. doi: 10.1016/S1470-2045(16)30297-2
22. Hainsworth JD, Schnabel CA, Erlander MG, Haines DW, Greco FA. A Retrospective Study of Treatment Outcomes in Patients With Carcinoma of Unknown Primary Site and a Colorectal Cancer Molecular Profile. *Clin Colorectal Cancer* (2012) 11(2):112–8. doi: 10.1016/j.clcc.2011.08.001
23. Bugat R, Bataillard A, Lesimple T, Voigt JJ, Culine S, Lortholary A, et al. Summary of the Standards, Options and Recommendations for the management of patients with carcinoma of unknown primary site (2002). *Br J Cancer* (2003) 89(S1):S59–66. doi: 10.1038/sj.bjc.6601085
24. Golfinopoulos V, Pentheroudakis G, Salanti G, Nearchou AD, Ioannidis JPA, Pavlidis N. Comparative survival with diverse chemotherapy regimens for cancer of unknown primary site: Multiple-treatments meta-analysis. *Cancer Treat Rev* (2009) 35(7):570–3. doi: 10.1016/j.ctrv.2009.05.005
25. Tanizaki J, Yonemori K, Akiyoshi K, Minami H, Ueda H, Takiguchi Y, et al. NivoCUP: An open-label phase II study on the efficacy of nivolumab in cancer of unknown primary. *J Clin Oncol* (2020) 38(15_suppl):106–6. doi: 10.1200/JCO.2020.38.15_suppl.106

26. El Rassy E, Pavlidis N. The current evidence for a biomarker-based approach in cancer of unknown primary. *Cancer Treat Rev* (2018) 67:21–8. doi: 10.1016/j.ctrv.2018.04.011
27. Mileshkin LR, Posner A, Pattison A, Balachander S, Etemadmoghadam D, Fellowes A, et al. Identifying biomarkers to guide immunotherapy treatment of cancer of unknown primary. *J Clin Oncol* (2020) 38(15_suppl):e15252–2. doi: 10.1200/JCO.2020.38.15_suppl.e15252
28. Verver D, Veldt A, Akkooi A, Verhoef C, Grünhagen D, Louwman W. Treatment of melanoma of unknown primary in the era of immunotherapy and targeted therapy: A Dutch population-based study. *Int J Cancer* (2020) 146(1):26–34. doi: 10.1002/ijc.32229

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 in Genetics

ORIGINAL RESEARCH
published: 09 March 2021
doi: 10.3389/fgene.2021.608311

Serum Uric Acid Levels and Risk of Eight Site-Specific Cancers: A Mendelian Randomization Study

Minxiao Jiang[1†], Liangliang Ren[1†], Songzan Chen[2] and Gonghui Li[1]*

[1]*Department of Urology, Sir Run-Run Shaw Hospital College of Medicine, Zhejiang University, Hangzhou, China,*
[2]*Department of Cardiology, Key Laboratory of Biotherapy of Zhejiang Province, Sir Run-Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China*

The relationship between serum uric acid (UA) levels and cancer risk remains controversial. Here, a two-sample Mendelian randomization analysis was performed to identify a causal effect of serum UA levels on cancer risk. Twenty-six single nucleotide polymorphisms strongly associated with serum UA levels were screened as genetic variants from large-scale meta-analysis data of a genome-wide association study of 110,347 European individuals. Genetic associations with eight common site-specific cancers were subsequently explored. A total of six Mendelian randomization methods were used to estimate the potential effect of serum UA levels on cancer risk, including random effects inverse variance weighting, fix effects inverse variance weighting, MR-Egger, median weighting, mode weighting, and simple mode analysis. Our primary random effects inverse variance weighted analysis revealed that no significant associations with cancers was found (all $p > 0.05$). Sensitivity analyses and additional analyses also showed similar pooled results. In conclusion, no significant causality between serum UA levels and cancer risk was evidenced.

Keywords: Mendelian randomization, uric acid, cancer, risk, causality

OPEN ACCESS

Edited by:
Qinghua Xu,
Canhelp Genomics, China

Reviewed by:
Shuai Yuan,
Karolinska Institutet (KI), Sweden
Xiaoyan Zhou,
Fudan University, China

****Correspondence:***
Gonghui Li
3193119@zju.edu.cn

[†]These authors have contributed equally to this work

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

Received: *20 September 2020*
Accepted: *15 February 2021*
Published: *09 March 2021*

Citation:
Jiang M, Ren L, Chen S and Li G (2021) Serum Uric Acid Levels and Risk of Eight Site-Specific Cancers: A Mendelian Randomization Study. Front. Genet. 12:608311. doi: 10.3389/fgene.2021.608311

INTRODUCTION

Uric acid (UA) is a byproduct of purine metabolism, with both endogenous and exogenous purines degraded to UA by xanthine oxidase (Benn et al., 2018). Serum UA homeostasis is maintained *via* its production and excretion (Maiuolo et al., 2016), the latter in humans being primarily renal and hepatic (Su et al., 2020). Purine-rich diets, alcohol consumption, obesity, and hypertension are considered to be risk factors that lead to elevated serum UA, in turn resulting in hyperuricemia and even gout (Roddy and Choi, 2014; Li et al., 2020). Hyperuricemia is a common chronic illness defined by a serum UA level >7.0 mg/dl among men and >5.7 mg/dl among women. The incidence of hyperuricemia in the United States is 20.2% in men and 20.0% in women (Chen-Xu et al., 2019).

Previous studies have reported UA levels to be associated with the incidence of diabetes, cardiovascular disease, kidney disease, and malignancies (Weiner et al., 2008; Battelli et al., 2016; Wang et al., 2018; Borghi et al., 2020). The precise mechanistic role UA plays in the occurrence of malignancies, however, remains unclear. Conventional observational studies have reported that higher serum UA levels are protective against cancer (Horsfall et al., 2014;

Taghizadeh et al., 2014), while other studies reported higher serum UA levels to increase the risk of a number of malignancies (Strasak et al., 2007; Wang et al., 2015). As observational studies are frequently subject to confounding and a variety of biases, it is difficult to determine whether any causality between serum UA levels and cancer risk exists.

The randomized controlled trial is the gold standard for demonstrating epidemiological causality between exposures and outcomes (Klungel et al., 2004). However, the cost of such trials is high and their strict criteria also produce biases, thus limiting the robustness of results (Evans and Davey Smith, 2015). Mendelian randomization (MR) is a relatively novel and effective analytical method which can reveal causality between exposures and outcomes by considering genetic variants as instrumental variables (Smith and Ebrahim, 2003). Given that genetic variants are randomly distributed, determined at conception, and not associated with other confounders, MR reduces confounding and, to an extent, overcomes reverse causality bias (Emdin et al., 2017).

The role of UA in the pathogenesis of malignancies remains unclear. Here, we designed a two-sample MR study to analyze summary genetic data for the purposes of investigating any potential causal associations of genetically-proxied UA levels and the incidence of eight distinct malignancies.

MATERIALS AND METHODS

Study Design

To identify the potential effect of serum UA levels on cancer risk, we designed a two-sample MR study. Single nucleotide polymorphisms (SNPs) for serum UA levels were selected as instrumental variables from previously published genome-wide association study (GWAS) analyses. Three key assumptions were to be satisfied: first, the SNPs should have been associated with serum UA levels; second, the chosen SNPs should have been independent of confounders; and third, the SNPs should have affected cancer only *via* UA concentrations and could not have a direct correlation (**Figure 1**; Little, 2018).

Exposure Measure

We systematically extracted significant genome-wide SNPs related to serum UA levels from a large-scale GWAS meta-analysis of 110,347 European individuals (49,825 women and 60,522 men). The average age of the participants was 52.12 years. The GWAS data were obtained from the Global Urate Genetics Consortium (GUGC; **Table 1**; Köttgen et al., 2013). A total of 26 SNPs passed our *p*-value threshold of 5×10^{-8}, detailing a 7.0% phenotypic variance in serum UA levels. These genetic variants were pruned for linkage disequilibrium using LD-link (https://ldlink.nci.nih.gov/) with an r^2 threshold of 0.01 (Junqueira et al., 2017). After LD pruning, 26 SNPs remained as genetic instrumental variables to proxy serum UA levels. The average values of serum UA concentrations in these studies were recorded and ranged from 3.9 to 6.1 mg/dl (standard deviation (SD): 0.92–1.68 mg/dl). In addition, the strength of each SNP was evaluated by F-statistic values and the instrument with an F-statistic value larger than 10 was regarded as having strong potential to predict UA levels (Lawlor et al., 2008).

Outcome Measure

Data from eight, large-scale meta-analyses of GWASs studying eight common cancers were used to explore the association of genetically-proxied serum UA levels with risk of malignancy incidence rates; namely bladder, breast, colorectal, lung, prostate, renal cell, skin, and thyroid cancers. The breast cancer outcome dataset was composed of summary genetic data obtained from the Breast Cancer Association Consortium (BCAC) and consisted of a meta-analysis of 11 GWASs (15,748 cases, 18,084 controls) in addition to 41 studies (46,785 cases, 42,892 controls) genotyped on the iCOGs custom array (Michailidou et al., 2015). The prostate cancer dataset consisted of 79,148 cases and 61,106 controls and was obtained from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the

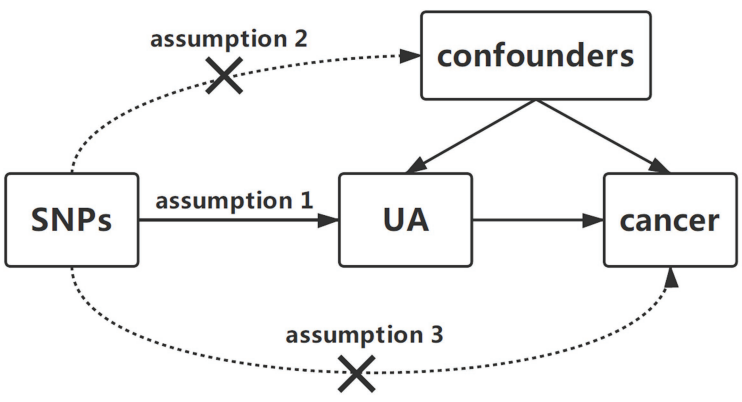


FIGURE 1 | Diagram of two-sample Mendelian randomization analysis of serum uric acid levels and cancer risk. Three key assumptions in the Mendelian randomization analysis are as follows: (1) the SNPs should be related to serum UA levels, (2) the SNPs should be independent of confounders, and (3) the SNPs could affect cancer only by UA. SNP, single-nucleotide polymorphism; UA, uric acid.

TABLE 1 | Characteristics of exposure dataset and outcome datasets.

Phenotype	Consortium	Sample size	Cases	Controls	Ethnicity	References
Serum uric acid	GUGC	110,347			European	Köttgen et al. (2013)
Bladder cancer	UK biobank	361,194	1,367	359,827	European	The Neale lab (2018)
Breast cancer	BCAC	123,509	62,533	60,976	European	Michailidou et al. (2015)
Colorectal cancer	UK biobank	408,458	4,562	382,756	European	Zhou et al. (2018)
Lung cancer	UK biobank	361,194	329	360,865	European	The Neale lab (2018)
Prostate cancer	PRACTICAL	140,254	79,148	61,106	European	Schumacher et al. (2018)
Renal cell cancer	UK biobank	361,194	727	360,467	European	The Neale lab (2018)
Skin cancer	UK biobank	300,791	9,950	290,841	European	Watanabe et al. (2019)
Thyroid cancer	UK biobank	407,757	358	407,399	European	Zhou et al. (2018)

Genome (PRACTICAL; Schumacher et al., 2018). The colorectal and thyroid cancer datasets of GWAS meta-analysis data were obtained from the UK Biobank. The colorectal cancer dataset included 4,562 cases and 382,756 controls while the thyroid cancer dataset included 358 cases and 407,399 controls (Zhou et al., 2018). The skin cancer dataset included 9,950 cases and 290,841 controls; this GWAS meta-analysis was performed by the UK Biobank (Watanabe et al., 2019). Bladder (1,367 cases, 359,827 controls), lung (329 cases, 360,865 controls) and renal cell (727 cases, 360,467 controls) cancer datasets were downloaded from the Neale Lab. All aforementioned GWAS meta-analyses only evaluated participants of European descent (**Table 1**).

Statistical Analysis

In this study, we used five different methods of MR analysis to evaluate the causal effect of serum UA levels on cancer risk. Here, the random-effects inverse variance weighted (IVW) method was used as the primary analysis. The Wald estimator was used to calculate the ratio of the SNP-outcome estimate over the SNP-exposure estimate, while the Delta method was employed to calculate the standard errors (Teumer, 2018). The overall estimate was subsequently obtained by pooling the Wald ratio estimates of each SNP weighted by inverse variances of the SNP-outcome associations (Little, 2018). As this method assumes that the intercept is constrained to the origin [0,0], the presence of horizontal pleiotropy makes this method susceptible to bias. To supplement calculations, we used MR-Egger regression, where intercept and slope represent the average horizontal pleiotropy and the pleiotropy-adjusted MR estimate, respectively. In addition, we utilized weighted median analysis to estimate the effects of all MR estimates that every individual instrument was weighted equally to the inverse of the standard error. Weighted median analysis served as an important method of estimating the causal effect if over 50% of SNPs met the "no horizontal pleiotropy" assumption (Burgess et al., 2017). Finally, weighted mode and simple mode analyses were used to estimate the causal effect (Hemani et al., 2018a).

For the individual variants in the genetic instrumental variable model for serum UA levels, we examined whether some SNPs had a significantly independent influence on results *via* leave-one-out sensitivity analysis. The remaining estimate effect was shown when one SNP was excluded (Hemani et al., 2018b). Cochran's Q statistics were used to estimate the level of heterogeneity.

In addition, we searched traits associated with all 26 SNPs on the PhenoScanner website.[1] After excluding SNPs that were not exclusively associated with UA levels, MR analysis was repeated for the purposes of improving result robustness and deal with potential horizontal pleiotropy.

All statistical analyses in this paper were performed using R software (version 4.0.2; http://www.rproject.org) with the "TwoSampleMR" package (version 0.5.4). Results were considered to show strong evidence of an association between serum UA levels and cancer incidence if they surpassed a stringent Bonferroni-corrected p-value threshold of 6.25×10^{-3} (0.05/8 cancer outcomes).

RESULTS

Here, 26 SNPs strongly related to serum UA levels were extracted from a GWAS meta-analysis based on GUGC data ($p < 5 \times 10^{-8}$; **Supplementary Table S1**). No linkage disequilibrium ($r^2 < 0.01$) was observed. The minimum F-statistic value of these 26 SNPs was 30.05, suggesting that they were sufficiently effective in this study. All SNPs could thus be used to identify the potential effect of serum UA levels on cancer risk. Scatter plots were shown in the supplementary materials (**Supplementary Figures S1–S8a**).

Using a genetic instrumental variable for serum UA levels consisting of 26 SNPs, we estimated the association of serum UA levels against the incidence of eight distinct cancers *via* MR analysis. Associations with individual cancers were described below.

Effect of Uric Acid on Cancers

The main MR results detailing the influence of UA levels on cancers were obtained using random-effects IVW methodology (**Figure 2**). Our primary results did not reveal any association between serum UA levels and the risk of any other cancer type (all $p > 0.05$).

The MR estimates of UA levels on cancer risk obtained using other sensitivity MR approaches are shown in the supplementary materials (**Supplementary Table S2**). Consistent results using MR-Egger, weighted median, weighted mode, and simple mode analyses were not obtained regarding the risk of all cancer types.

[1]http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner

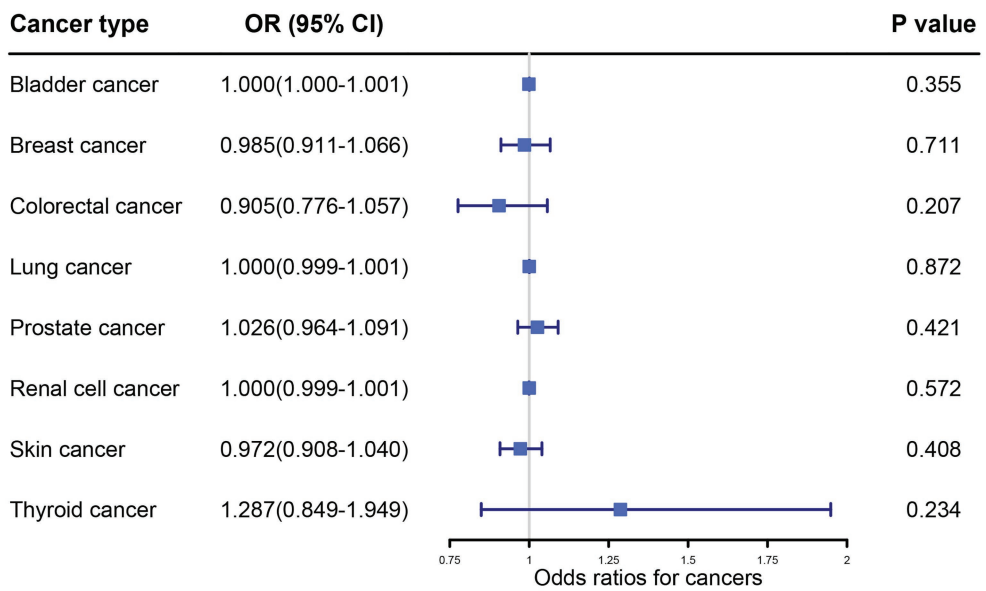


FIGURE 2 | Primary results of the causal associations between serum uric acid levels and cancer risk by random effects inverse variance weighted method. OR, odds ratio; IVW, inverse variance weighted; CI, confidence interval.

In addition, leave-one-out sensitivity analysis was performed and revealed that no SNP could independently drive MR analysis results for most of cancers (**Supplementary Figures S1–S8b**). However, it was observed that rs12498742 had a significant effect on the results for colorectal and skin cancers. Therefore, we performed analysis based on SNPs excluding the rs12498742. Similarly, no causal effect was observed in the results after Bonferroni correction (**Supplementary Figure S9**).

Other Analysis

The results of MR-Egger regression for the assessment of pleiotropy are listed in **Supplementary Table S3**, suggesting that the non-pleiotropy assumption was satisfied in all of the aforementioned MR methods for most cancers. However, we detected significant pleiotropy when testing for a causal effect of UA on skin cancer risk (intercept = −0.01, p = 0.03). In addition, some evidence of heterogeneity was also found using Cochran's Q statistics for some cancers, including breast, colorectal and prostate cancer (**Supplementary Table S4**). To deal with heterogeneity and potential horizontal pleiotropy, all 26 SNPs were searched on the PhenoScanner website and six SNPs were found to be exclusively associated with serum UA levels (**Supplementary Table S5**). The entire analysis was subsequently repeated using these six SNPs as instruments. Similarly, no significant causal relationship was observed between serum UA levels and the eight site-specific cancers in question (**Supplementary Table S6**). These results strongly suggested that the observed associations were not biased by pleiotropic effects (**Supplementary Table S3**). Results of heterogeneity testing also revealed a significant decrease of heterogeneity after excluding SNPs associated with the phenomenon, apart from serum UA levels (**Supplementary Table S4**).

DISCUSSION

This study explored the relationship between serum UA levels and cancer risk *via* a two-sample MR analysis and did not identify strong evidence supporting causality between serum UA levels and cancer risk, including that of bladder, breast, colorectal, lung, prostate, renal cell, skin, and thyroid cancer. The sensitivity analyses and other analyses supported these findings.

Ames et al. first reported that serum UA was an excellent scavenger of singlet oxygen and hydroxyl radicals, and could be a protective factor against cancer in humans (Ames et al., 1981). Evidence for the antioxidant function of UA has continued to increase over the recent decades (Peden et al., 1990; Becker, 1993; Liu et al., 2019). Itahana et al. demonstrated that SLC2A9 was a key downstream target of p53, already well known as a typical UA transporter. This pathway was found to be protective from ROS-induced damage and cancer pathogenesis in humans (Itahana et al., 2015). In contrast, UA levels are also regarded to be a risk factor for cancer due to its function in inducing chronic inflammation and increasing ROS production (Mi et al., 2020). Chronic inflammation and tissue infiltration by neutrophils, macrophages and monocytes (Grainger et al., 2013; Weigt et al., 2017), in turn, promote carcinogenesis (Fini et al., 2012; Braga et al., 2017; Ahechu et al., 2018). Due to the complex roles UA plays in cancer occurrence, associations remain unconfirmed in previous epidemiological literature (Strasak et al., 2007; Dziaman et al., 2014; Horsfall et al., 2014; Szkandera et al., 2015; Battelli et al., 2016; Mi et al., 2020).

A recently published prospective population-based study demonstrated associations between serum UA levels and the risks of common cancers (Kuhn et al., 2017). Consistent with our findings, serum UA levels were reported not to be associated

with risks of lung, prostate, and colorectal cancer. In addition, higher serum UA levels were associated with lower breast cancer risk. The implications of these results, however, were limited due to the number of cases in this cohort study (lung cancer, n = 195; colorectal cancer, n = 256; breast cancer, n = 627; prostate cancer, n = 554). One MR study also revealed causal relationships between serum UA levels and cancer risks. The study was based on the Copenhagen General Population Study and selected rs7442295 (gene: SLC2A9, position: chr4:9964756) as the instrument for serum UA level representation. Results revealed causal relationships between high serum UA levels, high cancer incidence and high all-cause mortality, contrary to our findings. Possible explanations of such a paradoxical finding were that this study only used rs7442295 as the instrument which could only explain 2% of the variation in serum UA levels, and SLC2A9 was expressed differently in various organs, thus there might some biases in that study (Kobylecki et al., 2017). Li et al. additionally reviewed the relationship between serum UA levels and multiple health outcomes recently. Similarly, the review did not identify convincing evidence supporting a clear influence of serum UA levels on cancer outcomes (Li et al., 2017).

Here, we designed an MR study to investigate potential causality between serum UA levels and malignancy risk. To meet the first assumption, the genetic instrument chosen in the MR study should be strongly associated with serum UA levels. A previous study has reported that almost 40–80% of variation of serum UA levels could be explained by genetic factors (Krishnan et al., 2012), while the strength of genetic instruments used in MR studies was still small and accounted for only 7% of serum UA variance. The power of genetic instruments to detect causal associations with serum UA levels was thereby limited. Nevertheless, all SNPs chosen in this study passed our p-value threshold of 5×10^{-8} and F-statistics values threshold of 10, indicating all instruments were sufficiently effective. Sensitivity analyses, including five MR methods and pleiotropy analyses, were subsequently carried out to evaluate for potential violation of the second assumption. Results revealed some evidence for the existence of horizontal pleiotropy and heterogeneity in the analysis by using 26 SNPs, indicting this assumption may be violated. After exclusion of the SNPs not exclusively associated with serum UA levels, we found the heterogeneity and pleiotropy decreased significantly and the results remained unchanged, which proved the robustness of the results. In addition, no SNP was found to be associated directly with cancers. Therefore, the likelihood of biases in this paper is low.

This study has several advantages. First, almost all prior studies were observational and incorporated a limited quantity of patients, thus likely causing observation bias and increasing the risk of confounding. Furthermore, few studies have demonstrated potential causality between serum UA levels and cancer risk. Our study used a novel method, MR analysis, to assess any potential causal relationship, thereby minimizing confounding and overcoming reverse causality. The two-sample MR method also allowed us to integrate several independent GWAS datasets with large sample sizes and yield more precise results. To the best of our knowledge, this study was the largest such MR analysis focused on the relationship between serum UA levels and malignancy. Moreover, six different MR methods were employed in this study, thus increasing result robustness.

Our study, however, was not without limitations. First, the proportions of cases for some site-specific cancers were low, and it might result in a low precision of the estimates (Zhou et al., 2018). Second, data for most cancers was downloaded from the UK Biobank. The individuals in the UK Biobank are healthier than the general population, and we cannot rule out the "healthy volunteer" selection bias (Fry et al., 2017). Third, due to a lack of detailed information in the datasets, we were unable to conduct more refined analyses (e.g., stratification analysis). In addition, the GWAS datasets in this study only contained data from European individuals. Our findings may thus not be applicable to other races. Future studies should evaluate patients from different ethnicities and in wider age ranges.

CONCLUSION

In conclusion, we did not find any consistently strong evidence supporting causality between serum UA levels and cancer risk. However, the potential causal role of serum UA levels in the risk of malignancy warrants further investigation by studying a greater number of cancer types and employing larger MR analyses.

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

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

MJ and GL designed the study. MJ and LR assembled and analyzed the data. LR and SC visualized the data. MJ, LR, and SC wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Joint construction project of Zhejiang Province and Ministry (grant number: 2020388200) and Key R&D plan of Zhejiang Province (grant number: 2019C03089).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.608311/full#supplementary-material

REFERENCES

Ahechu, P., Zozaya, G., Marti, P., Hernandez-Lizoain, J. L., Baixauli, J., Unamuno, X., et al. (2018). NLRP3 Inflammasome: a possible link between obesity-associated low-grade chronic inflammation and colorectal cancer development. *Front. Immunol.* 9:2918. doi: 10.3389/fimmu.2018.02918

Ames, B. N., Cathcart, R., Schwiers, E., and Hochstein, P. (1981). Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 78, 6858–6862. doi: 10.1073/pnas.78.11.6858

Battelli, M. G., Polito, L., Bortolotti, M., and Bolognesi, A. (2016). Xanthine oxidoreductase in cancer: more than a differentiation marker. *Cancer Med.* 5, 546–557. doi: 10.1002/cam4.601

Becker, B. F. (1993). Towards the physiological function of uric acid. *Free Radic. Biol. Med.* 14, 615–631. doi: 10.1016/0891-5849(93)90143-I

Benn, C. L., Dua, P., Gurrell, R., Loudon, P., Pike, A., Storer, R. I., et al. (2018). Physiology of hyperuricemia and urate-lowering treatments. *Front. Med.* 5:160. doi: 10.3389/fmed.2018.00160

Borghi, C., Agabiti-Rosei, E., Johnson, R. J., Kielstein, J. T., Lurbe, E., Mancia, G., et al. (2020). Hyperuricaemia and gout in cardiovascular, metabolic and kidney disease. *Eur. J. Intern. Med.* 80, 1–11. doi: 10.1016/j.ejim.2020.07.006

Braga, T. T., Forni, M. F., Correa-Costa, M., Ramos, R. N., Barbuto, J. A., Branco, P., et al. (2017). Soluble uric acid activates the NLRP3 Inflammasome. *Sci. Rep.* 7:39884. doi: 10.1038/srep39884

Burgess, S., Bowden, J., Fall, T., Ingelsson, E., and Thompson, S. G. (2017). Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* 28, 30–42. doi: 10.1097/EDE.0000000000000559

Chen-Xu, M., Yokose, C., Rai, S. K., Pillinger, M. H., and Choi, H. K. (2019). Contemporary prevalence of gout and hyperuricemia in the United States and decadal trends: the national health and nutrition examination survey, 2007-2016. *Arthritis Rheum.* 71, 991–999. doi: 10.1002/art.40807

Dziaman, T., Banaszkiewicz, Z., Roszkowski, K., Gackowski, D., Wisniewska, E., Rozalski, R., et al. (2014). 8-Oxo-7,8-dihydroguanine and uric acid as efficient predictors of survival in colon cancer patients. *Int. J. Cancer* 134, 376–383. doi: 10.1002/ijc.28374

Emdin, C. A., Khera, A. V., and Kathiresan, S. (2017). Mendelian randomization. *JAMA* 318, 1925–1926. doi: 10.1001/jama.2017.17219

Evans, D. M., and Davey Smith, G. (2015). Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu. Rev. Genomics Hum. Genet.* 16, 327–350. doi: 10.1146/annurev-genom-090314-050016

Fini, M. A., Elias, A., Johnson, R. J., and Wright, R. M. (2012). Contribution of uric acid to cancer risk, recurrence, and mortality. *Clin. Transl. Med.* 1:16. doi: 10.1186/2001-1326-1-16

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., et al. (2017). Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am. J. Epidemiol.* 186, 1026–1034. doi: 10.1093/aje/kwx246

Grainger, R., McLaughlin, R. J., Harrison, A. A., and Harper, J. L. (2013). Hyperuricaemia elevates circulating CCL2 levels and primes monocyte trafficking in subjects with inter-critical gout. *Rheumatology* 52, 1018–1021. doi: 10.1093/rheumatology/kes326

Hemani, G., Bowden, J., and Davey Smith, G. (2018a). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* 27, R195–R208. doi: 10.1093/hmg/ddy163

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., et al. (2018b). The MR-base platform supports systematic causal inference across the human phenome. *elife* 7:e34408. doi: 10.7554/eLife.34408

Horsfall, L. J., Nazareth, I., and Petersen, I. (2014). Serum uric acid and the risk of respiratory disease: a population-based cohort study. *Thorax* 69, 1021–1026. doi: 10.1136/thoraxjnl-2014-205271

Itahana, Y., Han, R., Barbier, S., Lei, Z., Rozen, S., and Itahana, K. (2015). The uric acid transporter SLC2A9 is a direct target gene of the tumor suppressor p53 contributing to antioxidant defense. *Oncogene* 34, 1799–1810. doi: 10.1038/onc.2014.119

Junqueira, S. C., Dos Santos Coelho, I., Lieberknecht, V., Cunha, M. P., Calixto, J. B., Rodrigues, A. L. S., et al. (2017). Inosine, an endogenous purine nucleoside, suppresses immune responses and protects mice from experimental autoimmune encephalomyelitis: a role for A2A adenosine receptor. *Mol. Neurobiol.* 54, 3271–3285. doi: 10.1007/s12035-016-9893-3

Klungel, O. H., Martens, E. P., Psaty, B. M., Grobbee, D. E., Sullivan, S. D., Stricker, B. H., et al. (2004). Methods to assess intended effects of drug treatment in observational studies are reviewed. *J. Clin. Epidemiol.* 57, 1223–1231. doi: 10.1016/j.jclinepi.2004.03.011

Kobylecki, C. J., Afzal, S., and Nordestgaard, B. G. (2017). Plasma urate, cancer incidence, and all-cause mortality: a Mendelian randomization study. *Clin. Chem.* 63, 1151–1160. doi: 10.1373/clinchem.2016.268185

Köttgen, A., Albrecht, E., Teumer, A., Vitart, V., Krumsiek, J., Hundertmark, C., et al. (2013). Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* 45, 145–154. doi: 10.1038/ng.2500

Krishnan, E., Lessov-Schlaggar, C. N., Krasnow, R. E., and Swan, G. E. (2012). Nature versus nurture in gout: a twin study. *Am. J. Med.* 125, 499–504. doi: 10.1016/j.amjmed.2011.11.010

Kuhn, T., Sookthai, D., Graf, M. E., Schubel, R., Freisling, H., Johnson, T., et al. (2017). ALbumin, bilirubin, uric acid and cancer risk: results from a prospective population-based study. *Br. J. Cancer* 117, 1572–1579. doi: 10.1038/bjc.2017.313.

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., and Smith, G. D. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* 27, 1133–1163. doi: 10.1002/sim.3034

Li, X., Meng, X., Timofeeva, M., Tzoulaki, I., Tsilidis, K. K., Ioannidis, J. P., et al. (2017). Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* 357:j2376. doi: 10.1136/bmj.j2376

Li, L., Zhang, Y., and Zeng, C. (2020). Update on the epidemiology, genetics, and therapeutic options of hyperuricemia. *Am. J. Transl. Res.* 12, 3167–3181.

Little, M. (2018). Mendelian randomization: methods for using genetic variants in causal estimation. *J. Roy. Stat. Soc. Stat. Soc.* 181, 549–550. doi: 10.1111/rssa.12343

Liu, D. Q., Yun, Y., Yang, D. C., Hu, X. Y., Dong, X. X., Zhang, N., et al. (2019). What is the biological function of uric acid? An antioxidant for neural protection or a biomarker for cell death. *Dis. Markers* 2019, 1–9. doi: 10.1155/2019/4081962

Maiuolo, J., Oppedisano, F., Gratteri, S., Muscoli, C., and Mollace, V. (2016). Regulation of uric acid metabolism and excretion. *Int. J. Cardiol.* 213, 8–14. doi: 10.1016/j.ijcard.2015.08.109

Mi, S., Gong, L., and Sui, Z. (2020). Friend or foe? An unrecognized role of uric acid in cancer development and the potential anticancer effects of uric acid-lowering drugs. *J. Cancer* 11, 5236–5244. doi: 10.7150/jca.46200

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., et al. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* 47, 373–U127. doi: 10.1038/ng.3242

Peden, D. B., Hohman, R., Brown, M. E., Mason, R. T., Berkebile, C., Fales, H. M., et al. (1990). Uric-acid is a major antioxidant in human nasal airway secretions. *Proc. Natl. Acad. Sci. U. S. A.* 87, 7638–7642. doi: 10.1073/pnas.87.19.7638

Roddy, E., and Choi, H. K. (2014). Epidemiology of gout. *Rheum. Dis. Clin. N. Am.* 40, 155–175. doi: 10.1016/j.rdc.2014.01.001

Schumacher, F. R., Al Olama, A. A., Berndt, S. I., Benlloch, S., Ahmed, M., Saunders, E. J., et al. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* 50, 928–936. doi: 10.1038/s41588-018-0142-8

Smith, G. D., and Ebrahim, S. (2003). Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. doi: 10.1093/ije/dyg070

Strasak, A. M., Rapp, K., Hilbe, W., Oberaigner, W., Ruttmann, E., Concin, H., et al. (2007). The role of serum uric acid as an antioxidant protecting against cancer: prospective study in more than 28000 older Austrian women. *Ann. Oncol.* 18, 1893–1897. doi: 10.1093/annonc/mdm338

Su, H. Y., Yang, C., Liang, D., and Liu, H. F. (2020). Research advances in the mechanisms of hyperuricemia-induced renal injury. *Biomed. Res. Int.* 2020, 1–12. doi: 10.1155/2020/5817348

Szkandera, J., Gerger, A., Liegl-Atzwanger, B., Stotz, M., Samonigg, H., Ploner, F., et al. (2015). Uric acid levels in blood are associated with clinical outcome in soft-tissue sarcoma patients. *Clin. Chem. Lab. Med.* 53, 493–497. doi: 10.1515/cclm-2014-0486

Taghizadeh, N., Vonk, J. M., and Boezen, H. M. (2014). Serum uric acid levels and cancer mortality risk among males in a large general population-based cohort study. *Cancer Causes Control* 25, 1075–1080. doi: 10.1007/s10552-014-0408-0

Teumer, A. (2018). Common methods for performing Mendelian randomization. *Front. Cardiovasc. Med.* 5:51. doi: 10.3389/fcvm.2018.00051

The Neale lab (2018). http://www.nealelab.is/uk-biobank/ (Accessed August 15, 2020).

Wang, W., Xu, D., Wang, B., Yan, S., Wang, X., Yin, Y., et al. (2015). Increased risk of cancer in relation to gout: a review of three prospective cohort studies with 50,358 subjects. *Mediat. Inflamm.* 2015:680853. doi: 10.1155/2015/680853

Wang, H., Zhang, H., Sun, L., and Guo, W. (2018). Roles of hyperuricemia in metabolic syndrome and cardiac-kidney-vascular system diseases. *Am. J. Transl. Res.* 10, 2749–2763.

Watanabe, K., Stringer, S., Frei, O., Umicevic Mirkov, M., de Leeuw, C., Polderman, T. J. C., et al. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* 51, 1339–1348. doi: 10.1038/s41588-019-0481-0

Weigt, S. S., Palchevskiy, V., and Belperio, J. A. (2017). Inflammasomes and IL-1 biology in the pathogenesis of allograft dysfunction. *J. Clin. Invest.* 127, 2022–2029. doi: 10.1172/JCI93537

Weiner, D. E., Tighiouart, H., Elsayed, E. F., Griffith, J. L., Salem, D. N., and Levey, A. S. (2008). Uric acid and incident kidney disease in the community. *J. Am. Soc. Nephrol.* 19, 1204–1211. doi: 10.1681/ASN.2007101075

Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* 50, 1335–1341. doi: 10.1038/s41588-018-0184-y

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 in Genetics

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fgene.2021.650416

Ferroptosis-Related Gene-Based Prognostic Model and Immune Infiltration in Clear Cell Renal Cell Carcinoma

Guo-Jiang Zhao, Zonglong Wu, Liyuan Ge, Feilong Yang, Kai Hong, Shudong Zhang and Lulin Ma**

Department of Urology, Peking University Third Hospital, Beijing, China

Clear cell renal cell carcinoma (ccRCC) is one of the most common tumors in the urinary system. Ferroptosis plays a vital role in ccRCC development and progression. We did an update of ferroptosis-related multigene expression signature for individualized prognosis prediction in patients with ccRCC. Differentially expressed ferroptosis-related genes in ccRCC and normal samples were screened using The Cancer Genome Atlas. Univariate and multivariate Cox regression analyses and machine learning methods were employed to identify optimal prognosis-related genes. *CARS1*, *CD44*, *FANCD2*, *HMGCR*, *NCOA4*, *SLC7A11*, and *ACACA* were selected to establish a prognostic risk score model. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses revealed that these genes were mainly enriched in immune-related pathways; single-sample Gene Set Enrichment Analysis revealed several immune cells potentially related to ferroptosis. Kaplan–Meier survival analysis demonstrated that patients with high-risk scores had significantly poor overall survival (log-rank $P = 7.815 \times 10^{-11}$). The ferroptosis signature was identified as an independent prognostic factor. Finally, a prognostic nomogram, including the ferroptosis signature, age, histological grade, and stage status, was constructed. Analysis of The Cancer Genome Atlas-based calibration plots, C-index, and decision curve indicated the excellent predictive performance of the nomogram. The ferroptosis-related seven-gene risk score model is useful as a prognostic biomarker and suggests therapeutic targets for ccRCC. The prognostic nomogram may assist in individualized survival prediction and improve treatment strategies.

Keywords: ferroptosis, clear cell renal cell carcinoma, risk score, The Cancer Genome Atlas, nomogram, machine learning

OPEN ACCESS

Edited by:
Yi Ba,
Tianjin Medical University, China

Reviewed by:
Chin-Yo Lin,
University of Houston, United States
Xiao-Qing Li,
Tianjin Medical University Cancer Institute and Hospital, China

****Correspondence:***
Lulin Ma
doctor.lulin.ma@gmail.com
Shudong Zhang
shootong@163.com

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

Received: *07 January 2021*
Accepted: *27 April 2021*
Published: *09 June 2021*

Citation:
Zhao G-J, Wu Z, Ge L, Yang F, Hong K, Zhang S and Ma L (2021) Ferroptosis-Related Gene-Based Prognostic Model and Immune Infiltration in Clear Cell Renal Cell Carcinoma. Front. Genet. 12:650416. doi: 10.3389/fgene.2021.650416

INTRODUCTION

Renal cell carcinoma (RCC), which occurs in the proximal convoluted tubule, is among the top 20 most common cancers worldwide, affecting more than 400,000 individuals each year and accounting for 2.2% of malignant tumors among all new cancer cases (Bray et al., 2018). There are three main types of RCC: clear cell (ccRCC), papillary (types I and II), and chromophobe

(Frew and Moch, 2015; Moch et al., 2016), with ccRCC accounting for approximately 70% of RCC cases (Frew and Moch, 2015; Moch et al., 2016). Compared with papillary RCC and chromophobe RCC, ccRCC shows the worst prognosis according to univariate analysis (Capitanio et al., 2009). This type of renal cell carcinoma is called clear cell because most of the lipids and glycogen are dissolved in conventional histological treatments and are vacuoles under a microscope. The main risk factors for RCC are tobacco smoking, overweight or obesity, and hypertension (Frew and Moch, 2015). Deletion of *PBRM1*, *SETD2*, and *BAP1* on the 3p chromosome and mutation or deletion of *VHL* are also important driving factors for ccRCC (Mitchell et al., 2018; Turajlic et al., 2018). Loss of *VHL* leads to the accumulation of HIF, which inhibits the metabolism of fatty acids in ccRCC and leads to the accumulation of lipids (Du et al., 2017). Although tumor resection in the early stage shows beneficial effects for patients, because of the lack of early symptoms, around 25–30% of patients with RCC are diagnosed in stages III and IV, making resection difficult (Keegan et al., 2012; Capitanio and Montorsi, 2016). The 5-year survival rate of patients with metastatic RCC is only approximately 20%[1], and ccRCC is not sensitive to radiotherapy and chemotherapy. Therefore, methods for identifying key biological markers are urgently needed to facilitate the early diagnosis of ccRCC and development of new drugs for these biological markers.

Programmed cell death (PCD) is a normal physiological process occurring in cells, and ferroptosis is a type of programmed cell death. Ferroptosis, which can be blocked by iron chelators, was first described in 2012 (Dixon et al., 2012) and then defined by the Nomenclature Committee on Cell Death as a form of PCD in 2018 (Galluzzi et al., 2018). This process differs from apoptosis, necrosis, autophagy, and other forms of regulatory cell necrosis at the morphological, biological, and gene levels. Ferroptosis is induced by the accumulation of ferric ion (Hirschhorn and Stockwell, 2019). Morphologically, according to electron microscopy analysis, iron death leads to a decrease in mitochondria and increase in membrane density (Dixon et al., 2012). The small molecule RSL3 can induce cell death, which cannot be inhibited by necrosis inhibitors such as caspase inhibitors and necrostatins (Dixon et al., 2012). When the cell cystine transporter protein is inhibited (e.g., erastin), intracellular glutathione is depleted, eventually leading to inactivation of glutathione peroxidase (GPX4) and resulting in lipid peroxidation accumulation; these factors induce iron-induced cell death (Stockwell et al., 2017). Ferroptosis helps maintain cell-death-related homeostasis in normal cells and tissues and the homeostasis of some carcinoma gene pathways. In people with iron metabolism disorder, Fe^{3+} accumulates by binding to transferrin (TF) and enters the cell through transferrin receptor-mediated endosomes. Fe^{3+} is reduced to Fe^{2+} iron by STEAP3 metalloreductase in the endosome and is then released into the intracellular environment via solute carrier family 11 member 2. After entering the cytoplasm, Fe^{2+} is stored by ferritin, which is formed by ferritin light chain and ferritin heavy chain 1. Solute carrier family 40 member 1 (also known as ferroportin-1) acts as pump to transport iron ions out of the cell (Liu et al., 2020). Increasing the release of stored intracellular iron into the cytoplasm and reducing the discharge of iron ions can increase the intracellular concentration of these ions, inducing intracellular reactive oxygen species and leading to ferroptosis (Dixon et al., 2012; Bogdan et al., 2016; Hou et al., 2016). Ferroptosis is mainly regulated by two parallel pathways involved in lipid peroxidation (Wu Y. N. et al., 2020): that involving glutathione/GPX4 (Stockwell et al., 2017) and another involving ferroptosis suppressor protein 1, ubiquinone (CoQ10), and NAD(P)H (Bersuker et al., 2019; Doll et al., 2019). If either of these systems is inactivated, Reactive oxygen species (ROS), which include superoxide and hydroxyl radicals, become active. In the cancer cells, some ferroptosis key genes are altered, reducing programmed cell death (Wu G. et al., 2020). A study had shown that glutamine can inhibit oxidation in clear cell renal cell carcinoma by ferroptosis pathway (Abu Aboud et al., 2017).

In previous research, different studies from different perspectives have explored the value of prognostic markers in ccRCC, like long non-coding RNAs (Cheng et al., 2019; Yang et al., 2020), some key genes (Yang et al., 2018; Luo et al., 2019), immune-related genes (Zhang et al., 2019), etc. Some studies have focused on the relationship between ferroptosis and ccRCC (Abu Aboud et al., 2017; Wu G. et al., 2020; Mou et al., 2021; Wang et al., 2021). Wu and colleagues did a remarkable work in clinical significance of ferroptosis-related genes (FRGs) in pan cancer and conducted a new survival model based on five risk-related FRGs for ccRCC (Wu G. et al., 2020). With the development of research, more and more genes related to ferroptosis have been found. In our study, we included more FRGs and updated FRGs survival model. Furthermore, we analyzed these genes by ssGSEA to explore the role of these genes in the immune microenvironment of ccRCC first. Considering the pathological features and the metabolic characteristics of lipid and glutamine of renal clear cell carcinoma, we tried to explore the role of ferroptosis pathway in ccRCC. Therefore, we updated on measuring ferroptosis-associated biomarkers for predicting the survival of patients with ccRCC. Furthermore, in this study, we explored the relationships between ferroptosis and the immune microenvironment in patients with ccRCC. We used these hub genes to establish a nomogram for clinical use to improve individualized prognosis assessment. By building and validating the nomogram, the prognosis prediction of patients with ccRCC in the clinic can be improved.

MATERIALS AND METHODS

Acquisition of Transcriptome and Clinical Data

We downloaded transcriptome profiles from the publicly available The Cancer Genome Atlas (TCGA) database via the GDC data portal[2], with the HTSeq-FPKM workflow type of all available ccRCC samples compared with normal tissues. Corresponding clinical information was obtained from the GDC

[1]https://seer.cancer.gov/

[2]https://portal.gdc.cancer.gov/

portal, including age, gender, tumor grade, stage, and survival outcomes (updated by September 2020). According to TCGA data access policies and publication guidelines, there was no ethical conflict to declare, and this study did not require the approval of the ethics committee.

Differential Ferroptosis-Related Genes Analysis and Survival Analysis

We evaluated 60 ferroptosis-related genes (Chen et al., 2020; Elgendy et al., 2020). "Limma" was used to identify differentially expressed genes (DEGs) in two groups, normal samples and tumor samples, with a false discovery rate <0.05. The "org.Hs.eg.db" package was used to obtain the Entrez ID for each DEG, and significant enrichment pathways were considered only with a false discovery rate <0.05. We selected FRGs in ccRCC to assess survival outcomes by univariate Cox regression analysis, and the *P* value was shown in the plot determined by "survival" package. Both DEGs and survival-related genes between the two groups were determined with the "VennDiagram" package. Furthermore, a heatmap was drawn by "pheatmap" package, and a Forest plot was drawn by the "survival" package to show the difference between groups.

Determination and Validation of FRGs as Independent Prognostic Factors

Least absolute shrinkage and selection operator (LASSO) analyses were performed to identify key FRGs and establish a clinical prognostic model. Kaplan–Meier survival curves were generated to assess the model based on overall survival (OS). Time-dependent receiver operating characteristic (ROC) curves were plotted to validate the FRG prognostic model. To classify the risk of FRGs in ccRCC, we used the optimal risk cutoff to set the risk score model. The risk score for each patient, with β indicating the regression coefficient, was calculated as follows: risk score = expression level of $FRG_1 \times \beta_1$ + expression level of $FRG_2 \times \beta_2 + \ldots +$ expression level of FRGn × βn (Kidd et al., 2018). According to the median expression levels of prognosis-related genes, patients were divided into high- and low-risk groups. Both principal component analysis for linear data dimension reduction and visualization by t-distributed stochastic neighbor embedding plots for non-linear dimension reduction were used to distinguish high− and low−risk groups. We also tested the expression and prognostic ability of each gene in high− and low−risk groups. Our model was verified by randomly sampling 70% of the total samples as validation set, which can be put back. We attach the verification results to the **Supplementary Materials**.

Molecular Mechanism and Immune Infiltrate Analysis

Gene Ontology (GO) was constructed in 2,000 as a structured standard biological model to build a standard vocabulary of knowledge about genes and their products, covering the cellular component, molecular function, and biological process[3]. The main features of KEGG are to link genes with various biochemical reactions[4], and KEGG is a comprehensive database integrating genomic, chemical, and systematic functional information. GO and KEGG can provide rapid gene function annotation and preliminary functional analysis. Both these methods are very popular for analyzing gene function. We analyzed the functions of FRGs by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) with "clusterProfiler" package (false discovery rate <0.05). Single-sample Gene Set Enrichment Analysis (ssGSEA) (Barbie et al., 2009) is implemented by the extended GSEA extension; ssGSEA allows the definition of an enrichment score that represents the absolute enrichment of the gene set in each sample within a given dataset. Rank normalization of gene expression values for a given sample and generating enrichment scores using the empirical cumulative distribution function (ECDF) of genes in the signature and the remaining genes. ssGSEA was used to calculate the enrichment score of each sample in different gene sets and explore the enrichment signaling pathways in the low− and high−risk groups. ssGSEA was performed to evaluate FRGs with immune activity in ccRCC based on "GSVA" package (Hänzelmann et al., 2013).

Establishment of the Nomogram for Clinical Application

Univariate and multivariate Cox regressions were performed to determine the association between OS and clinical characteristics. According to univariate and multivariate Cox regressions, we built a nomogram (Iasonos et al., 2008) to assess the probability of 1−, 3−, and 5-year OS of patients. To assess the predictive accuracy of this model, the C-index was used to estimate the probability of agreement between the nomogram and actual results. The C-index ranged from 0.5 to 1.0, with 1.0 indicating a perfect capacity to correctly distinguish the outcome with the model and 0.5 indicating random chance. Calibration plots were used to test the discriminative ability of the nomogram. Finally, decision curve analysis was performed to evaluate the clinical benefits of the nomogram.

Statistical Analysis

All statistical analyses were performed using R language (version 4.0.2). A two-sided *P* value of <0.05 was considered as statistically significant. The adjusted *P* value was determined by the Benjamini–Hochberg method.

RESULTS

Differentially Expressed FRGs and Survival Analysis

To explore the prognostic value of 60 FRGs (Stockwell et al., 2017; Chen et al., 2020) in ccRCC, 72 normal and 539 carcinoma samples were evaluated. Among them, 28 FRGs were upregulated and 22 FRGs were downregulated

[3]http://geneontology.org/

[4]https://www.kegg.jp/

(adjusted $P < 0.05$). Univariate Cox regression analysis was conducted, and 36 of these 60 FRGs were found to be related to the survival of patients with ccRCC. The intersection of the FRGs based the DEGs and FRG-based survival-related genes was determined by drawing a Venn diagram (**Figure 1A**). Twenty-eight FRGs were selected for further analysis, with their expression displayed in a heatmap (**Figure 1B**). The hazard ratio, 95% confidence interval (CI), and P value of each key FRG are illustrated in a forest plot (**Figure 1C**).

ROC Curve Indicated Good Performance for the Seven FRGs in the Risk Score Model for Predicting the OS of Patients With ccRCC

To more efficiently identify hub genes, a machine learning method (LASSO) was used. Seven FRGs – cysteinyl-tRNA synthetase 1 (*CARS1*), cluster of differentiation 44 (*CD44*), Fanconi anemia complementation group D2 (*FANCD2*), 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGCR*), nuclear

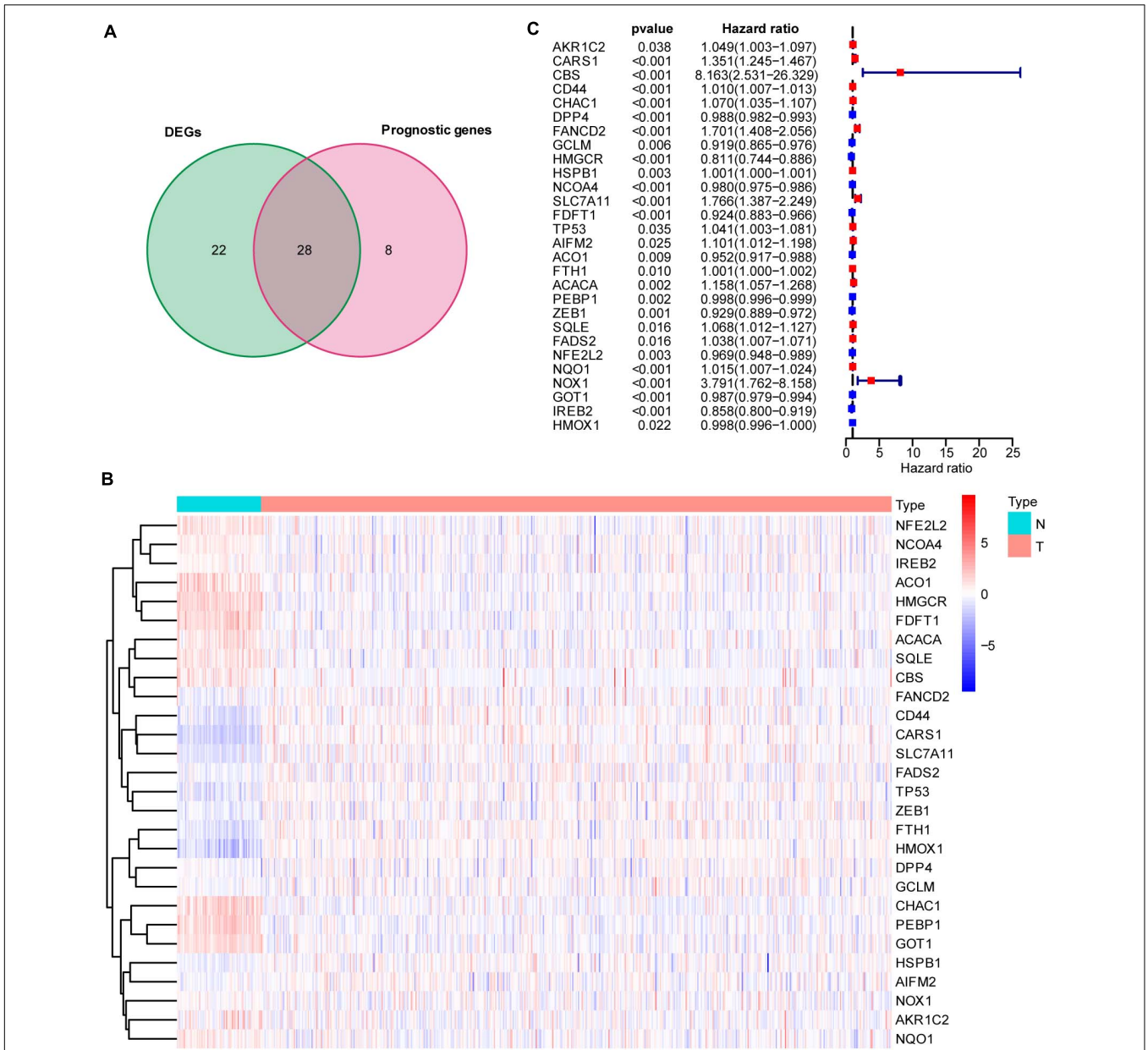


FIGURE 1 | Intersection of differentially expressed ferroptosis-related genes (FRGs) and survival-associated ferroptosis-related genes. **(A)** Venn diagram showing the 28 FRGs associated with survival [intersection of the differentially expressed genes (DEGs) and prognostic genes]. **(B)** Heatmap showing differentially expressed ferroptosis-related genes between normal and tumor samples (adjusted $P < 0.05$). **(C)** Hazard ratio (HR) and 95% CI of the 28 key FRGs in univariate Cox regression. A total of 72 normal and 539 carcinoma samples were included in the analysis.

receptor coactivator 4 (*NCOA4*), solute carrier family 7 member 11 (*SLC7A11*), and acetyl-CoA carboxylase alpha (*ACACA*) – were selected to build a ferroptosis signature model based on minimum criteria (**Figures 2A,B**). We also made a heatmap and a survival status plot of these seven genes (**Figures 3E,F**). According the median risk score, we divided patients with ccRCC into low− and high−risk groups and built a new survival model comprised of the seven FRGs (**Figures 2C,D**). The equation of the FRG-based prognostic signature model was as follows: risk score = (0.1581 × expression value of *CARS1*) + (0.0040 × expression value of *CD44*) + (0.1968 × expression value of *FANCD2*) + (−.0464 × expression value of *HMGCR*) + (−0.0091 × expression value of *NCOA4*) + (0.0225 × expression value of *SLC7A11*) + (0.0352 × expression value of *ACACA*). To investigate the prognostic prediction performance, Kaplan–Meier survival curves were used. Patients in the high-risk group showed significantly shorter survival rates than those in the low-risk group (**Figure 3A**). The heatmaps showed the difference from the high− and low−risk groups (**Figure 3E**). Meanwhile, the outcomes of Kaplan–Meier survival curves of the seven genes in ccRCC also showed that the low-risk group have a better prognosis result (**Figure 3F**). ROC curve analysis performed to determine the prognostic prediction performance of the new survival model in patients with ccRCC revealed area under the curve (AUC) scores of 0.780, 0.736, and 0.747, for 1−, 3−, and 5-year survival, respectively (**Figure 3B**). As a novel prognostic factor, the FRG-based risk score can effectively help to predict the survival status of patients with ccRCC. Furthermore, we applied two widely applied dimension reduction methods, principal component analysis for unsupervised learning, and linear and t-distributed stochastic neighbor embedding for non-linear dimension reduction. These two plots revealed good distinction between the high− and low−risk groups (**Figures 3C,D**). Validation set also showed that the low-risk

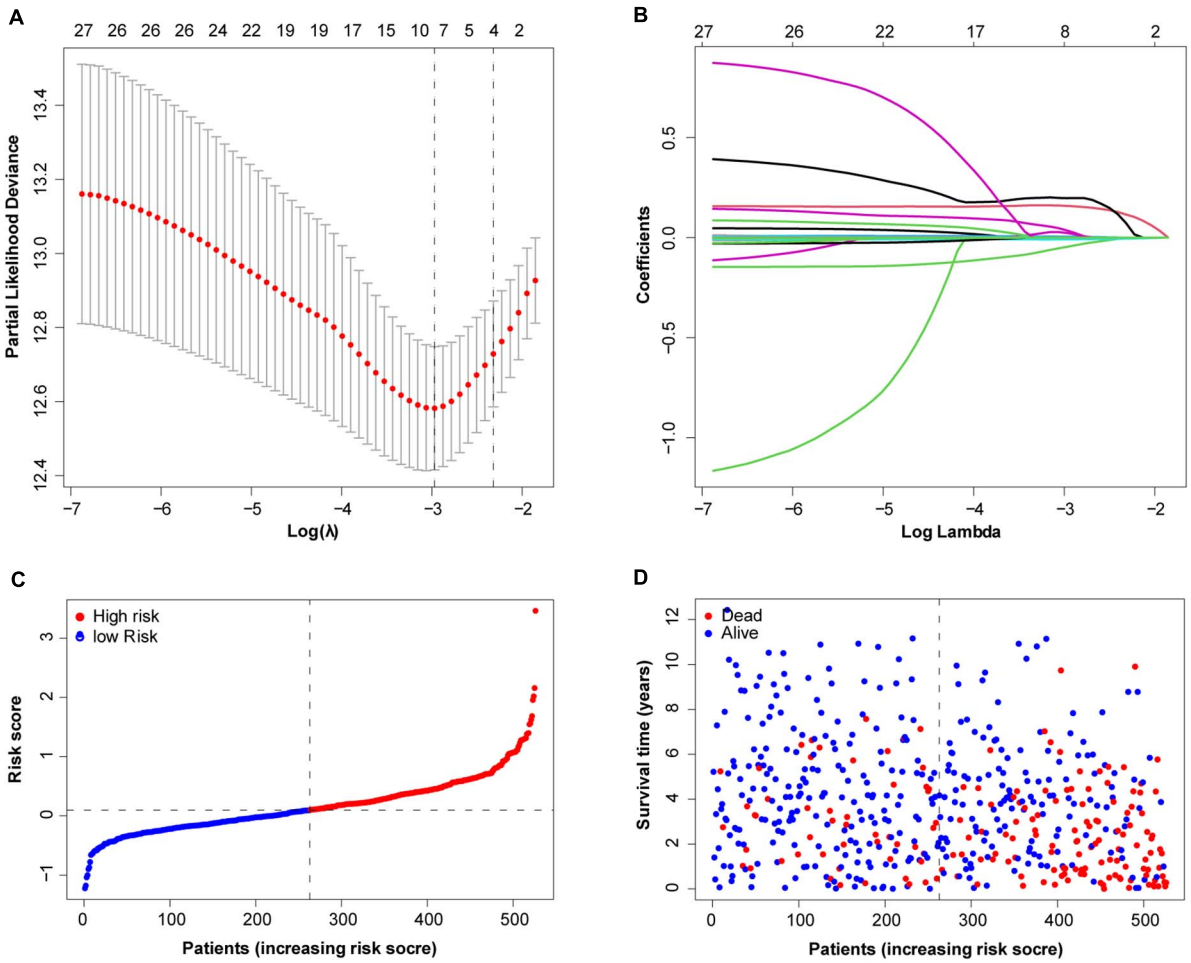


FIGURE 2 | Determination of ferroptosis-related genes (FRGs) as an independent prognostic factor. **(A)** Partial likelihood deviance was plotted against log (lambda). Vertical dotted lines indicate the lambda value with minimum error. The largest lambda value is where the deviation is within one standard error (SE) of the minimum. **(B)** Least absolute shrinkage and selection operator (LASSO) coefficient profiles of FRGs and selected seven FRGs. **(C)** Risk score curve of ferroptosis signature. **(D)** Patient survival status and time distributed by FRG-based risk score.

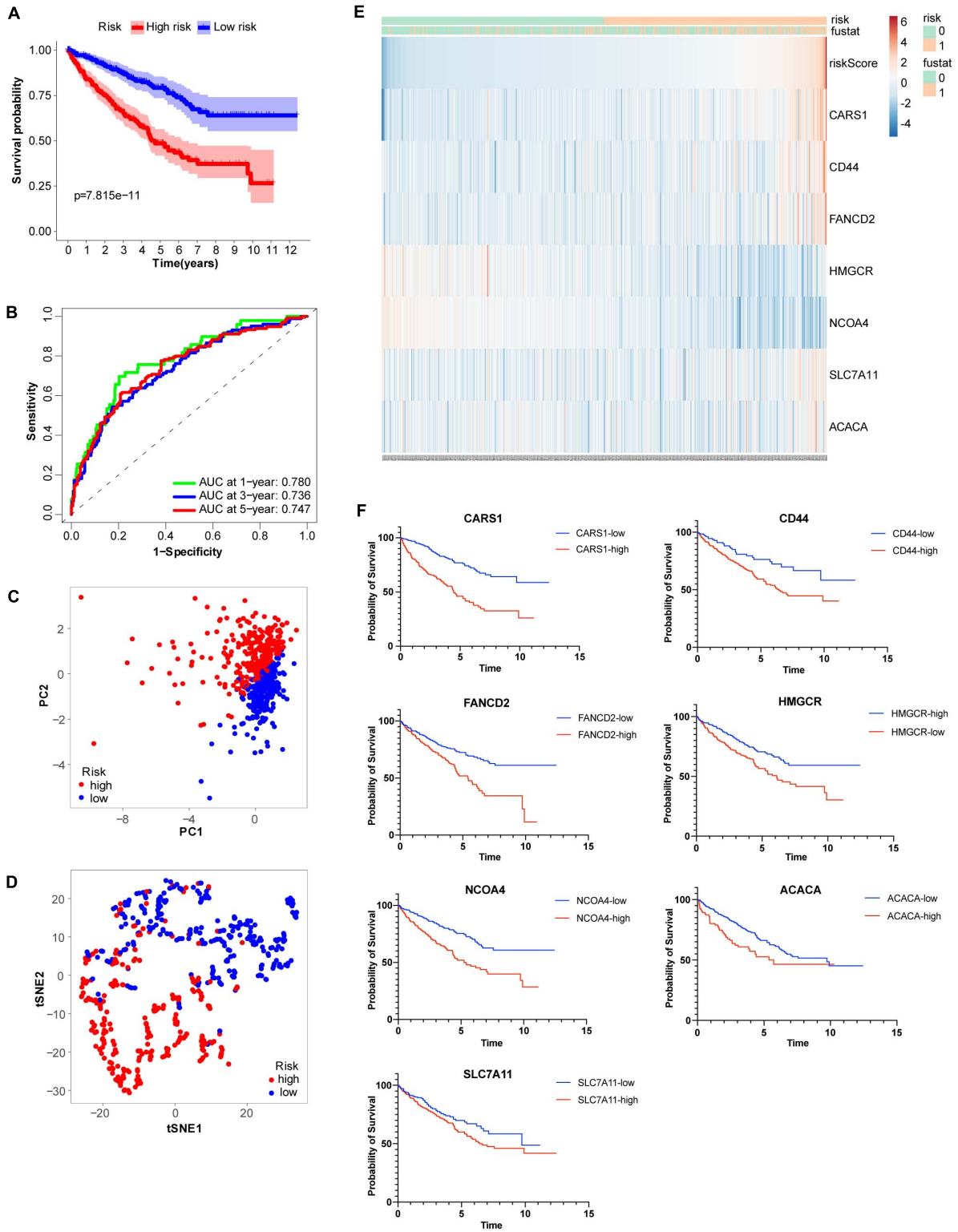


FIGURE 3 | Analysis of ferroptosis-related genes (FRGs) as an independent prognostic factor. **(A)** Kaplan–Meier survival curves show overall *P* values of survival and 95% CI for high− and low−risk patients with clear cell renal cell carcinoma (ccRCC) based on the novel model. **(B)** One−, 3−, and 5-year receiver operating characteristic (ROC) curves and area under the curve (AUC) = 0.780, 0.736, and 0.747, respectively. **(C)** Principal component analysis of FRGs showed two patient clusters. **(D)** t-Distributed stochastic neighbor embedding plots for non-linear dimension divided ccRCC into two groups by the new risk score model. **(E)** Heatmap showing differentially expressed ferroptosis-related genes between high− and low−risk. **(F)** Kaplan-Meier survival curves of the seven genes in ccRCC.

group has better prognosis result than the high-risk group ($P < 0.05$, **Supplementary Materials 1**).

Molecular Function Analysis Indicated the Relationships Between FRGs and Immune Infiltrates

To explore the molecular function and potential signaling pathways related to the seven prognostic genes, we conducted GO and KEGG analyses. GO enrichment analysis (biological process, cellular component, and molecular function) of the seven FRGs revealed a significant relationship between humoral immune response, immunoglobulin complex, and antigen binding (**Figure 4A**). KEGG analysis showed that these genes were significantly associated with the cytokine–cytokine receptor interaction (**Figure 4B**). Furthermore, ssGSEA was conducted to detect potential relationships between immune-related signaling pathways and the seven prognostic genes. As shown in **Figures 4C,D**, among the low− and high-risk groups, significant differences ($P < 0.001$) were observed in some immune-function-related pathways such as cytolytic activity, inflammation-promoting, T-cell costimulation, and type II interferon response. The cell proportions were significantly different for lymphoid dendritic cells, mast cells, neutrophils, follicular helper T cells, and Th2 cells between the high− and low−risk groups ($P < 0.001$).

Construction and Validation of FRG-Based Risk Score Prognostic Model

To better understand the prognostic value of the risk score prognostic model and other clinical characteristics in ccRCC, we performed both univariate and multivariate Cox regression analyses to identify factors affecting OS in TCGA dataset (**Figures 5A,B**). As staging is based on tumor–node–metastasis (TNM) classes, only stage was included in our analysis. Four independent prognostic factors – age, seven FRG-based risk score, grade, and stage – were included in the prediction model. We estimated the nomogram associated with FRGs to predict the OS probability at 1, 3, and 5 years and evaluated its predictive ability (**Figure 5C**). The C-index of the nomogram was 0.705 (95% CI, 0.663−0.747; $P = 6.885 \times 10^{-21}$). The calibration plots for 1−, 3−, 5-year survival showed a favorable predictive ability of the nomogram in predicting the OS of patients with ccRCC. Compared with other clinical prognostic factors, decision curve analysis showed that the nomogram was more beneficial for patients (**Figure 6**).

DISCUSSION

Renal cell carcinoma is a malignant tumor with an occult incidence and is difficult to diagnose in early stages. Although previous studies have investigated numerous molecular biomarkers and multiple gene expression signatures for ccRCC (Wu G. et al., 2020), the clinical utility patterns and immune microenvironment of FRGs have not been comprehensively analyzed. A Cox proportional-hazards model was developed and validated to predict the survival probabilities of patients, resulting in the selection of seven FRGs. These genes may play important roles in the occurrence and development of ccRCC. Based on related studies of FRGs, we explored the mechanism and immune microenvironment of ccRCC from the perspective of ferroptosis. We hypothesized that the combination of FRGs and clinical characteristics could accurately predict the prognosis of ccRCC. Thus, we constructed a risk score model based on these genes and applied it to establish a nomogram.

Ferroptosis disrupts iron and lipid metabolism, leading to the accumulation of reactive oxygen species and lipid peroxidation products. The molecular biological functions of seven FRGs were reported previously. We evaluated the roles of these genes in ccRCC; NCOA4 is a cargo receptor that is highly enriched in autophagosomes. The combination of NCOA4 and ferritin heavy chain 1 promotes ferroptosis by releasing free iron stored in ferritin heavy chain 1. Our risk score analysis showed that high expression of NCOA4 in patients has a negative risk score. Thus, NCOA4 may have some protective effects in patients. Polyunsaturated fatty acids are the main oxidants in the cytoplasm and are synthesized by acetyl-CoA. ACACA/ACC1 catalyzes the rate-limiting step of fatty acid synthesis and triggers cell death (Dixon et al., 2015). However, phosphorylation of ACACA limits its function, thus protecting cells from cell death (Lee et al., 2020). GPX4 mediates the conversion of glutathione to GSSG, which can reduce lipid peroxidation and inhibit ferroptosis. Glutathione is an intracellular antioxidant that can convert to GSSG by GPX4. As the precursor material of glutathione, cysteine can be converted from cystine that is transported into cells by the x_c^- system; this system consists of light-chain subunit SLC7A11 and heavy-chain subunit SLC3A2. Extracellular cystine is imported with intracellular glutamate release, and SLC7A11 is specific for the x_c^- system (Lin et al., 2020). Thus, SLC7A11 is an important transporter that maintains intracellular cystine homeostasis (Bannai, 1986). Transport of cystine inhibits peroxidation and reduces cell death. SLC7A11 can also affect the occurrence and progression of tumors in other ways (Lin et al., 2020). CD44 is a common biomarker of cancer stem cells as well as tumor cell invasion and metastasis (Xu H. et al., 2020). Moreover, CD44 acts on SLC7A1 to stabilize SLC7A11 and inhibit iron death in combination with the x_c^- system (Ishimoto et al., 2011). The HMGCR enzyme catalyzes the rate-limiting step of mevalonate-derived terpene biosynthesis including isopentenyl pyrophosphate to promote the maturation of GPX4 (Warner et al., 2000; Viswanathan et al., 2017). Previous studies showed that an increase in GPX4 also inhibits cell death and promotes the tolerance of tumor cells to lipid peroxidation, contrasting our results. According to our risk score, high HMGCR expression, which may cause hypercholesterolemia, is a protective factor in renal ccRCC. This outcome was also observed previously (Wu G. et al., 2020). Statins can inhibit the function of HMGCR and are widely used to treat hypercholesterolemia. Statins are administered as a protective factor to patients with ccRCC (Neumann et al., 2019; Okubo et al., 2020). As no medication information was given in the clinical data, the impact of statin use on our results is unclear. CARS, the cysteinyl-tRNA synthetase, suppresses ferroptosis induced by erastin,

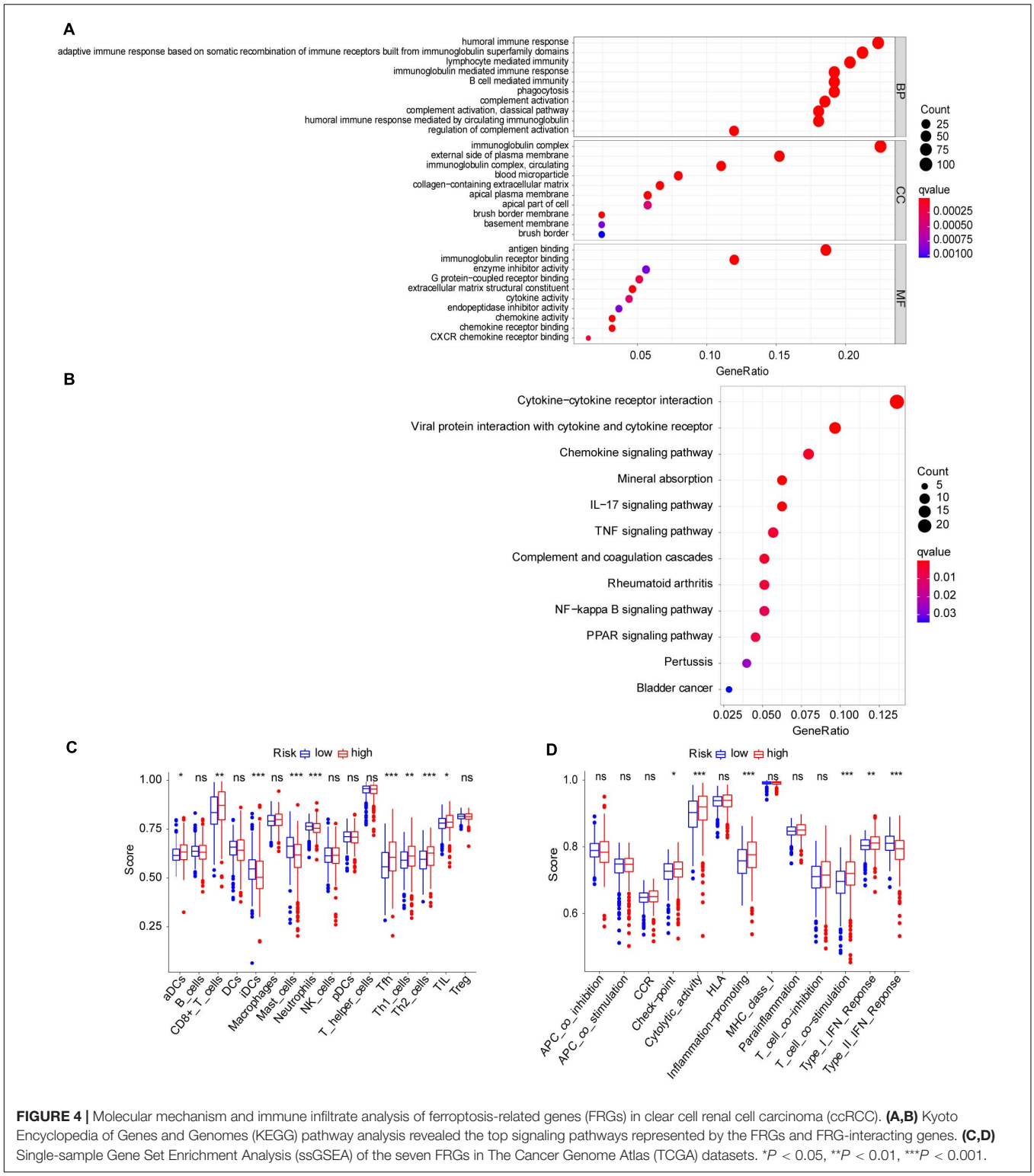


FIGURE 4 | Molecular mechanism and immune infiltrate analysis of ferroptosis-related genes (FRGs) in clear cell renal cell carcinoma (ccRCC). **(A,B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed the top signaling pathways represented by the FRGs and FRG-interacting genes. **(C,D)** Single-sample Gene Set Enrichment Analysis (ssGSEA) of the seven FRGs in The Cancer Genome Atlas (TCGA) datasets. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

which inhibits the cystine–glutamate antiporter known as the x_c^- system (Hayano et al., 2016). FANCD2 is a nuclear protein that may regulate iron metabolism by ferritin heavy chain 1, TF, and lipid peroxidation by GPX4 to prevent ferroptosis (Song et al., 2016).

The cross-talk between tumor cells and non-tumor cells like immune cells, fibroblasts, and myeloid lineage cells consists the tumor microenvironment, which plays an important role in tumor occurrence and progression. The immune cells including T cells, dendritic cells, and macrophages are both

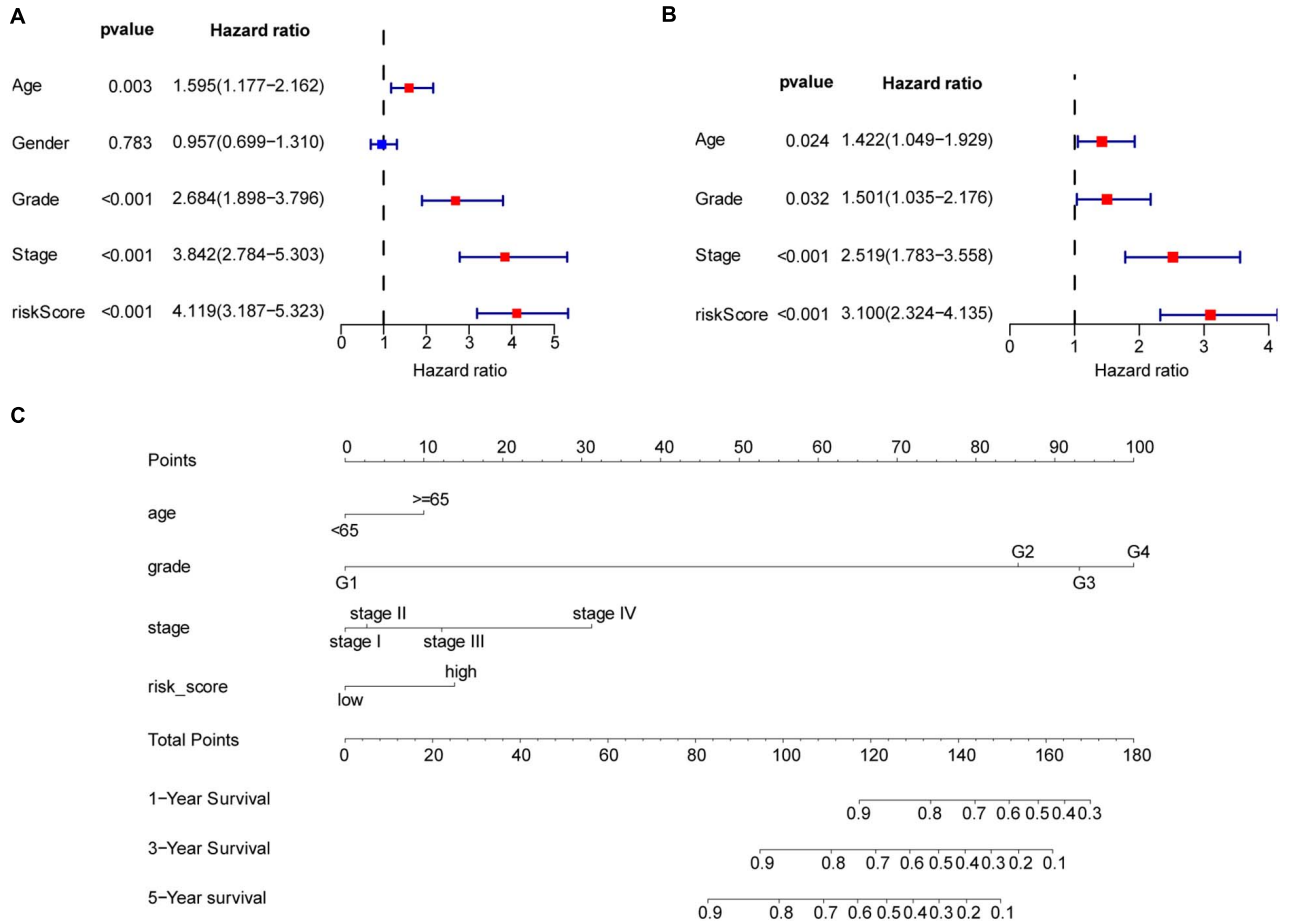


FIGURE 5 | Nomogram to predict the survival probability of patients with clear cell renal cell carcinoma (ccRCC). **(A)** Univariate Cox regression of clinical characteristic and risk score model. **(B)** Based on the results of univariate Cox regression, hazard ratio (HR), and 95% CI of multivariate Cox regression analysis, age, grade, stage, and risk score showed significant differences between the high− and low-risk groups. **(C)** Prognostic nomogram for predicting the survival of patients with ccRCC. Every predict factor corresponds to a score. Clinically, the corresponding score can be obtained according to the patient's condition, and the total score can be obtained by adding each score, which corresponds to the corresponding survival probability.

prohibited and promoted tumorigenic depending on the complex cross-talk (Schulz et al., 2019; Yang et al., 2021). With tumor growth, fibroblasts and macrophages began to have affinity to tumor cells, and immune-suppressor cells, including myeloid-derived suppressor cells, and Treg cells were mobilized. As the consequence, many immune survey pathways were blocked. In ccRCC, there are abundant infiltration of immune cells, such as CD4+ T cells, CD8+ T cells, and natural killer cells (Komohara et al., 2011). However, contrary to what people expected, the number of CD4+ T cells and CD8+ T cells was negatively correlated with the survival time of patients (Nakano et al., 2001). With the further study of the interaction between tumor cells and immune cells, the prospect of effective immunotherapies for the treatment of patients with cancer is now becoming a clinical reality. The most widely studied is the expression of key receptors on the surface of T cells that prevent them from becoming activated, cytotoxic T lymphocyte antigen 4 (CTLA4), programmed cell death-1 (PD-1), and programmed death-ligand 1 (PD-L1). Based on the above two immune checkpoints, a series of monoclonal antibody drugs targeting CTLA4, PD-1, and PD-L1 have been developed, which have all been shown to prevent the interaction of these molecules with their respective inhibitory target proteins, thereby restoring antitumor immune responses (Xu W. et al., 2020).

Our study also found the association between FRGs and immunity, so we further conducted ssGSEA. ssGSEA is used to calculate enrichment scores for each paired sample and gene set and is an extension of GSEA for systematically analyzing the association of FRDs and the immune microenvironment (Barbie et al., 2009). Our results partially explain the possible reaction between the immune system and ferroptosis in ccRCC. As shown in **Figure 4C**, the ssGSEA score of CD8+ T cells was higher in the high-risk group, and the ferroptosis signaling pathway represented by FRGs was more active in CD8+ cells. This may imply that CD8+ T cells are more likely to die of programmed cell death. Additionally, studies suggested that T-cell activation state are strong prognostic determinants of

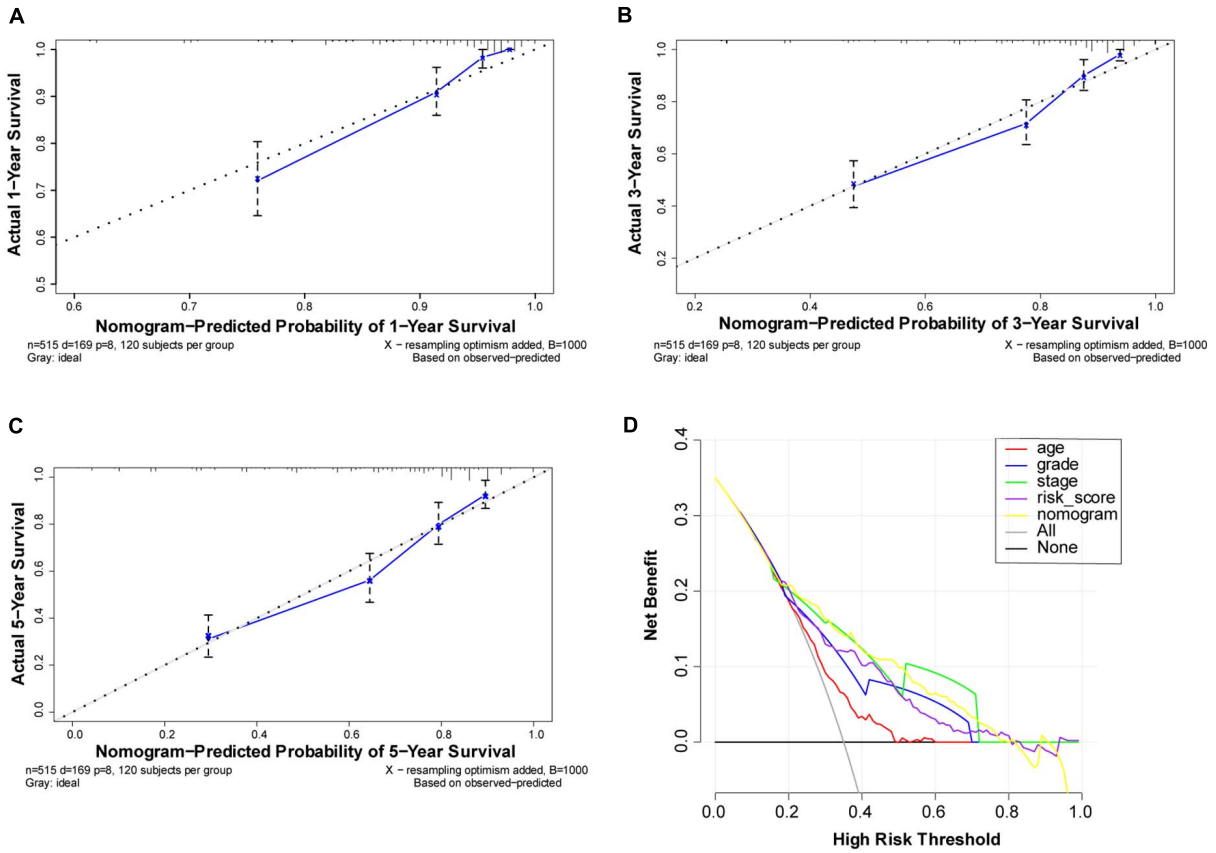


FIGURE 6 | Performance of ferroptosis-related gene (FRG)-based risk score prognostic model. **(A–C)** Calibration curves of the nomogram for predicting survival at 1, 3, and 5 years. If the actual curve is closer to the ideal curve, the nomogram prediction accuracy is higher **(D)** Decision curve analysis of this nomogram. Including the risk score model, the nomogram shows that advanced age, grade, risk score, and partially were better than stage for predicting survival.

ccRCC (Nakano et al., 2001; Adotevi et al., 2010). Interferon-γ is secreted by CD8+ T cells, leading to the downregulation of the expression of SLC7A11 of tumor cells, decline in levels of intracellular cystine, and prompting ferroptosis (Wang et al., 2019). By exporting glutamate, the x_c^- system also affects the tumor microenvironment (Lin et al., 2020). CD44 not only affects the immune system by affecting SLC7A11 but also regulates the migration and activation of T cells in a variety of ways (Kano et al., 2014). NCOA4 can also affect the recognition of tumor cells by interfering with interferon-γ (IFN-γ) receptor signaling (Sottile et al., 2019). CARS could be secreted from cancer cells to activate immune responses via specific interactions with TLR2/6 of dendritic cells (Cho et al., 2020). Inhibition of ACACA function was reported to enhance the formation of CD4+ T memory cells by disturbing the fatty acid metabolism pathway (Endo et al., 2019). Iron death is a double-edged sword. If the iron death pathway is activated in T cells, the effect of T cells on tumor is decreased. Recent studies have shown that T cells can also induce iron death in tumor cells, thus inhibiting tumor cells (Li and Li, 2020). The function of immune cells in tumor may be related to their activation status and regulatory factors and present affinity and antitumor characteristics at the same time (Zamarron and Chen, 2011; Gajewski et al., 2013).

The ferroptosis provides a new idea for the study of tumor immune infiltration. Future research may reveal the therapeutic direction of tumor immunotherapy by clarifying the mechanism of cytokines. Based on the function of GO and KEGG, our cellular function analysis is to explore the function of these FRGs. The individual function of these genes needs to be verified and identified in further experiments.

The nomogram is based on regression analysis, which integrates multiple clinical prognostic predictors. According to the contribution of each predictor to the overall survival rate, the points are scored, and then, the total points are obtained by adding the scores. According to the total score, the 1−, 3−, or 5-year overall survival rate of patients is calculated. The nomogram makes the prognosis model more readable and convenient for clinicians to evaluate the survival prognosis of patients.

To establish the nomogram, we detected seven FRGs most closely related to the clinical prognosis of ccRCC. Patients were divided into high− and low−risk groups according to the risk score. Clinical stage was determined based on TNM classification. Therefore, to avoid duplication of cancer classification, only stage classification was considered in this study. Age, gender, histological grade, and stage were integrated into the Cox proportional hazards models to achieve better predictive

performance. FRG prognostic nomograms based on the results of the Cox regression model were developed and validated to quantitatively estimate the survival probabilities of patients with ccRCC. The C-index, decision curve analysis, and calibration plots demonstrated favorable consistency between the actual and predicted survival. Thus, our gene signature and nomogram may provide an accurate and reliable prediction approach for the prognosis of patients with ccRCC and help clinicians optimize and personalize treatment strategies.

There were some limitations to our research, which will be further explored in our future work. We attempted to verify the above model by proteomics analysis but did not find all seven FRGs in the Clinical Proteomic Tumor Analysis Consortium. In some studies, ACACA was shown to promote programmed cell death, whereas other studies showed that phosphorylation of ACACA inhibits iron-related cell death (Lee et al., 2020). We hypothesized that the phosphorylation of ACACA in ccRCC could cause loss of function of this enzyme. We were unable to verify this hypothesis, as our study was performed at the RNA level. Many messenger RNA (mRNA) isomers play different roles in tumor, which could not be examined using databases. Thus, mRNA isomers were not considered in our research. Among the numerous forms of CD44 produced by alternative mRNA splicing, CD44v functions closely integrated to induce the x_c^- system and responds to SLC7A1 (Ishimoto et al., 2011). HMGCR has been predicted to inhibit ferroptosis, which contrasts the results observed using our risk score. Whether HMGCR is affected by statin administration to patients requires further analysis. Additionally, our research and hypothesis are based on mRNA level; the functions of these seven FRGs in ccRCC should be examined in further *in vivo* and *in vitro* studies. We searched several databases but did not find other ccRCC sequencing data with similar patient baseline and sequencing platform. The testing dataset in this study should be evaluated in further research.

In conclusion, this study explored the role of FRGs in ccRCC and the relationship between FRGs and the immune environment in ccRCC. We developed an FRG-based prognostic nomogram to improve estimation of the survival rate of patients with ccRCC. This model may be used by medical professionals to develop further treatment options and perform in-depth studies of the molecular biology of ccRCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

ETHICS STATEMENT

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

LM, SZ, and G-JZ designed and conducted the study. G-JZ, ZW, and FY analyzed the data. G-JZ wrote the manuscript. LM, KH, and LG reviewed the draft. All authors read and approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (no. 81972381), National Key Research and Development Program of China (grant nos. 2017YFC1002001 and SQ2018YFC100243), Beijing Natural Science Foundation of China (7182177), and Beijing Municipal Science and Technology Commission (grant no. Z151100003915105).

ACKNOWLEDGMENTS

The authors thank The Cancer Genome Atlas (TCGA) project for its valuable public dataset. The authors also thank Yujun Liu (Department of Epidemiology and Hygienic Statistics) for helping us with the drawings.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.650416/full#supplementary-material

REFERENCES

Abu Aboud, O., Habib, S. L., Trott, J., Stewart, B., Liang, S., Chaudhari, A. J., et al. (2017). Glutamine addiction in kidney cancer suppresses oxidative stress and can be exploited for real-time imaging. *Cancer Res.* 77, 6746–6758. doi: 10.1158/0008-5472.Can-17-0930

Adotevi, O., Pere, H., Ravel, P., Haicheur, N., Badoual, C., Merillon, N., et al. (2010). A decrease of regulatory T cells correlates with overall survival after sunitinib-based antiangiogenic therapy in metastatic renal cancer patients. *J. Immunother.* 33, 991–998. doi: 10.1097/CJI.0b013e3181f4c208

Bannai, S. (1986). Exchange of cystine and glutamate across plasma membrane of human fibroblasts. *J. Biol. Chem.* 261, 2256–2263.

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112. doi: 10.1038/nature08460

Bersuker, K., Hendricks, J. M., Li, Z., Magtanong, L., Ford, B., Tang, P. H., et al. (2019). The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit ferroptosis. *Nature* 575, 688–692. doi: 10.1038/s41586-019-1705-2

Bogdan, A. R., Miyazawa, M., Hashimoto, K., and Tsuji, Y. (2016). Regulators of iron homeostasis: new players in metabolism, cell death, and disease. *Trends Biochem. Sci.* 41, 274–286. doi: 10.1016/j.tibs.2015.11.012

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Capitanio, U., and Montorsi, F. (2016). Renal cancer. *Lancet* 387, 894–906. doi: 10.1016/S0140-6736(15)00046-X

Capitanio, U., Cloutier, V., Zini, L., Isbarn, H., Jeldres, C., Shariat, S. F., et al. (2009). A critical assessment of the prognostic value of clear cell, papillary and chromophobe histological subtypes in renal cell carcinoma: a population-based study. *BJU Int.* 103, 1496–1500. doi: 10.1111/j.1464-410X.2008.08259.x

Chen, X., Li, J., Kang, R., Klionsky, D. J., and Tang, D. (2020). Ferroptosis: machinery and regulation. *Autophagy* 26, 1–28. doi: 10.1080/15548627.2020.1810918

Cheng, G., Liu, D., Liang, H., Yang, H., Chen, K., and Zhang, X. (2019). A cluster of long non-coding RNAs exhibit diagnostic and prognostic values in renal cell carcinoma. *Aging (Albany NY)* 11, 9597–9615. doi: 10.18632/aging.102407

Cho, S., Kim, S. B., Lee, Y., Song, E. C., Kim, U., Kim, H. Y., et al. (2020). Endogenous TLR2 ligand embedded in the catalytic region of human cysteinyl-tRNA synthetase 1. *J. Immunother. Cancer* 8, e000277. doi: 10.1136/jitc-2019-000277

Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., et al. (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* 149, 1060–1072. doi: 10.1016/j.cell.2012.03.042

Dixon, S. J., Winter, G. E., Musavi, L. S., Lee, E. D., Snijder, B., Rebsamen, M., et al. (2015). Human haploid cell genetics reveals roles for lipid metabolism genes in nonapoptotic cell death. *ACS Chem. Biol.* 10, 1604–1609. doi: 10.1021/acschembio.5b00245

Doll, S., Freitas, F. P., Shah, R., Aldrovandi, M., da Silva, M. C., Ingold, I., et al. (2019). FSP1 is a glutathione-independent ferroptosis suppressor. *Nature* 575, 693–698. doi: 10.1038/s41586-019-1707-0

Du, W., Zhang, L., Brett-Morris, A., Aguila, B., Kerner, J., Hoppel, C. L., et al. (2017). HIF drives lipid deposition and cancer in ccRCC via repression of fatty acid metabolism. *Nat. Commun.* 8:1769. doi: 10.1038/s41467-017-01965-8

Elgendy, S. M., Alyammahi, S. K., Alhamad, D. W., Abdin, S. M., and Omar, H. A. (2020). Ferroptosis: an emerging approach for targeting cancer stem cells and drug resistance. *Crit. Rev. Oncol. Hematol.* 155:103095. doi: 10.1016/j.critrevonc.2020.103095

Endo, Y., Onodera, A., Obata-Ninomiya, K., Koyama-Nasu, R., Asou, H. K., Ito, T., et al. (2019). ACC1 determines memory potential of individual CD4(+) T cells by regulating de novo fatty acid biosynthesis. *Nat. Metab.* 1, 261–275. doi: 10.1038/s42255-018-0025-4

Frew, I. J., and Moch, H. (2015). A clearer view of the molecular complexity of clear cell renal cell carcinoma. *Annu. Rev. Pathol.* 10, 263–289. doi: 10.1146/annurev-pathol-012414-040306

Gajewski, T. F., Schreiber, H., and Fu, Y. X. (2013). Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 14, 1014–1022. doi: 10.1038/ni.2703

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., et al. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* 25, 486–541. doi: 10.1038/s41418-017-0012-4

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 14:7. doi: 10.1186/1471-2105-14-7

Hayano, M., Yang, W. S., Corn, C. K., Pagano, N. C., and Stockwell, B. R. (2016). Loss of cysteinyl-tRNA synthetase (CARS) induces the transsulfuration pathway and inhibits ferroptosis induced by cystine deprivation. *Cell Death Differ.* 23, 270–278. doi: 10.1038/cdd.2015.93

Hirschhorn, T., and Stockwell, B. R. (2019). The development of the concept of ferroptosis. *Free Radic. Biol. Med.* 133, 130–143. doi: 10.1016/j.freeradbiomed.2018.09.043

Hou, W., Xie, Y., Song, X., Sun, X., Lotze, M. T., Zeh, H. J. III, et al. (2016). Autophagy promotes ferroptosis by degradation of ferritin. *Autophagy* 12, 1425–1428. doi: 10.1080/15548627.2016.1187366

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26, 1364–1370. doi: 10.1200/JCO.2007.12.9791

Ishimoto, T., Nagano, O., Yae, T., Tamada, M., Motohara, T., Oshima, H., et al. (2011). CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc(-) and thereby promotes tumor growth. *Cancer Cell* 19, 387–400. doi: 10.1016/j.ccr.2011.01.038

Kano, Y., Ishii, H., Konno, M., Yamasaki, M., Miyata, H., Nishikawa, S., et al. (2014). Cells of origin of squamous epithelium, dysplasia and cancer in the head and neck region after bone marrow transplantation. *Int. J. Oncol.* 44, 443–450. doi: 10.3892/ijo.2013.2206

Keegan, K. A., Schupp, C. W., Chamie, K., Hellenthal, N. J., Evans, C. P., and Koppie, T. M. (2012). Histopathology of surgically treated renal cell carcinoma: survival differences by subtype and stage. *J. Urol.* 188, 391–397. doi: 10.1016/j.juro.2012.04.006

Kidd, A. C., McGettrick, M., Tsim, S., Halligan, D. L., Bylesjo, M., and Blyth, K. G. (2018). Survival prediction in mesothelioma using a scalable Lasso regression model: instructions for use and initial performance using clinical predictors. *BMJ Open Respir. Res.* 5:e000240. doi: 10.1136/bmjresp-2017-000240

Komohara, Y., Hasita, H., Ohnishi, K., Fujiwara, Y., Suzu, S., Eto, M., et al. (2011). Macrophage infiltration and its prognostic relevance in clear cell renal cell carcinoma. *Cancer Sci.* 102, 1424–1431. doi: 10.1111/j.1349-7006.2011.01945.x

Lee, H., Zandkarimi, F., Zhang, Y., Meena, J. K., Kim, J., Zhuang, L., et al. (2020). Energy-stress-mediated AMPK activation inhibits ferroptosis. *Nat. Cell Biol.* 22, 225–234. doi: 10.1038/s41556-020-0461-8

Li, D. S., and Li, Y. S. (2020). The interaction between ferroptosis and lipid metabolism in cancer. *Signal Transduct. Target. Ther.* 5:108. doi: 10.1038/s41392-020-00216-5

Lin, W., Wang, C., Liu, G., Bi, C., Wang, X., Zhou, Q., et al. (2020). SLC7A11/xCT in cancer: biological functions and therapeutic implications. *Am. J. Cancer Res.* 10, 3106–3126.

Liu, J., Kuang, F., Kroemer, G., Klionsky, D. J., Kang, R., and Tang, D. (2020). Autophagy-dependent ferroptosis: machinery and regulation. *Cell Chem. Biol.* 27, 420–435. doi: 10.1016/j.chembiol.2020.02.005

Luo, Y., Shen, D., Chen, L., Wang, G., Liu, X., Qian, K., et al. (2019). Identification of 9 key genes and small molecule drugs in clear cell renal cell carcinoma. *Aging (Albany NY)* 11, 6029–6052. doi: 10.18632/aging.102161

Mitchell, T. J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J. H. R., O'Brien, T., et al. (2018). Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell* 173, 611–623.e17. doi: 10.1016/j.cell.2018.02.020

Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E., and Ulbright, T. M. (2016). The 2016 WHO classification of tumours of the urinary system and male genital organs-part a: renal, penile, and testicular tumours. *Eur. Urol.* 70, 93–105. doi: 10.1016/j.eururo.2016.02.029

Mou, Y., Wu, J., Zhang, Y., Abdihamid, O., Duan, C., and Li, B. (2021). Low expression of ferritinophagy-related NCOA4 gene in relation to unfavorable outcome and defective immune cells infiltration in clear cell renal carcinoma. *BMC Cancer* 21:18. doi: 10.1186/s12885-020-07726-z

Nakano, O., Sato, M., Naito, Y., Suzuki, K., Orikasa, S., Aizawa, M., et al. (2001). Proliferative activity of intratumoral CD8(+) T-lymphocytes as a prognostic factor in human renal cell carcinoma: clinicopathologic demonstration of antitumor immunity. *Cancer Res.* 61, 5132–5136.

Neumann, E., Klaiber, P., Freitag, K., Schwab, M., Schaeffeler, E., Hennenlotter, J., et al. (2019). Assessment of concomitant non-oncologic medication in patients with surgically treated renal cell carcinoma: impact on prognosis, cell-cycle progression and proliferation. *J. Cancer Res. Clin. Oncol.* 145, 1835–1843. doi: 10.1007/s00432-019-02914-2

Okubo, K., Isono, M., Miyai, K., Asano, T., and Sato, A. (2020). Fluvastatin potentiates anticancer activity of vorinostat in renal cancer cells. *Cancer Sci.* 111, 112–126. doi: 10.1111/cas.14225

Schulz, M., Salamero-Boix, A., Niesel, K., Alekseeva, T., and Sevenich, L. (2019). Microenvironmental regulation of tumor progression and therapeutic response in brain metastasis. *Front. Immunol.* 10:1713. doi: 10.3389/fimmu.2019.01713

Song, X., Xie, Y., Kang, R., Hou, W., Sun, X., Epperly, M. W., et al. (2016). FANCD2 protects against bone marrow injury from ferroptosis. *Biochem. Biophys. Res. Commun.* 480, 443–449. doi: 10.1016/j.bbrc.2016.10.068

Sottile, R., Federico, G., Garofalo, C., Tallerico, R., Faniello, M. C., Quaresima, B., et al. (2019). Iron and ferritin modulate MHC class I expression and NK cell recognition. *Front. Immunol.* 10:224. doi: 10.3389/fimmu.2019.00224

Stockwell, B. R., Friedmann Angeli, J. P., Bayir, H., Bush, A. I., Conrad, M., Dixon, S. J., et al. (2017). Ferroptosis: a regulated cell death nexus linking metabolism, redox biology, and disease. *Cell* 171, 273–285. doi: 10.1016/j.cell.2017.09.021

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., et al. (2018). Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* 173, 595–610.e11. doi: 10.1016/j.cell.2018.03.043

Viswanathan, V. S., Ryan, M. J., Dhruv, H. D., Gill, S., Eichhoff, O. M., Seashore-Ludlow, B., et al. (2017). Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* 547, 453–457. doi: 10.1038/nature23007

Wang, J., Yin, X., He, W., Xue, W., Zhang, J., and Huang, Y. (2021). SUV39H1 deficiency suppresses clear cell renal cell carcinoma growth by inducing ferroptosis. *Acta Pharm. Sin B* 11, 406–419. doi: 10.1016/j.apsb.2020.09.015

Wang, W., Green, M., Choi, J. E., Gijón, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8+ T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* 569, 270–274. doi: 10.1038/s41586-019-1170-y

Warner, G. J., Berry, M. J., Moustafa, M. E., Carlson, B. A., Hatfield, D. L., and Faust, J. R. (2000). Inhibition of selenoprotein synthesis by selenocysteine tRNA[Ser]Sec lacking isopentenyladenosine. *J. Biol. Chem.* 275, 28110–28119. doi: 10.1074/jbc.M001280200

Wu, G., Wang, Q., Xu, Y., Li, Q., and Cheng, L. (2020). A new survival model based on ferroptosis-related genes for prognostic prediction in clear cell renal cell carcinoma. *Aging (Albany, NY)* 12, 14933–14948. doi: 10.18632/aging.103553

Wu, Y. N., Yu, C. C., Luo, M., Cen, C., Qiu, J. L., Zhang, S. Z., et al. (2020). Ferroptosis in cancer treatment: another way to Rome. *Front. Oncol.* 10:571127. doi: 10.3389/fonc.2020.571127

Xu, H., Niu, M., Yuan, X., Wu, K., and Liu, A. (2020). CD44 as a tumor biomarker and therapeutic target. *Exp. Hematol. Oncol.* 9:36. doi: 10.1186/s40164-020-00192-0

Xu, W., Atkins, M. B., and McDermott, D. F. (2020). Checkpoint inhibitor immunotherapy in kidney cancer. *Nat. Rev. Urol.* 17, 137–150. doi: 10.1038/s41585-020-0282-3

Yang, C. A., Huang, H. Y., Yen, J. C., and Chang, J. G. (2018). Prognostic value of RNASEH2A-, CDK1-, and CD151-related pathway gene profiling for kidney cancers. *Int. J. Mol. Sci.* 19;1586. doi: 10.3390/ijms19061586

Yang, F., Liu, C., Zhao, G., Ge, L., Song, Y., Chen, Z., et al. (2020). Long non-coding RNA LINC01234 regulates proliferation, migration and invasion via HIF-2alpha pathways in clear cell renal cell carcinoma cells. *PeerJ* 8:e10149. doi: 10.7717/peerj.10149

Yang, M., Li, J., Gu, P., and Fan, X. (2021). The application of nanoparticles in cancer immunotherapy: targeting tumor microenvironment. *Bioact. Mater.* 6, 1973–1987. doi: 10.1016/j.bioactmat.2020.12.010

Zamarron, B. F., and Chen, W. J. (2011). Dual roles of immune cells and their factors in cancer development and progression. *Int. J. Biol.l Sci.* 7, 651–658. doi: 10.7150/ijbs.7.651

Zhang, C., Li, Z., Qi, F., Hu, X., and Luo, J. (2019). Exploration of the relationships between tumor mutation burden with immune infiltrates in clear cell renal cell carcinoma. *Ann. Transl. Med.* 7:648. doi: 10.21037/atm.2019.10.84

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 in Oncology

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fonc.2021.691798

An Improved Detection of Circulating Tumor DNA in Extracellular Vesicles-Depleted Plasma

Li Sun[1,2,3,4], Meijun Du[3†], Manish Kohli[5], Chiang-Ching Huang[6], Xiaoxiang Chen[2], Mu Xu[2], Hongbing Shen[4], Shukui Wang[2] and Liang Wang[3,7*]*

[1] Laboratory Medicine Center, the Second Affiliated Hospital, Nanjing Medical University, Nanjing, China, [2] Department of General Clinical Research Center, Nanjing First Hospital, Nanjing Medical University, Nanjing, China, [3] Department of Pathology, Medical College of Wisconsin, Milwaukee, WI, United States, [4] Department of Epidemiology and Biostatistics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, School of Public Health, Nanjing Medical University, Nanjing, China, [5] Division of Oncology, University of Utah Huntsman Cancer Institute, Salt Lake City, UT, United States, [6] Zilber School of Public Health, University of Wisconsin, Milwaukee, WI, United States, [7] Department of Tumor Biology, H. Lee Moffitt Cancer Center, Tampa, FL, United States

OPEN ACCESS

Edited by:
Qinghua Xu, Canhelp Genomics, China

Reviewed by:
Massimo Broggini, Istituto di Ricerche Farmacologiche Mario Negri (IRCCS), Italy
Beifang Niu, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Liang Wang
liang.wang@moffitt.org
Shukui Wang
sk_wang@njmu.edu.cn

†Present address:
Meijun Du, Department of the Genomic Sciences and Precision Medicine Center (GSPMC), Medical College of Wisconsin, Milwaukee, WI, United States

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

Received: *07 April 2021*
Accepted: *17 May 2021*
Published: *11 June 2021*

Citation:
Sun L, Du M, Kohli M, Huang C-C, Chen X, Xu M, Shen H, Wang S and Wang L (2021) An Improved Detection of Circulating Tumor DNA in Extracellular Vesicles-Depleted Plasma. Front. Oncol. 11:691798. doi: 10.3389/fonc.2021.691798

Circulating tumor DNA (ctDNA) in plasma has been used as a biomarker for cancer detection and outcome prediction. In this study, we collected the five precipitates (fractions 1–5) and leftover supernatant plasma component (fraction 6) by a sequential centrifugation in plasma samples from nine small cell lung cancer (SCLC) patients. The fractions 3, 5 and 6 were large vesicles, exosomes and extracellular vesicles (EVs)-depleted plasma, respectively. Fragment size analysis using DNAs from these fractions showed dramatical differences from a peak of 7–10 kb in fraction 1 to 140–160 bp in fraction 6. To determine ctDNA content, we performed whole genome sequencing and applied copy number-based algorithm to calculate ctDNA percentage. This analysis showed the highest ctDNA content in EV-depleted plasma (average = 27.22%), followed by exosomes (average = 22.09%) and large vesicles (average = 19.70%). Comparatively, whole plasma, which has been used in most ctDNA studies, showed an average of 23.84% ctDNA content in the same group of patients. To further demonstrate higher ctDNA content in fraction 6, we performed mutational analysis in the plasma samples from 22 non-small cell lung cancer (NSCLC) patients with known EGFR mutations. This analysis confirmed higher mutation detection rates in fraction 6 (14/22) than whole plasma (10/22). This study provides a new insight into potential application of using fractionated plasma for an improved ctDNA detection.

Keywords: ctDNA, liquid biopsy, plasma, exosome, copy number variation

INTRODUCTION

Cancer is a serious public burden with an estimation of 1,898,160 new cancer cases and 608,570 cancer deaths in the United States in 2021 (1). To reduce cancer-related morbidity and mortality, more effective approaches in diagnosis and treatment are urgently needed. It is well known that genomic abnormalities are not only hallmarks of cancers but also in evolution during cancer

progression (2). Due to intratumor heterogeneity, however, genomic sequencing from a single tumor biopsy may not fully capture the genomic profile of tumors (3). Moreover, tissue biopsy is limited on tissue availability and sampling frequency. It may increase patients' risk of complication because of the invasive procedure. To address these issues, analysis of circulating cell-free DNA (cfDNA) in blood has been used as a non-invasive method for molecular characterization of tumor genome variations. This blood-based approach has been referred to as liquid biopsy and has demonstrated great potential in cancer diagnosis and outcome prediction (4–6).

Due to lack of adequate oxygen and nutrition, rapidly growing tumor cells often become stressed, and experience apoptosis and necrosis. DNA fragments released from these dead cells eventually end up in circulating blood (7). In patients with cancer, a fraction of cfDNA is tumor-derived and is termed circulating tumor DNA (ctDNA). Analysis of ctDNA has an advantage of identifying genomic alterations that are specific to tumor (8, 9). Interestingly, ctDNA has also been reported in isolated extracellular vesicles (EVs) (10–12). Analysis of vesicles-associated nucleic acids for BRAF, KRAS, and EGFR mutations has shown higher sensitivity compared to plasma ctDNA in non-small cell lung cancer (NSCLC) patients (13). Microvesicles isolated from plasma of NSCLC patients can be used for EGFR genotyping for the detection of drug-resistance mutations, demonstrating improved concordance with tumor tissue compared to a conventional ctDNA (14). Additionally, exosomes from patients with metastatic pancreatic cancer showed a higher mutant *KRAS* allele frequency than exosomes from patients with local disease (15). These studies suggest that EVs may enrich ctDNA and may be used as a preferred source of material for cancer biomarker discovery. However, a recent study showed that the extracellular DNA may not be associated with exosomes, but could instead be co-purified with the small EV fraction during standard isolation protocols (16). Nevertheless, these studies suggest that EVs and/or their co-precipitates enrich ctDNA and may be used to increase sensitivity of cancer biomarker detection.

To systematically determine DNA size distribution and ctDNA content in different fractions of plasma, in this study, we collected plasma from nine small cell lung cancer (SCLC) patients with known high ctDNA content (17). We performed five consecutive centrifugations, collected each of precipitates and analyzed DNA size distribution in each collection. We also performed low-pass whole genome sequencing and estimated ctDNA content using a novel copy number-based algorithm in each of these fractions. In a separate set of plasma samples consisting of 22 non-small cell lung cancer (NSCLC) patients with known EGFR mutations, we compared the mutation detection rate in the fractionated plasma and whole plasma.

MATERIALS AND METHODS

Patients and Plasma Collection

We selected nine SCLC patients whose plasma demonstrated relatively high ctDNA content based on our previous study (17), and 22 NSCLC patients with known EGFR mutations in tumor tissues. We collected the plasma samples from the Medical College of Wisconsin Tissue Bank and the Second Affiliated Hospital of Nanjing Medical University. Original plasma samples (platelet-rich) were prepared by one time 3,000 rpm for 10 min as previously described (17–19). All samples were uniformly processed and stored at −80°C prior to this study. Cancer diagnosis was confirmed in all cases by routine histopathologic examination. All participants provided written informed consent. This study was approved by the Medical College of Wisconsin Institutional Review Broad and the Research and Ethical Committee of the Second Affiliated Hospital of Nanjing Medical University.

Characterization of Fractions 3 (Large EVs) and 5 (Exosomes)

A total of 10 ml pooled plasma from 20 healthy individuals (0.5 ml of each) was used for fractions 3 and 5 preparation (**Figure 1** for detail). Transmission electron microscopy (TEM), Nanosight and flow cytometry were used to characterize the two fractions. The fractionated samples were first fixed in 2.5% glutaraldehyde solution for 2 h. About 10 μl of the diluted mixtures were then transferred to a cleaned copper net and images were obtained by TEM (JEM-1010, JEOL, Japan) after staining with 2% phosphotungstic acid solution. For Nanosight analysis, fractionated samples were diluted 2,000-fold in PBS for size distribution analysis using a Zetasizer Nano ZS (Malvern Instruments Ltd, UK). For flow cytometry analysis, the fractions 3 and 5 were first resuspended in 100 μl PBS and then incubated with anti-CD63 and anti-CD81 specific monoclonal antibodies (BD Biosciences, San Jose, CA, USA) with fluorescent direct labeling. BD Accuri C6 Flow Cytomenter (BD Accuri, San Jose, CA, USA) was used to examine the characteristic protein markers of the two fractions.

DNA Extraction and Quantification

For each SCLC patient, 1 ml platelet-rich plasma (one spin at 3,000 rpm for 10 min) was used for fraction separation. For each NSCLC patient, 2 ml platelet-poor plasma (double spins at 3,000 rpm for 10 min) was used for DNA extraction and subsequent mutational analysis. DNAs from all samples were extracted using DNA Blood Mini Kit (Qiagen, Valencia, CA, USA). Final DNA eluent (50 μl) was quantified by Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). High Sensitivity DNA Analysis Chip (Agilent Technologies, Santa Clara, CA, USA) was used to examine the DNA size distribution of different fractions. The extracted DNA was stored at −20°C until use.

Library Preparation and Whole Genome Sequencing

DNA libraries were prepared using a ThruPLEX DNA-seq Library Kit (Takara Bio, Mountain View, CA) according to the manufacturer's instructions. About 0.5–1 ng DNA was used for library preparation including end repair, adaptor addition, and 15 cycles of high-fidelity amplification. Following amplification, libraries were purified using a 1:1 ratio of DNA sample to Agencourt AMPure XP Beads (Beckman Coulter, Indianapolis,

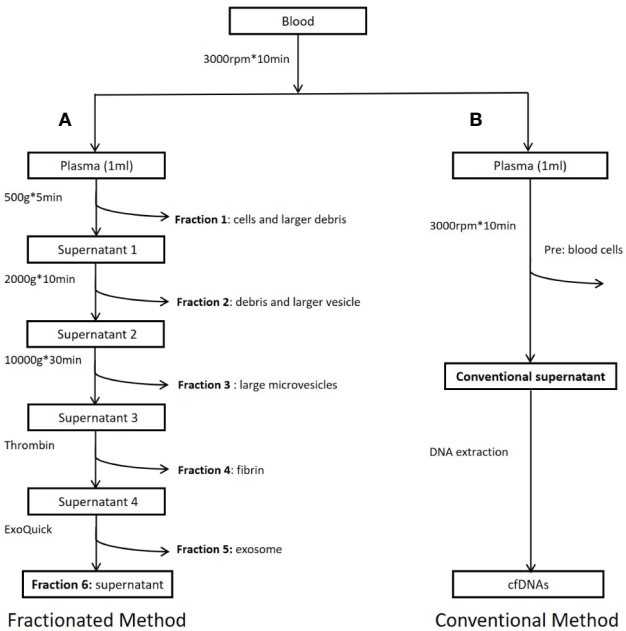


FIGURE 1 | Workflow of study design. **(A)** Preparation of plasma fractions by five consecutive centrifugations. **(B)** Conventional plasma cfDNA extraction. Fraction 1: Precipitates after centrifugation at 500g. Major components of this fraction are some cells and large cell debris. Fraction 2: Precipitates after centrifugation at 2,000*g* from fraction 1 supernatant. Major components of this fraction are small cell debris and large vesicles, such as apoptotic bodies. Fraction 3: Precipitates after centrifugation at 10,000*g* from fraction 2 supernatant. Major component of this fraction is large microvesicles. Fraction 4: Precipitates after centrifugation at 10,000*g* from fraction 3 supernatant treated with Thrombin at room temperature for 5 min (9.5 ul Thrombin per 950 ul sample). Major component of this fraction is fibrin. Fraction 5: Precipitates after centrifugation at 1,500*g* from fraction 4 supernatant treated with Exoquick (System Biosciences, Mountain View, CA, USA) at 4°C overnight. Major component of this fraction is believed to be exosomes. Fraction 6: Supernatant from fraction 5. This fraction is the leftover supernatant after removing precipitates from five consecutive centrifugations.

IN, USA). The library quality and insert size were examined using High Sensitivity DNA Analysis Chip. Sequencing Libraries were diluted to a concentration of 10 nM and every 12 index libraries were pooled for 50 bp single-read sequencing on a HiSeq2500 Sequencing System (Illumina, San Diego, CA, USA).

Copy Number Variation (CNV) and ctDNA Content Calculation

Raw sequencing data (fastq files) were first mapped to the human reference genome (NCBI37/hg19) using SeqMan NGen 12 (DNASTAR, Madison, WI, USA) and assembled in Partek Genomics Suite (St. Louis, MO, USA). The mapped reads were then binned into 1 Mb genomic bins and rescaled to 10 million reads after excluding sex chromosomes. Read count in each genomic window was normalized to mean read count from 33 healthy controls as previously described (19). The resulting ratios were further transformed with log2 and adjusted for GC content (20). The fully normalized log2 ratios in genomic bins were subjected to segmentation using the copy number analysis method (CNAM) algorithm (Golden Helix, Bozeman, MT, USA). To estimate ctDNA content, we developed a CNV-based algorithm to quantify ctDNA percentage in plasma cfDNA (21). In these studies, we used mean log2 values of genomic segments generated from CNAM algorithm for ctDNA content calculation. Segment sizes were evaluated to test ctDNA content stability. We selected mean log2 values from most significant deletion segments (>20 Mb in size) in each patient and calculated ctDNA content by $1–2^{\text{segment log2 ratio}}$ (21).

EGFR Mutational Analysis

For 22 NSCLC patients with known EGFRE19del/L858R mutations in tumor tissues, we applied ARMS-PCR method and tested their mutational status in fraction 6 and platelet-poor plasma samples using EGFR Mutations Detection Kit (Amoy Diagnostics, Xiamen, Fujian, China). Mutational analysis was performed on the ABI 7500 Real-Time PCR System (Thermo Fisher Scientific, Foster City, CA, USA). PCR was prepared by mixing 5 µl DNA with 5 µl control reaction mix or mutation mix (E19del, L858R). PCR was set up as follows: 95°C/5 min; 15 cycles of 95°C/25 s, 64°C/20 s and 72°C/20 s; 31 cycles of 93°C/25 s, 60°C/35 s and 72°C/20 s. The signal is collected at 60°C in the third stage.

RESULTS

Plasma Fractions

To estimate cfDNA size and ctDNA content from different plasma fractions, we selected nine SCLC patients (**Table S1**) who have been previously analyzed and showed relatively high tumor burden (17). As described in previous publications,

increasing centrifugation speed may generate sequential precipitations of different components including cells, cell debris, larger vesicles, apoptotic bodies, and microvesicles (11, 12, 22, 23). Based on these publications, we performed sequential centrifugation using 1 ml plasma sample/patient and collected six fractions for separate cfDNA extraction. **Figure 1** shows overall workflow of this study. To test higher ctDNA content in a plasma fraction, we compared mutation detection rate between the plasma fraction and whole plasma in 22 NSCLC patients with known EGFR mutations in tumor tissues.

Characterization of Extracellular Vesicles in Fractions 3 and 5

We first applied transmission electron microscopy and nanoparticle tracking analysis (Nanosight) to estimate vesicle size in the fractions 3 and 5 (**Figures 2A, B**). This analysis showed that all particles in fraction 3 were within the range 200–600 nm with peak size at 405.5 nm while 83.7% particles in fraction 5 were within 20–200 nm with main size at 100.3 nm. We then applied flow cytometry to examine characteristic protein markers in the two fractions. This analysis showed that CD63 and CD81 positive ratio were 26.4 and 11.3% in fraction 3, and 66.4 and 88.2% in fraction 5, respectively (**Figures 2C, D**). Clearly, the fraction 3 is featured as large microvesicles while fraction 5 is featured as exosomes.

DNA Yield in Each Plasma Fraction

To evaluate DNA yield from each plasma fraction, we isolated DNA and quantified the DNA concentration using a high sensitive Qubit assay in a total of 54 fractionated biospecimens. From 1 ml starting plasma, an average yield of each individual fractions was 5.03 ng (median = 1.66 ng, range 0.23–17.01 ng) in fraction 1, 1.73 ng (median = 0.86 ng, range 0.47–5.15 ng) in fraction 2, 0.99 ng (median = 0.50 ng, range 0.18–2.84 ng) in fraction 3, 0.68 ng (median = 0.40ng, range 0.18–1.48 ng) in fraction 4, 4.17 ng (median = 2.65 ng, range 0.58–13.05 ng) in fraction 5 and 4.28 ng (median = 1.55 ng, range 0.22–12.15 ng) in fraction 6. Although the average DNA yields in fractions 1, 5 and 6 were among top three and accounted for 79.9% of all DNA yields, their variations were also among the top three. In contrast, fractions 2–4 showed relatively low but stable DNA yield (**Figure 3A**).

DNA Size Distribution in Different Plasma Fractions

To investigate DNA size distribution, we measured each of the six DNA samples using Agilent Bioanalyzer. This analysis revealed distinct peak sizes in different fractions. The fraction 1 showed a peak size of 7,000–10,000 bp which was gradually reduced in fractions 2–3. Although barely seen in fraction 1, the density of a smaller fragment at ~160 bp was gradually intensified from fractions 3 to 6 (**Figure 3B**). In some samples, the fraction 6 showed a peak size at ~140 bp (**Figure 3C**). The fragment sizes are similar to the DNA length of a mono-nucleosome (~147 bp) (23). Overall, we observed clear trend that the larger fragment (~10,000 bp) was slowly diminished from fractions 1 to 3 while smaller fragment (~160 bp) was gradually increased from fractions 4 to 6. Clearly, larger DNA fragments in fractions 1 and 2 are more likely derived from genomic DNA contamination of cell debris and platelets.

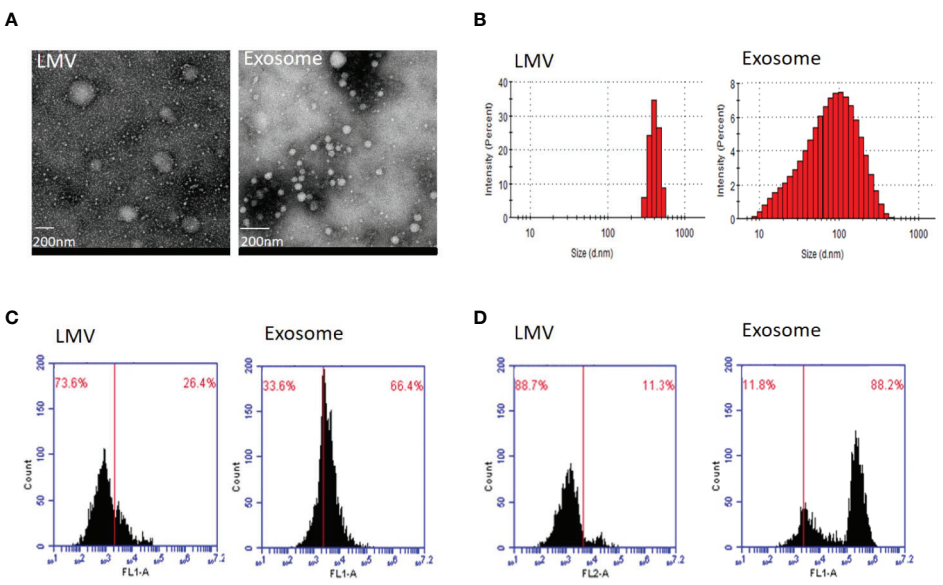


FIGURE 2 | Identity analysis of fraction 3 (large microvesicles) and fraction 5 (exosomes). **(A)** Transmission electron microscopy. The round shape of large microvesicles (LMV) and exosomes by negatively staining the background with phosphotungstic acid. The bar represents 200 nm. **(B)** Nanosight analysis. Particle sizes of fractions 3 and 5 are different with 405.5 nm and 100.3 nm in the main peak value, respectively. **(C, D)** Flow cytometry of characteristic protein analysis. Results of CD63 **(C)** and CD81 **(D)** positive ratio show 26.4 and 11.3% in fraction 3, and 66.4 and 88.2% in fraction 5, respectively.

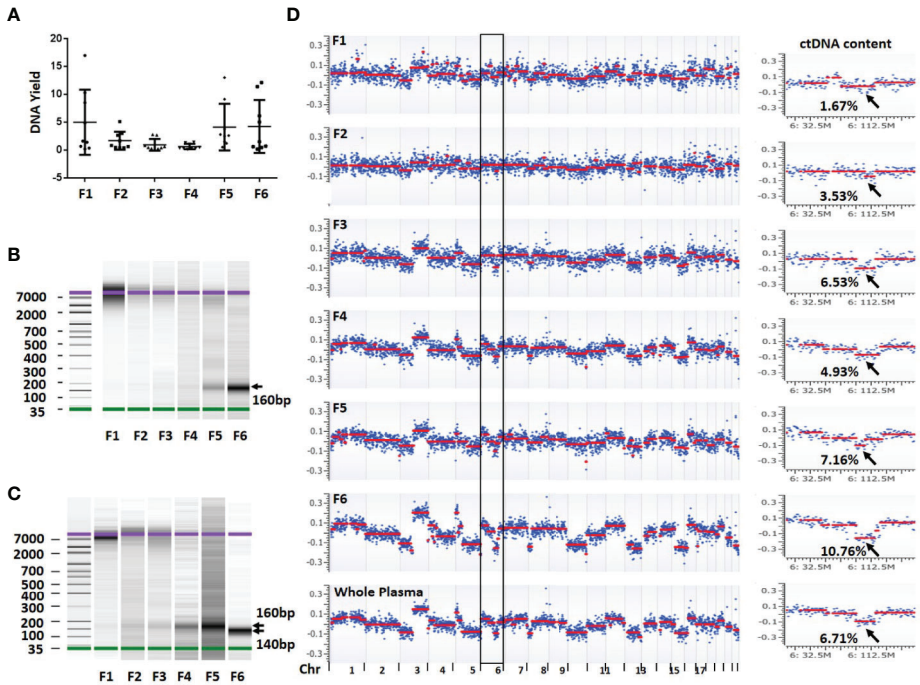


FIGURE 3 | DNA yield, size and overall view of genomic alterations in six fractions collected from a sequential centrifugation of 1 ml plasma sample. **(A)** DNA yield (ng) from six different fractions. **(B)** DNA size (bp) distribution in six plasma fractions of patient 1. **(C)** DNA size (bp) distribution in six plasma fractions of patient 7. **(D)** Overall view of genomic alterations in six plasma fractions and their corresponding whole plasma from patient 1. Segmentation-based copy number variation analysis shows different genomic variations across chromosomes 1–22. Most significant segments losses (arrows) on chromosome 6 were used to calculate ctDNA content. The log2 ratio scale in y axis was from –0.4 to 0.4. F1–6 represent fractions 1-6, respectively.

ctDNA Content in Different Fractions of Plasma

To estimate ctDNA content (defined as ctDNA percentage in a total cfDNA) from each plasma fraction, we first performed low-pass whole genome sequencing and received approximately 20 million (range 9.4–42.1 million) mappable reads per fraction (**Table S2**). We then performed log2 ratio-based segmentation analysis using 1 Mb genomic windows. This analysis showed a significant change of detectable copy number among different fractions. In general, the fractions 1 and 2 might show detectable copy number variations (CNVs) but call confidence was relatively low. In contrast, the fractions 3 (large EVs), 5 (exosomes) and 6 (EVs-depleted plasma) were more likely to demonstrate detectable CNVs with high confidence call (**Figure 3D**).

To calculate ctDNA content, we selected mean value from most significantly deleted segments in each patient and estimated ctDNA proportion in each individual fraction (21). This analysis showed that the average ctDNA content was 12.12% (median = 9.30%, range 0.39–27.09%) in fraction 1, 14.25% (median = 16.93%, range 3.53–22.82%) in fraction 2, 19.70% (median = 20.18%, range 6.53–37.94%) in fraction 3, 19.23% (median = 18.42%, range 4.93–38.60%) in fraction 4, 22.09% (median = 20.45%, range 7.16–40.51%) in fraction 5, and 27.22% (median = 27.04%, range 10.76–40.12%) in fraction 6. Clearly, ctDNA content in EVs-depleted fraction 6 was the highest among the six components. Fractions 5 (exosome) and 3 (large EVs) showed the 2nd and 3rd highest ctDNA content, respectively (**Figures 4A, B** and **Table S3**). We also compared the fractionated plasma DNA to platelet-poor plasma DNA for their ctDNA content differences. The ctDNA content in platelet-poor plasma showed an average of 23.84% (median = 23.34%, range 6.71–41.22%), further support that the fraction 6 has the highest ctDNA content among all plasma fractions and whole platelet-poor plasma.

To further demonstrate ctDNA content difference between fraction 6 and whole plasma, we performed clustering analysis using GC-corrected log2 ratio as input. Although fraction 6 and platelet-poor plasma from the same patients clustered perfectly across all chromosome regions, the heatmap showed clear intensity differences in most regions showing CNVs (**Figure 4C**). Of nine cases, seven showed higher intensity (hence, higher ctDNA content) in fraction 6 than in whole plasma. For example, based on mean absolute log2 ratios at these selected genomic segments (**Table S4**), we estimated that ctDNA content in patient 7 was 37.2% in fraction 6 while 23.3% in platelet-poor plasma, indicating 13.9% more ctDNA content in fraction 6 than whole plasma sample in the patient.

Detection of EGFR Mutations in Fractionated Plasma and Platelet-Poor Plasma

Since fraction 6 showed the highest ctDNA content, we hypothesized that the fraction 6 had higher sensitivity in

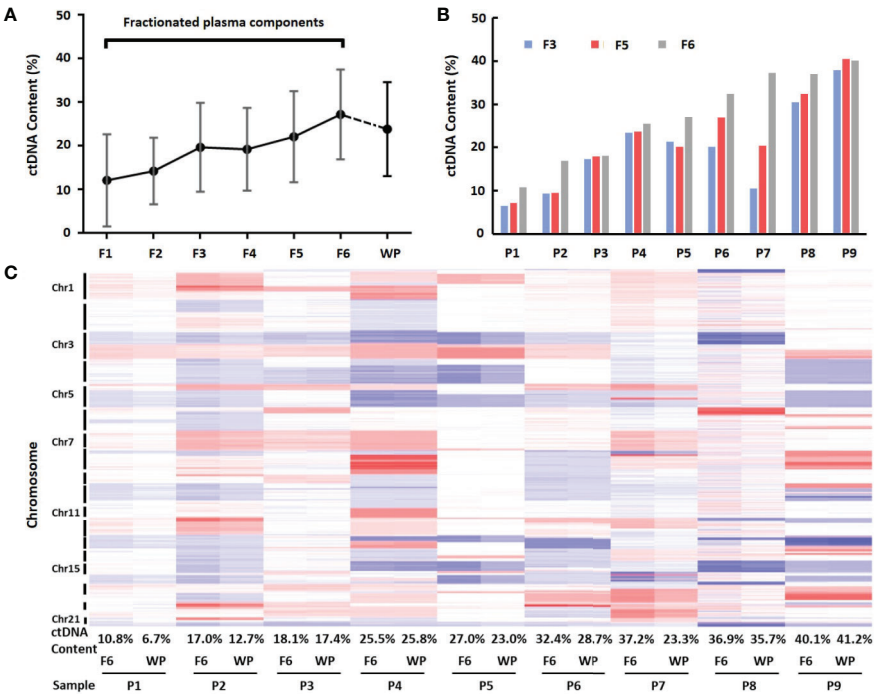


FIGURE 4 | Differences of ctDNA content in fractionated plasma components and whole plasma. **(A)** Overall view of ctDNA content from fractionated plasma components and whole plasma. The average ctDNA content is the highest in EVs-depleted fraction 6. **(B)** Differences of ctDNA content among fractions 3, 5 and 6. **(C)** Heatmap of log2 ratio in 1 Mb genomic window across chromosomes 1–22 in fraction 6 (F6) and whole plasma (WP) from nine patients (P1–P9). Red color represents copy number gain, while blue represents loss. Intensity of the color is proportional to the value of log2 ratio and reflects the weight of ctDNA in overall background cfDNA.

mutation detection. To test this, we selected 22 non-small cell lung cancer (NSCLC) patients with known EGFR E19del/L858R mutations in tumor tissues (**Table S5**). The positive percentage of serum tumor biomarkers including CEA (Carcino-embryonic antigen) and CYFRA 21-1(Cytokeratin-19-fragment) was 36.4 and 40.9%, respectively, in these NSCLC patients. We applied the amplification-refractory mutation system (ARMS)-PCR assays to detect these mutations in the DNAs derived from the fraction 6 and platelet-poor plasma. Among the 22 patients, we identified EGFR mutations that matched to tumor tissues in 14 of fraction 6 samples and 10 of platelet-poor plasma samples. Sensitivity of the EGFR mutation detection was 63.6% (95% CI: 40.8 to 82.0%) in fraction 6 and 45.5% (95% CI: 25.1 to 67.3%) in platelet-poor plasma, respectively (**Table 1**). This result suggests that compared to traditionally used platelet-poor plasma, the fraction 6 derived from a series of centrifugations including removal of EVs may improve EGFR mutation detection.

DISCUSSION

It is well known that ctDNAs are detectable in plasma samples of peripheral blood (24–26). The ctDNAs appear to demonstrate unique DNA fragmentation pattern and smaller fragment size (27). However, effect of plasma preparation methods on ctDNA content has not been reported. In this study, we isolated DNA from six fractions of plasma samples by multiple physical and chemical precipitations. We applied low-pass whole genome sequencing technology to determine CNVs for ctDNA content estimation. Our results showed that DNA fragment size and ctDNA content varied among the six fractions with fraction 6 showing enrichment of smaller DNA fragments and tumor-derived cfDNA. Fraction 6 also showed higher sensitivity in mutation detection than whole (unfractionated) plasma. These results suggest that plasma preparation before DNA extraction is an important step for sensitive detection of low level ctDNA in peripheral blood.

By separating whole plasma into six fractions, we were able to compare DNA yield, size distribution and ctDNA content differences among these fractionated samples. For fractions 1–3, ~10,000 bp DNA fragments are dominant but total DNA yields are gradually decreased. Since fractions 1–2 are primarily composed of contaminated cell debris, platelets and larger vesicles such as apoptotic bodies, it is not surprised to see higher molecule weight DNA fragments. The fraction 3 is believed to contain primarily large EVs, which have shown predominantly large size (~10,000 bp) dsDNA by chip-based capillary electrophoresis (11), which is consistent with our observation. Additionally, a recent report showed that centrifugation protocols had an effect on DNA integrity (28).

TABLE 1 | Comparison of the EGFR mutation status between fraction 6 DNA and cell free DNA in NSCLC patients.

EGFR genotype	Tissue	Plasma (n = 22)			
		Fraction 6 DNA		cfDNA	
		Mutant type	Wild type	Mutant type	Wild type
Mutant type	22 (100.0%)	14 (63.6%)	0	10 (45.5%)	0
Wild type	0	0	8 (36.4%)	0	12 (54.5%)
Sensitivity (%) (95% CI)		63.6% (40.8–82.0)		45.5% (25.1–67.3)	
Specificity (%) (95% CI)		NA		NA	

CI, Confidence Interval; NA, not available.

This study reported longer DNA fragments almost exclusively in CPBasic Fraction (plasma after 400*g* for 10 min centrifugation) and CPAdBasic_P Fraction (pellet after 400*g* for 10 min and max speed for 1 min), which is also consistent with our observation. Fractions 4–5 are dominated by 160 bp fragments. This DNA size is similar to commonly reported cfDNA and is corresponding to the size of chromatosomes (nucleosome + linker histone; ~167 bp) (23). Fraction 4 is thrombin-precipitated fibrin. Thrombin acts as a serine protease that converts soluble fibrinogen into insoluble strands of fibrin, as well as catalyzing many other coagulation-related reactions. Fraction 5 is derived from Exoquick-precipitated exosomes and other small EVs. Studies have shown that the exosomes contain DNA from parent tumor cells (10–12). Interestingly, DNA sizes from fraction 6 seem sample-dependent with some samples being at ~160 bp while others showing ~140 bp. Fraction 6 is the leftover supernatant after five consecutive precipitations and can be considered as EVs-depleted plasma. Clearly, the smaller mono-nucleosome derived DNA has been preserved after multiple centrifugations. Further study is needed to determine whether the mono-nucleosome sized DNA is free in true free state or in histone-bound state.

By comparing different plasma fractions, we observed a clear trend of higher ctDNA content in fractions 3 (large EVs), 5 (exosomes) and 6 (EVs-depleted plasma). Consistent with our findings, a previous study showed that large vesicles from cancer patients enriched ctDNA (11). Another study showed that DNA from nanoscale vesicles (30–220 nm, the size of exosomes) is better than whole plasma cfDNA for mutation detection in early stage NSCLC (12). It is known that ctDNA tends to be shorter than normal cfDNA in plasma (29). An animal model-based study demonstrated that the most common fragment length of ctDNA was 134–144 bp, which is significantly smaller than the most common 167 bp fragment present in noncancer cfDNA (30). Since fraction 6 has smaller fragments (~140 bp) than any other fractions, it may be one reason to explain why fraction 6 shows an increased ctDNA content and thus higher sensitivity in mutation detection. Additionally, detection of EGFR mutations in plasma samples of NSCLC patients are predictive of survival and resistance to EGFR TKI (31). Therefore, our results strongly support that the enriched ctDNA in fraction 6 will increase mutation detection sensitivity and facilitate identification of tumor-specific biomarkers.

An innovative feature of this study is the assessment of ctDNA content in multiple fractions of plasma. Since all fractions were derived from the same 1 ml of plasma by consecutive centrifugations, we were able to directly compare ctDNA contents in different fractions from the same patients. Another feature is ctDNA content estimation using a novel algorithm. To determine ctDNA content, mutant allele frequency is commonly used. However, it is difficult to calculate ctDNA content when cfDNA input is low. In this study, we applied a CNV-based algorithm to estimate ctDNA content (21). The CNV-based method uses an average log2 ratio values across multiple genomic bins (windows). Therefore, the estimate is expected to be more stable when compared to single mutant allele-based method. Additionally, we selected patients with high tumor burden from our previous study (17). The high tumor burden is necessary to accurately determine ctDNA content and to demonstrate difference of the ctDNA content among these fractions. It is worth mentioning that the plasma fractionation may allow maximum use of valuable plasma samples for a wide variety of studies. For example, supernatant fraction of a plasma sample may be used for ctDNA-based genetic analysis while exosome fraction of the same plasma sample may be used for microRNA-based biomarker study. This approach resembles blood transfusion of components in clinic to efficiently use different blood fractions.

Although we observed significant ctDNA content differences in fractionated plasma components and showed higher sensitivity in mutation detection in EV-depleted plasma fraction, this study also has some limitations. First, we were not able to remove any possible DNA that may be co-purified with exosomes in fraction 5. A new study has shown that extracellular DNA could be co-purified with the small EV fraction during standard isolation protocols (16). Therefore, the origin of ctDNAs detected in fraction 5 needs further investigation. Second, we made double stranded DNA library for sequencing analysis. It seems that double stranded DNA might not be associated with exosomes or with any small EVs at all (32). Single strand DNA library preparation method may be needed to evaluate the presence of ctDNA in exosomes. Third, we tested plasma EGFR mutations in 22 patients only. Although the EV-depleted plasma fraction showed higher mutation detection rate than whole plasma (63.6% vs 45.5%), the difference did not reach statistical significance. Further study in large sample size is needed. Fourth, the current study used a sequential centrifugation process, which was involved in multiple pipetting and sample transfer, significantly increasing risk of sample contamination. Future study to optimize the

centrifugation steps will help simplify the sample processing. Finally, because tumor genome is evolving during disease progression, it is interesting to analyze ctDNA content changes in different plasma fractions at different blood draw time points. Nevertheless, our study provided a new insight into potential application of fractionated plasma for an improved ctDNA detection. The result supports that different plasma fractions may enrich different types of tumor-associated molecules. Further understanding of DNA origins in different plasma fractions will facilitate cancer biomarker discovery.

CONCLUSIONS

cfDNA from different fractions of plasma varies in fragmentation sizes and ctDNA contents. Due to its higher ctDNA content and increased sensitivity of mutation detection, the fraction 6 is the preferred source of material for ctDNA-based genomic analysis.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical College of Wisconsin Institutional Review Broad. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Conceptualization, LS and LW. Methodology, LS and MD. Writing—original draft preparation, LS. Writing—review and editing. MK, LW, HS, SW, C-CH, XC, and MX. Funding acquisition, LW, SW, and LS. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by National Institute of Health (grant number R01CA212097) to LW, Key Project of Science and National Natural Science Foundation of China (No. 81972806) to SW, Key Project of Science and National Natural Science Foundation of China (No. 82002230), Medical Scientific Research Guidance Project of the Health Commission of Jiangsu (No. Z2020007), 789 Outstanding Talent Program of SAHNMU (No. 789ZYRC202080338), and International Exchange and Cooperation Foundation of Nanjing Medical University (No. C063, D009) to LS.

ACKNOWLEDGMENTS

We thank the MCW Tissue Bank for biospecimens, the Human Molecular Genetics Center Sequencing Core at Medical College of Wisconsin for providing DNA sequencing service, and the Jiangsu Cancer Personalized Medicine Collaborative Innovation Center for supports.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.691798/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic Instability–An Evolving Hallmark of Cancer. *Nat Rev Mol Cell Biol* (2010) 11(3):220–8. doi: 10.1038/nrm2858
3. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* (2012) 366(10):883–92. doi: 10.1056/NEJMoa1113205
4. Kaiser J. Medicine. Keeping Tabs on Tumor DNA. *Science* (2010) 327 (5969):1074. doi: 10.1126/science.327.5969.1074
5. Almodovar K, Iams WT, Meador CB, Zhao Z, York S, Horn L, et al. Longitudinal Cell-Free DNA Analysis in Patients With Small Cell Lung Cancer Reveals Dynamic Insights Into Treatment Efficacy and Disease Relapse. *J Thorac Oncol* (2018) 13(1):112–23. doi: 10.1016/j.jtho.2017.09.1951
6. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable Whole-Exome Sequencing of Cell-Free DNA Reveals High Concordance With Metastatic Tumors. *Nat Commun* (2017) 8(1):1324. doi: 10.1038/s41467-017-00965-y
7. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating Liquid Biopsies Into the Management of Cancer. *Nat Rev Clin Oncol* (2017) 14(9):531–48. doi: 10.1038/nrclinonc.2017.14
8. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct Detection of Early-Stage Cancers Using Circulating Tumor DNA. *Sci Transl Med* (2017) 9(403):eaan2415. doi: 10.1126/scitranslmed.aan2415
9. Choudhury AD, Werner L, Francini E, Wei XX, Ha G, Freeman SS, et al. Tumor Fraction in Cell-Free DNA as a Biomarker in Prostate Cancer. *JCI Insight* (2018) 3(21):e122109. doi: 10.1172/jci.insight.122109
10. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, et al. Double-Stranded DNA in Exosomes: A Novel Biomarker in Cancer Detection. *Cell Res* (2014) 24(6):766–9. doi: 10.1038/cr.2014.44
11. Vagner T, Spinelli C, Minciacchi VR, Balaj L, Zandian M, Conley A, et al. Large Extracellular Vesicles Carry Most of the Tumour DNA Circulating in Prostate Cancer Patient Plasma. *J Extracell Vesicles* (2018) 7(1):1505403. doi: 10.1080/20013078.2018.1505403
12. Wan Y, Liu B, Lei H, Zhang B, Wang Y, Huang H, et al. Nanoscale Extracellular Vesicle-Derived DNA is Superior to Circulating Cell-Free DNA for Mutation Detection in Early-Stage Non-Small-Cell Lung Cancer. *Ann Oncol* (2018) 29(12):2379–83. doi: 10.1093/annonc/mdy458
13. Mohrmann L, Huang HJ, Hong DS, Tsimberidou AM, Fu S, Piha-Paul SA, et al. Liquid Biopsies Using Plasma Exosomal Nucleic Acids and Plasma

Cell-Free DNA Compared With Clinical Outcomes of Patients With Advanced Cancers. *Clin Cancer Res* (2018) 24(1):181–8. doi: 10.1158/1078-0432.CCR-17-2007
14. Hur JY, Kim HJ, Lee JS, Choi CM, Lee JC, Jung MK, et al. Extracellular Vesicle-Derived DNA for Performing EGFR Genotyping of NSCLC Patients. *Mol Cancer* (2018) 17(1):15. doi: 10.1186/s12943-018-0772-6
15. Allenson K, Castillo J, San Lucas FA, Scelo G, Kim DU, Bernard V, et al. High Prevalence of Mutant KRAS in Circulating Exosome-Derived DNA From Early-Stage Pancreatic Cancer Patients. *Ann Oncol* (2017) 28(4):741–7. doi: 10.1093/annonc/mdx004
16. Jeppesen DK, Fenix AM, Franklin JL, Higginbotham JN, Zhang Q, Zimmerman LJ, et al. Reassessment of Exosome Composition. *Cell* (2019) 177(2):428–45.e18. doi: 10.1016/j.cell.2019.02.029
17. Du M, Thompson J, Fisher H, Zhang P, Huang CC, Wang L. Genomic Alterations of Plasma Cell-Free DNAs in Small Cell Lung Cancer and Their Clinical Relevance. *Lung Cancer* (2018) 120:113–21. doi: 10.1016/j.lungcan.2018.04.008
18. Xia S, Huang CC, Le M, Dittmar R, Du M, Yuan T, et al. Genomic Variations in Plasma Cell Free DNA Differentiate Early Stage Lung Cancers From Normal Controls. *Lung Cancer* (2015) 90(1):78–84. doi: 10.1016/j.lungcan.2015.07.002
19. Li J, Dittmar RL, Xia S, Zhang H, Du M, Huang CC, et al. Cell-Free DNA Copy Number Variations in Plasma From Colorectal Cancer Patients. *Mol Oncol* (2017) 11(8):1099–111. doi: 10.1002/1878-0261.12077
20. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of Genomic Waves in Signal Intensities From Whole-Genome SNP Genotyping Platforms. *Nucleic Acids Res* (2008) 36(19):e126. doi: 10.1093/nar/gkn556
21. Du M, Tian Y, Tan W, Wang L, Wang L, Kilari D, et al. Plasma Cell-Free DNA-Based Predictors of Response to Abiraterone Acetate/Prednisone and Prognostic Factors in Metastatic Castration-Resistant Prostate Cancer. *Prostate Cancer Prostatic Dis* (2020) 23(4):705–13. doi: 10.1038/s41391-020-0224-4
22. Rider MA, Hurwitz SN, Meckes DGJr. ExtraPEG: A Polyethylene Glycol-Based Method for Enrichment of Extracellular Vesicles. *Sci Rep* (2016) 6:23978. doi: 10.1038/srep23978
23. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-Free DNA Comprises an *In Vivo* Nucleosome Footprint That Informs its Tissues-of-Origin. *Cell* (2016) 164(1-2):57–68. doi: 10.1016/j.cell.2015.11.050
24. Siravegna G, Bardelli A. Genotyping Cell-Free Tumor DNA in the Blood to Detect Residual Disease and Drug Resistance. *Genome Biol* (2014) 15(8):449. doi: 10.1186/s13059-014-0449-4
25. Sakata-Yanagimoto M, Nakamoto-Matsubara R, Komori D, Nguyen TB, Hattori K, Nanmoku T, et al. Detection of the Circulating Tumor DNAs in Angioimmunoblastic T- Cell Lymphoma. *Ann Hematol* (2017) 96(9):1471–5. doi: 10.1007/s00277-017-3038-2
26. Oellerich M, Schutz E, Beck J, Walson PD. Circulating Cell-Free DNA-Diagnostic and Prognostic Applications in Personalized Cancer Therapy. *Ther Drug Monit* (2019) 41(2):115–20. doi: 10.1097/FTD.0000000000000566
27. Huang CC, Du M, Wang L. Bioinformatics Analysis for Circulating Cell-Free DNA in Cancer. *Cancers (Basel)* (2019) 11(6):805. doi: 10.3390/cancers11060805
28. Sorber L, Zwaenepoel K, Jacobs J, De Winne K, Goethals S, Reclusa P, et al. Circulating Cell-Free DNA and RNA Analysis as Liquid Biopsy: Optimal Centrifugation Protocol. *Cancers (Basel)* (2019) 11(4):458. doi: 10.3390/cancers11040458
29. Mouliere F, Rosenfeld N. Circulating Tumor-Derived DNA is Shorter Than Somatic DNA in Plasma. *Proc Natl Acad Sci USA* (2015) 112(11):3178–9. doi: 10.1073/pnas.1501321112
30. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment Length of Circulating Tumor DNA. *PloS Genet* (2016) 12(7):e1006162. doi: 10.1371/journal.pgen.1006162
31. Phan TT, Tran BT, Nguyen ST, Ho TT, Nguyen HT, Le VT, et al. EGFR Plasma Mutation in Prediction Models for Resistance With EGFR TKI and Survival of Non-Small Cell Lung Cancer. *Clin Transl Med* (2019) 8(1):4. doi: 10.1186/s40169-019-0219-8
32. Pasini L, Ulivi P. Extracellular Vesicles in Non-Small-Cell Lung Cancer: Functional Role and Involvement in Resistance to Targeted Treatment and Immunotherapy. *Cancers (Basel)* (2019) 12(1):40. doi: 10.3390/cancers12010040

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 in Genetics

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fgene.2021.669605

Predictive Biomarkers of Dicycloplatin Resistance or Susceptibility in Prostate Cancer

*Minglu Liu[1], Xiaoyu Zhou[2], Jun Liu[2], Chelong Lu[2], Guoqing Zhang[1], Jing Zhang[1] and Shunchang Jiao[1]**

[1] Department of Medical Oncology, Chinese PLA General Hospital, Beijing, China, [2] GenomiCare Biotechnology Co., Ltd., Shanghai, China

Background: Prostate cancer (PCa) is among the leading causes of cancer mortality. Dicycloplatin is a newer generation platinum-based drug that has less side effects than cisplatin and carboplatin. However, its effects in PCa is mixed due to lack of appropriate stratifying biomarkers. Aiming to search for such biomarkers, here, we analyze a group of PCa patients with different responses to dicycloplatin.

Methods: We carried out whole-exome sequencing on cell-free DNA (cfDNA) and matched leukocyte DNA from 16 PCa patients before treatment with dicycloplatin. We then compared the clinical characteristics, somatic mutations, copy number variants (CNVs), and mutational signatures between the dicycloplatin-sensitive (nine patients) and dicycloplatin-resistant (seven patients) groups and tested the identified mutations, CNV, and their combinations as marker of dicycloplatin response.

Results: The mutation frequency of seven genes (*SP8*, *HNRNPCL1*, *FRG1*, *RBM25*, *MUC16*, *ASTE1*, and *TMBIM4*) and CNV rate of four genes (*CTAGE4*, *GAGE2E*, *GAGE2C*, and *HORMAD1*) were higher in the resistant group than in the sensitive group, while the CNV rate in six genes (*CDSN*, *DPCR1*, *MUC22*, *TMSB4Y*, *VARS*, and *HISTCH2AC*) were lower in the resistant group than in the sensitive group. A combination of simultaneous mutation in two genes (*SP8*/*HNRNPCL1* or *SP8*/*FRG1*) and deletion of *GAGE2C* together were found capable to predict dicycloplatin resistance with 100% sensitivity and 100% specificity.

Conclusion: We successfully used cfDNA to monitor mutational profiles of PCa and designed an effective composite marker to select patients for dicycloplatin treatment based on their mutational profile.

Keywords: prostate cancer, dicycloplatin, whole-exome sequencing, biomarker, anti-cancer (anticancer) drugs

OPEN ACCESS

Edited by:
Xiaoyan Zhou,
Fudan University, China

Reviewed by:
Jian Yu,
Beihang University, China
Sanjay Rathod,
University of Pittsburgh, United States

****Correspondence:***
Shunchang Jiao
jiaoshunchang@csco.org.cn

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

Received: *19 February 2021*
Accepted: *09 June 2021*
Published: *27 July 2021*

Citation:
Liu M, Zhou X, Liu J, Lu C, Zhang G, Zhang J and Jiao S (2021) Predictive Biomarkers of Dicycloplatin Resistance or Susceptibility in Prostate Cancer. Front. Genet. 12:669605. doi: 10.3389/fgene.2021.669605

INTRODUCTION

Prostate cancer (PCa) is one of the most common cancers in men worldwide and has the second highest mortality (Siegel et al., 2020). It was estimated that there would be about 190,000 new cases of PCa (21% of all are male cancers) and 30,000 deaths (10% of all are male cancer deaths) in the United States alone in 2020 (Siegel et al., 2020). Hormone therapy is an effective therapy that

can improve the survival time and clinical benefits for early stage PCa. However, about 10–20% of patients will inevitably develop into drug resistance within 5 years during the course of treatment, leading to castration-resistant prostate cancer (Kirby et al., 2011). Platinum-based chemotherapies, such as cisplatin and carboplatin, are attracting more and more attention in the treatment of cancer (Apps et al., 2015). These drugs mainly function through induction of DNA cross-links, therefore inhibiting DNA synthesis, mitosis, and induce apoptosis (Lokich and Anderson, 1998; Ozols et al., 2003). However, the clinical benefits from these drug therapies are still low, they can prolong patients' overall survival for only 3–6 months (Fortin et al., 2013). Moreover, their clinical application is limited by severe adverse effects, including ototoxicity, neurotoxicity, and myelosuppression (Rossi et al., 2012). Therefore, there is an urgent need to develop more effective drugs to PCa with lower side effects.

Dicycloplatin is a derivative of carboplatin in which a carboxylic acid ligand is bound to the carboplatin moiety through hydrogen bonds. Therefore, it has a more stable chemical structure and better aqueous solubility than carboplatin (Yang et al., 2010). Previous studies showed that dicycloplatin has a better anticancer activity and much lower toxicity than cisplatin and carboplatin in patients with non-small cell lung cancer (Wang and Yu, 2004). *In vivo* and *in vitro* studies showed that dicycloplatin can induce cell cycle arrest and apoptosis and inhibit cell proliferation through reactive oxygen species stress-mediated death receptor pathway and mitochondrial pathway (Yu et al., 2014). Furthermore, a phase II clinical trial in non-small cell lung cancer has demonstrated the safety and efficacy of dicycloplatin in combination with paclitaxel (Liu et al., 2014). However, a considerable number of patients did not benefit from dicycloplatin treatment with some unknown reasons. Effective and reliable prognostic factors, therefore, are desperately needed in order to target dicycloplatin to the subset of patients who would benefit most from the treatment.

Whole exome sequencing (WES) is a high-throughput sequencing technology that can explore the whole functional DNA sequence and genetics variations of each patient to uncover novel molecules that may be related to therapies. In this retrospective study, we carried out WES on cell-free DNA (cfDNA) from blood- and patient-matched leukocyte DNA in 16 PCa patients before they received dicycloplatin monotherapy. A comprehensive analysis was performed to search for the association between the clinical outcomes of dicycloplatin treatment and molecular characterization of the patients, such as somatic mutations, copy number variants, and mutational signatures.

MATERIALS AND METHODS

Patients, Clinical Evaluation, and Sample Collection

Sixteen PCa patients were retrospectively enrolled from Chinese PLA General Hospital with the following criteria: (Siegel et al., 2020) Eastern Cooperative Oncology Group performance status of 0–2; (Kirby et al., 2011) patients with distant metastasis who have received surgical castration or medical castration (serum testosterone $\leq$ 50 ng/dl or 1.7 nmol/ml); (Apps et al., 2015) at least 4 weeks after antiandrogenic therapy; and (Lokich and Anderson, 1998) the value of prostate specific antigen (PSA) was more than 2 ng/ml, and sustained increase more than 50% one week. The patients were on dicycloplatin treatment during 2016–2019. The study protocol was approved by the Ethics Committee of Chinese PLA General Hospital (approval number S2017-032-02). All patients provided a written informed consent.

The pathological diagnosis was performed by experienced pathologists of the hospital. Tumor response to the treatment was evaluated based on the patients' radiological images [computed tomography (CT) and magnetic resonance imaging (MRI)] according to the Response Evaluation Criteria in Solid Tumors, version 1.1 (Tirkes et al., 2013).

Whole-Exome Sequencing

Ten milliliter of blood specimens was collected before dicycloplatin treatment and used to prepare cfDNA by MagMAX(Cell-Free DNA Isolation Kit (Applied Biosystems(A29319) and leukocyte DNA by Maxwell® RSC Blood DNA Kit (Promega AS1400). cfDNA for WES needed to meet the following conditions: there is an obvious peak in the region of 100–300 bp, and the regional molarity of the peak was >0.7 of all contents between 100 bp and 42 kb (indicating that there is no contamination from large genomic DNA). The purified DNA was sonicated using a Covaris L220 sonicator and hybridized to the probes in SureSelect Human All Exon V5 kit (cat. # 5190-6209 EN, Agilent Technologies, Sta. Clara, CA, United States) to capture exonic DNA, then prepared to libraries using the SureSelectXT Low Input Target Enrichment and Library Preparation system (cat. # G9703-90000, Agilent Technologies). Paired end reads of 150 × 150 bp were generated from the libraries using an Illumina NovaSeq-6000 sequencer (Illumina, San Diego, CA, United States). Image analysis and base calling were done using the onboard RTA3 software (Illumina). After removing adapters and low-quality reads, the reads were aligned to National Center for Biotechnology Information (NCBI) human genome reference assembly hg19 using the Burrows–Wheeler Aligner alignment algorithm and further processed using the Genome Analysis Toolkit (GATK, version 3.5), including the GATK Realigner Target Creator to identify regions that needed to be realigned. Somatic mutations, including single-nucleotide variants (SNVs), indel, and copy number variation (CNV) were determined by a comparison between the aligned sequences from cfDNA and patient-matched leukocyte DNA using the MuTect/ANNOVAR/dbNSFP31, VarscanIndel, and CNVnator software, respectively, as previously reported (Zang et al., 2019). The mutational signature classification was based on COSMIC Mutational Signature (version 2—March 2015), which was generated from studies performed by others (Maitra et al., 2013; Alexandrov et al., 2015; Nik-Zainal et al., 2016). Tumor mutation burden (TMB) was defined as the total number of somatic non-synonymous mutations in each sample according to a previous method for WES data (Chalmers et al., 2017). All autosomal

microsatellite tracts containing 1–5 bp repeating subunits in length and comprising five or more repeats in GRCh37/hg19 were identified using MISA[1] and used to calculate microsatellite instability score (MSI). MSI score was calculated by the number of unstable microsatellite sites/total valid sites.

The WES data of each patients were submitted in NCBI with submission number SUB9593847 and accession number PRJNA727718.

Statistical Analysis

All statistical analyses were performed using R[2] or SPSS 25 for Windows (SPSS Inc., Chicago, IL, United States). Differences in the distribution of somatic mutations, mutational signatures, and clinical characteristics between patient subgroups were evaluated by the Fisher' exact test and Mann–Whitney *U* test for categorical and continuous parameters, respectively, and events of $p < 0.05$ were considered statistically significant.

RESULTS

Clinicopathological Characteristics of the Patients

In total, 16 patients treated with dicycloplatin were enrolled, and their baseline characteristics are shown in **Table 1**. Nine of the patients were sensitive to the treatment, including six with partial response (PR) and three with complete response (CR) evaluated by CT and MRI scans. The remaining seven patients were resistant to the treatment, including six with progressive disease (PD) and one with stable disease (SD). The overall tumor response rate was 56.3%. The representative diagnostic images are shown in **Figure 1**. The PSA and free PSA (fPSA) levels of each patients during treatment course were analyzed; both PSA and fPSA were decreased in patients with PR and CR, while they were increased in patients with PD and SD (**Supplementary Figure 1**). Patients who received endocrine therapy before dicycloplatin treatment were more likely resistant to dicycloplatin (**Table 1**, Fisher's exact test, $p = 0.003$). Other factors, such as age, smoking and drinking history, tumor stage, Gleason score, pre-dicycloplatin treatment history of surgery, chemotherapy, and radiotherapy, all showed no significant difference between the dicycloplatin-sensitive and dicycloplatin-resistant groups.

The Mutational Landscape of the Patients

The whole-exome DNAs of the patients were captured and sequenced on an Illumina platform. The average sequencing depth was ×925 for cfDNA libraries and ×117 for leukocyte DNA libraries. The sequences from cfDNA were compared to matched leukocyte DNA to give somatic genetic changes including SNV, indel, and CNV (see section "Materials and Methods"). The median TMB was similar between treatment-sensitive group (3.3 mutations/Mb) and treatment-resistant group (4.4 mutations/Mb), with no statistical difference (Mann–Whitney *U* test, $p = 0.680$). There were also no differences in MSI score, the proportions of gene amplifications, and deletions

[1]http://pgrc.ipk-gatersleben.de/misa/misa.html

[2]https://cran.r-project.org

TABLE 1 | Comparison of baseline characteristics between dicycloplatin-sensitive and dicycloplatin-resistant patients.

Baseline characteristic	Sensitive (*n* = 9)	Resistant (*n* = 7)	*p* value	All patients (*n* = 16)
Age, median (range), years	66 (61–79)	63 (51–72)	0.210	66 (51–79)
Smoking history, no. (%)			0.633	
Yes	4 (44.4)	2 (28.6)		6 (37.5)
No	5 (55.5)	5 (71.4)		10 (62.5)
Drinking history, no. (%)			0.596	
Yes	2 (44.4)	3 (42.9)		6 (37.5)
No	7 (55.5)	4 (57.1)		10 (62.5)
Tumor stage at diagnosis, no. (%)			1	
II	1 (11.1)	0 (0.0)		1 (6.3)
IV	8 (88.8)	7 (100.0)		15 (93.8)
Gleason sum at diagnosis, no. (%)			1	
≤7	6 (66.6)	4 (57.1)		10 (62.5)
≥8	3 (33.3)	3 (42.9)		6 (37.5)
Prior treatment for PCa, no. (%)				
Surgery	2 (22.2)	1 (14.3)	1	3 (18.8)
Radiotherapy	1 (11.1)	2 (28.6)	0.4	3 (18.8)
Chemotherapy	3 (33.3)	5 (71.4)	0.157	10 (62.5)
Endocrine therapy	2 (22.2)	7 (100.0)	0.003	9 (56.3)
Pretreatment PSA level, median (range), ng/ml	87.6 (9.9–775.6)	79.3 (2.9–800)	0.837	86.0 (2.9–800)
Pretreatment-free PSA level, median (range), ng/ml	14.0 (0.5–48.6)	10.8 (0.9–29.8)	0.681	11.2 (0.5–48.6)

PSA, prostate-specific antigen.

p values are based on Fisher' exact test and Mann–Whitney U test for categorical and continuous parameters, respectively.

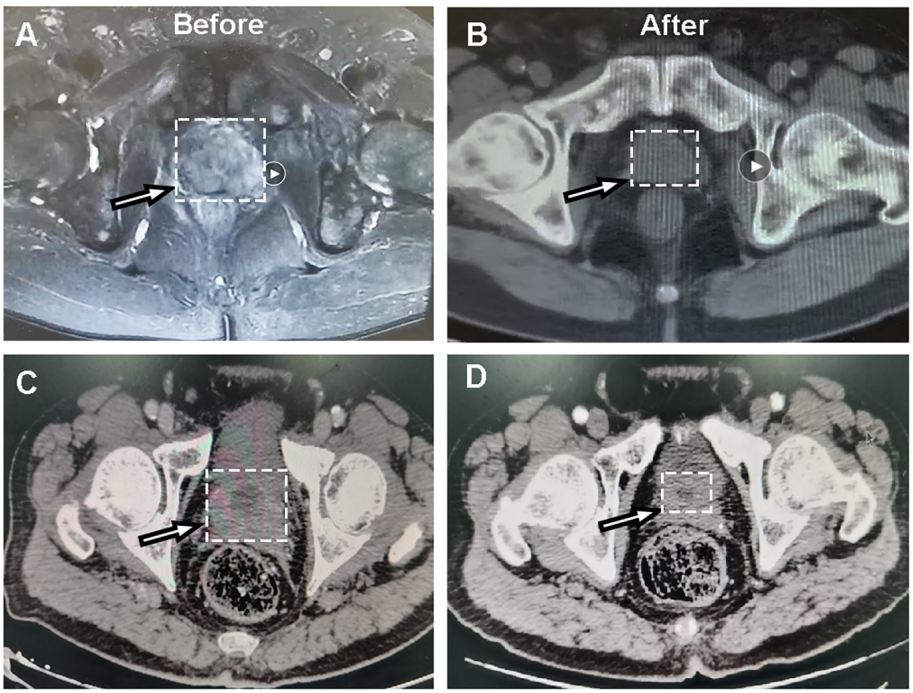


FIGURE 1 | Representative CT images of dicycloplatin-sensitive patients. **(A,B)** CT images of one patient with complete response taken before **(A)** and after **(B)** dicycloplatin treatment. **(C,D)** CT images of another patient with partial response taken before **(C)** and after **(D)** dicycloplatin treatment. White arrow represents the tumor localization.

between the two groups (Mann–Whitney *U* test, $p > 0.05$; **Figure 2**).

The most common somatic mutated genes were *ANKRD36C* (56.3%), *KIAA2018* (56.3%), *MUC4* (56.3%), *TMBIM4* (50.0%), and *RNF145* (50.0%) among all 16 patients (**Figure 3A**). We identified seven genes whose mutation rates were significantly higher in the resistant group than in the sensitive group: *SP8*, *HNRNPCL1*, *FRG1*, *RBM25*, *MUC16*, *ASTE1*, and *TMBIM4* (Fisher's exact test, $p < 0.05$; **Figure 3B**). The mutation rates of *SP8*, *HNRNPCL1*, and *FRG1* were 57.1% (4/7) in the sensitive group and 0% in the resistant group. The mutation rates of *ASTE1*, *MUC16*, and *RBM25* were 71.4% (5/7) in the sensitive group and 11.1% (1/9) in the resistant group.

All mutated genes were screened for a possible link to signaling pathways associated with platinum metabolism curated from the literature, such as cell cycle dependence, bicyclic platinum molecule activation, DNA damage repair, tumor cell apoptosis, drug transmembrane transport, platinum metabolizing drugs, DNA homeostasis disorders, and potential secondary drug resistance. The result showed that the mutation rates of these genes were similar in two groups, except for *MUC16* (**Supplementary Figure 2**).

Copy Number Variations

To further explore events that are related to dicycloplatin treatment, we analyzed CNVs harbored by the patients. *LAMTOR5-AS1*, *JUND,* and *SCAND1* were the most common genes with CNV deletion, and the number of CNV deleted genes was greater than the number of CNV-amplified genes (**Figure 3C**). The variation rates of CNV of four genes (*CTAGE4*, *GAGE2E*, *GAGE2C*, and *HORMAD1*) were higher in the resistant group than in the sensitive group (Fisher's exact test, $p < 0.05$). In contrast, the variation rates of CNV of six genes (*VARS, TMSB4Y*, *MUC22*, *DPCR1*, *CDSN*, and *HISTCH2AC*) were lower in the resistant group than in the sensitive group (Fisher's exact test, $p < 0.05$; **Figure 3D**).

Mutational Spectrum and Mutational Signatures

A scan of the observed mutations in the patients showed that C>T was the most common substitution in the cfDNA samples (**Figure 4A**). T>G substitution was the most common of the six base substitutions in the sensitive group. Its proportion decreased in the resistant group, but there was no statistical difference (Mann–Whitney *U* test, $p > 0.05$). The COSMIC mutational signatures of the patients are presented in **Figure 4B**. Signatures 3 and 1 were dominant in all samples with median percentages at 40.1 and 8.1%, respectively. The median percentages of the other mutational signatures ranged from 0 to 2.1%. Signature 12 had a higher proportion in the resistant group than in the sensitive group (Mann–Whitney *U* test, $p = 0.018$), while the other signatures did not show statistical differences between these two groups (Mann–Whitney *U* test, $p > 0.05$).

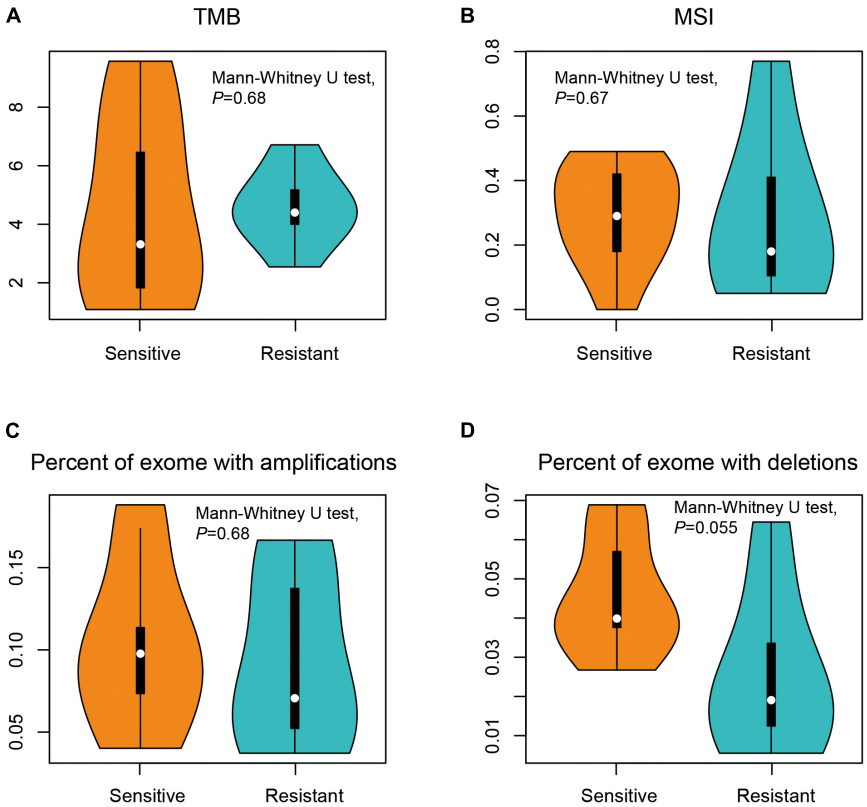


FIGURE 2 | Comparison of the genomic profiles between dicycloplatin-sensitive and dicycloplatin-resistant groups. **(A)** Tumor mutational burden (TMB). **(B)** Microsatellite instability (MSI) score. **(C)** Percentage of exons with amplification (defined as copy number ≥ 2 × average ploidy). **(D)** Percentage of exons with deletion (defined as copy number ≤ 0.5 × average ploidy).

Performance of Gene Classifiers in Predicting Resistance or Sensitivity to Dicycloplatin

There were differences in the frequency of multiple gene mutations and CNV variants between the resistant group and the sensitive group. The mutation frequency of seven genes (*SP8*, *HNRNPCL1*, *FRG1*, *RBM25*, *MUC16*, *ASTE1*, and *TMBIM4*) and the CNV variation frequency of four genes (*CTAGE4*, *GAGE2E*, *GAGE2C*, and *HORMAD1*) were higher in the resistant group than in the sensitive group. In order to search for markers of dicycloplatin treatment effects, we first calculated the sensitivity and specificity of these genes individually in assessing dicycloplatin resistance (**Table 2**, Nos. 1–11). Individually, none of these genes can satisfy a high sensitivity and high specificity at the same time. The sensitivity of three mutated genes (*SP8*, *HNRNPCL1*, and *FRG1*) and two CNV genes (*GAGE2E* and *GAGE2C*) was 57.1%, and their specificity was 100.0%. The sensitivity and specificity of three other mutated genes (*RBM25*, *MUC16*, and *ASTE1*) were 71.4 and 88.9%, respectively. Of the 11 genes tested, *TMBIM4* had the highest sensitivity (85.7%), but its specificity was the lowest (77.8%). The sensitivity of the remaining two copy number variated genes (*CTAGE4* and *HORMAD1*) was 71.4%, and their specificities were 100 and 88.9%, respectively.

The sensitivity and specificity of the CNV of six genes (*VARS*, *TMSB4Y*, *MUC22*, *DPCR1*, *CDSN*, and *HISTCH2AC*) in assessing dicycloplatin susceptibility were also calculated and shown on **Table 2** (Nos. 12–17). The sensitivity of five copy number variated genes (*VARS*, *TMSB4Y*, *MUC22*, *DPCR1*, and *CDSN*) was 55.6%, and their specificity was 100%. The sensitivity and specificity of *HISTCH2AC* were 100 and 71.4%, respectively.

We next calculated sensitivity and specificity of genes in different combinations to assess dicycloplatin resistance and susceptibility. We limited the number of genes in the combination within 3. Among all the tested combinations, one configuration stood out. It consisted of mutations of two genes (either *SP8*/*HNRNPCL1* or *SP8*/*FRG1*) and CNV loss of *GAGE2C* and successfully detected resistant cases with 100% sensitivity and 100% specificity (**Table 3**, Nos. 1–14). Another configuration of CNV loss of *DPCR1* and *TMSB4Y* together could detect sensitive cases with 88.9% sensitivity and 100% specificity (**Table 3**, Nos. 15–18).

DISCUSSION

In this study, we found several genes with mutation or CNV to predict the resistance or susceptibility to dicycloplatin

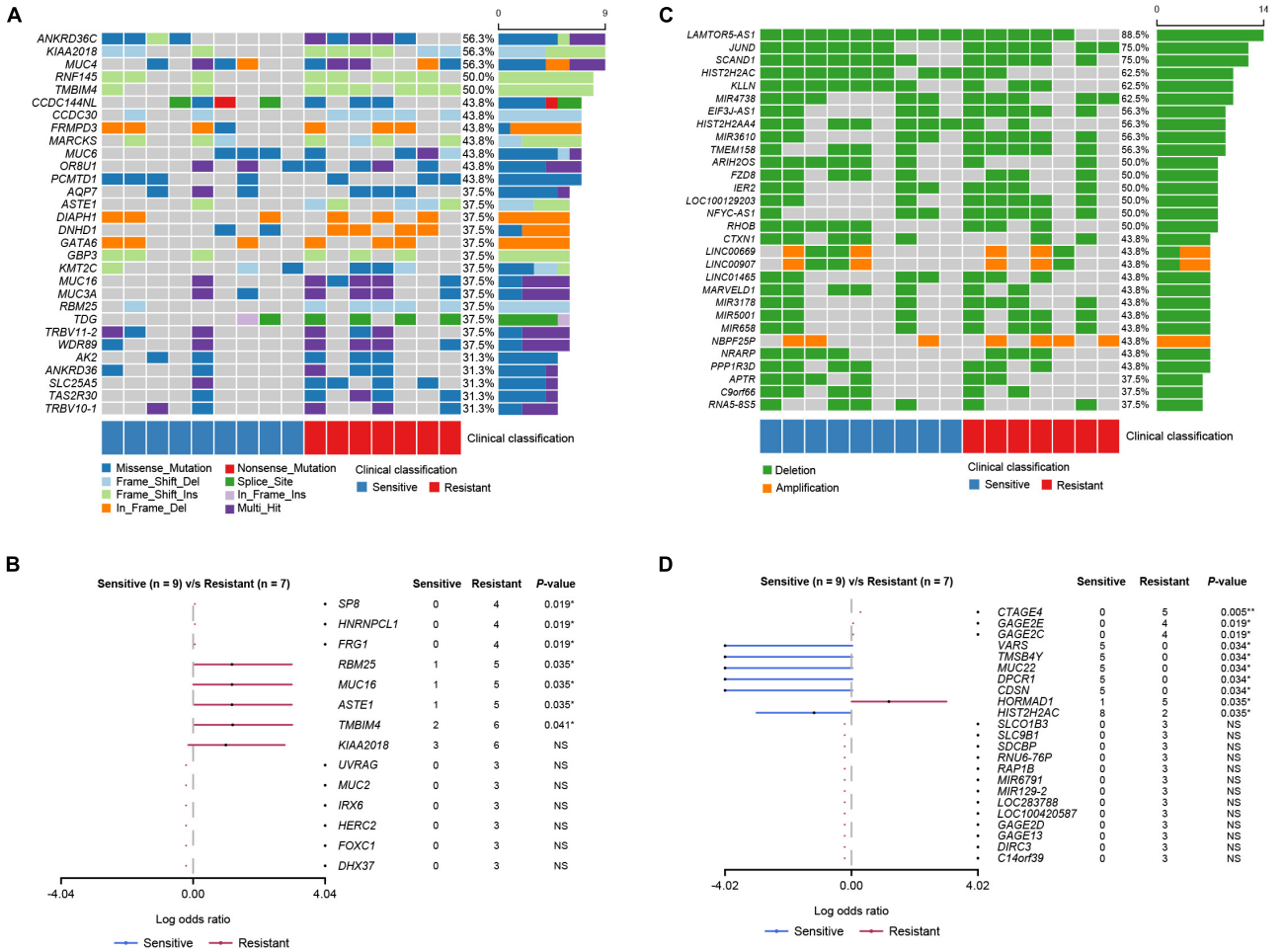


FIGURE 3 | Mutational landscape of the PCa patients. **(A)** The top 30 genes with SNV and Indel mutations. The percentage and horizontal bar on the right of each row indicate the fraction of patients with mutations in the corresponding genes and the composition of the types of mutations as color coded below the plot. **(B)** Somatic differentially mutated genes between the dicycloplatin-sensitive and dicycloplatin-resistant groups. The dots and horizontal bars denote the hazard rate and 5–95% confidence interval (CI), respectively. **(C)** The top 30 genes with CNV. The percentage and horizontal bar on the right of each row indicate the fraction of patients with CNV in the corresponding genes and composition of amplification (orange) or deletion (green). **(D)** Genes with differential CNV between the dicycloplatin-sensitive and dicycloplatin-resistant groups. The patients are grouped by their response to dicycloplatin. Blue: sensitive; red: resistant. The dots and horizontal bars denote the hazard rate and 5–95% CI in **(B,D)**. $^{**}p < 0.01$, $^{*}p < 0.05$. NS, not significant.

in PCa and evaluated the patients' response to a novel platinum drug. To our knowledge, this is the first predictive study in its category. Patients stratification by molecular subtyping would be helpful to improve the effectiveness of dicycloplatin in PCa.

Whole exome sequencing analysis in solid tumors is normally done with tissue samples. However, in the case of PCa, due to both the nature of the organ and surgical manipulation involved, tissue sample is often unavailable. Therefore, we extracted cfDNA from the blood of the PCa patients and used that in WES analysis instead of tumor tissue. Similar application of WES using cfDNA has been reported before and detected SNV, CNV, mutational signatures, TMB, and other genomic parameters in a reasonable accuracy, as verified by a positive correlation between the analysis results from cfDNA and tumor tissue samples (Bos et al., 2020).

In this study, all the patients were classified as stage IV, except one patient who was classified as stage II. This implies that the samples used in the current study were mostly from advanced tumors with high circulating tumor DNA (ctDNA) fractions. Therefore, the sensitivity of identified prediction factors in WES is expected to match to that from tissue samples. Our result indicates that cfDNA is a promising surrogate of tumor tissue because cfDNA can be obtained through a minimal-invasive procedure (blood drawing) and at the same time has the advantage to overcome tumor heterogeneity (Gonzalez-Billalabeitia et al., 2019). We expect that the surrogate strategy may even have expanded clinical uses beyond screening biomarkers for dicycloplatin response.

Among the potential predictive genes of dicycloplatin response that we identified, *MUC16* has been extensively studied

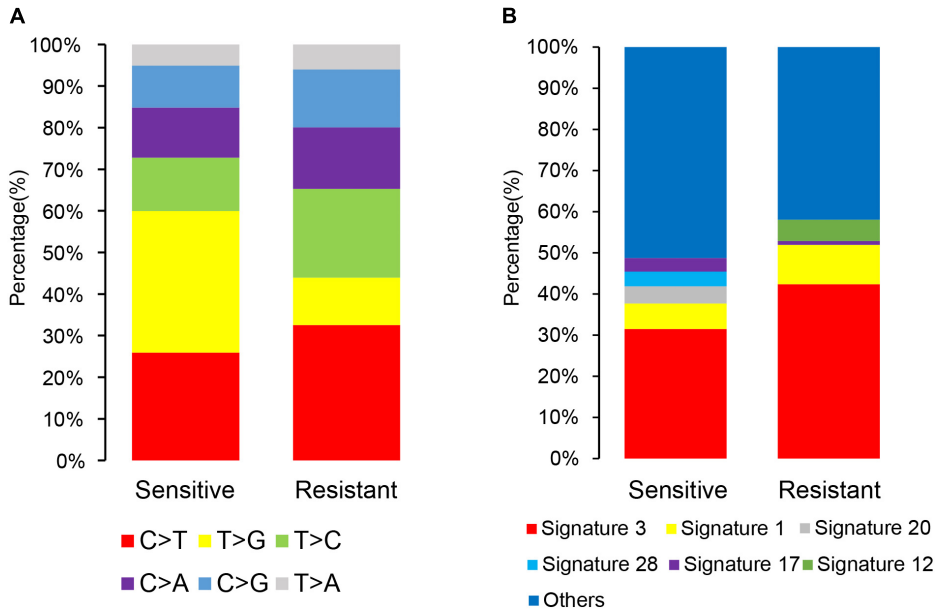


FIGURE 4 | Single-nucleotide and composite mutational signatures in dicycloplatin-sensitive and dicycloplatin-resistant groups. **(A)** Stacked bar graph of the percentages of six single-nucleotide substitutions in the sensitive (left) and resistance (right) groups. **(B)** Stacked bar graph of the percentages of COSMIC trinucleotide mutational signatures in the sensitive (left) and resistance (right) groups. The types of substitutions are color coded.

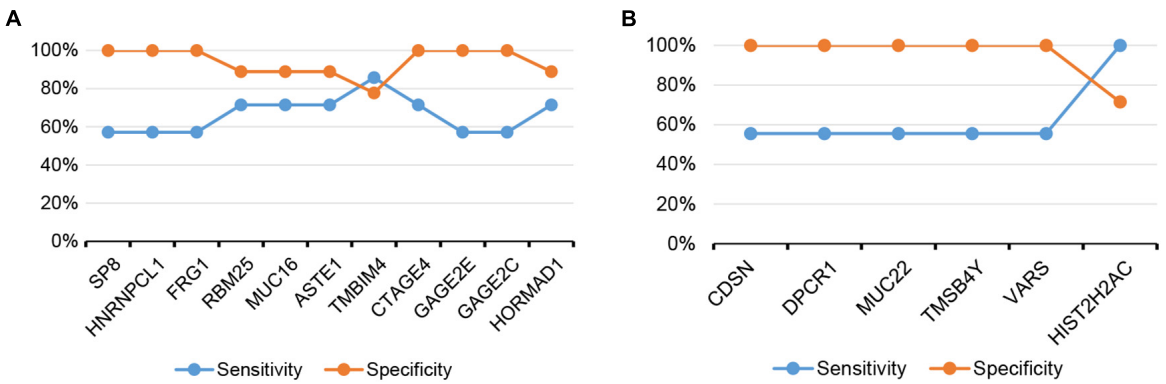


FIGURE 5 | The sensitivity and specificity of assessing dicycloplatin response using the mutation or CNV status of single gene. **(A)** Dicycloplatin resistance and **(B)** dicycloplatin susceptibility.

in several tumors. *MUC16* is known to promote the progression and metastasis of a variety of malignant tumors, and the abnormal expression of *MUC16* can lead to drug resistance to cytotoxic drugs and inhibition of apoptosis (Das et al., 2015). This is consistent with our findings that the mutation frequency of *MUC16* was higher in the dicycloplatin-resistant group. Most of the remaining genes that we identified have been found involved in tumorigenesis and disease progression. *SP8*, which encodes specificity protein 1/Klf-like zinc-finger transcription factor, inhibits KARS-mediated transformation and is also a tumor suppressor by itself (Fernandez-Zapico et al., 2011). The expression of SP8 was decreased in primary gastric cancer compared with normal gastric mucosa in a recent study (Chang et al., 2009). FRG1 expression was decreased in PCa tissues and that affected the migration and invasion of cancer cells (Tiwari et al., 2019). Deleterious mutations of FRG1, which had been identified in calcified pleura fibrous tumor and follicular thyroid cancer, were suggested to contribute to tumorigenesis (Erinjeri et al., 2018; Mehrad et al., 2018). Splicing regulator *RBM25* was identified as a tumor suppressor in acute myeloid leukemia, and the low level of RBM25 was associated with high MYC activity and poor prognosis of patients (Ge et al., 2019). In PCa, p53 regulates EMT by activating *RBM25*, thus promoting tumor progression and metastasis (Yang et al., 2019). However, the roles of the mentioned genes above in platinum metabolism and drug resistance remain unclear and need further investigation.

The sensitivity and specificity to predict dicycloplatin response by the individual actionable genes did not reach 100% (**Figure 5**).

TABLE 2 | The sensitivity and specificity of single gene in assessing dicycloplatin resistance (Nos. 1–11) and dicycloplatin susceptibility (Nos. 12–17).

No.	Genes	TP	FN	FP	TN	Sensitivity	Specificity	*p* value
1	*SP8*	4	3	0	9	57.1%	100.0%	0.019
2	*HNRNPCL1*	4	3	0	9	57.1%	100.0%	0.019
3	*FRG1*	4	3	0	9	57.1%	100.0%	0.019
4	*RBM25*	5	2	1	8	71.4%	88.9%	0.035
5	*MUC16*	5	2	1	8	71.4%	88.9%	0.035
6	*ASTE1*	5	2	1	8	71.4%	88.9%	0.035
7	*TMBIM4*	6	1	2	7	85.7%	77.8%	0.041
8	*CTAGE4*	5	2	0	9	71.4%	100.0%	0.005
9	*GAGE2E*	4	3	0	9	57.1%	100.0%	0.019
10	*GAGE2C*	4	3	0	9	57.1%	100.0%	0.019
11	*HORMAD1*	5	2	1	8	71.4%	88.9%	0.035
12	*VARS*	5	4	0	7	55.6%	100.0%	0.034
13	*TMSB4Y*	5	4	0	7	55.6%	100.0%	0.034
14	*MUC22*	5	4	0	7	55.6%	100.0%	0.034
15	*DPCR1*	5	4	0	7	55.6%	100.0%	0.034
16	*CDSN*	5	4	0	7	55.6%	100.0%	0.034
17	*HIST2H2AC*	8	0	2	5	100.0%	71.4%	0.035

TP, true positive; FN, false negative; FP, false positive; and TN, true negative.
The sensitivity in percentage is derived from the equation [TP/(TP + FN)]. The specificity in percentage is derived from the equation [TN/(TN + FP)]. p values are based on Fisher' exact test between the dicycloplatin-sensitive (nine patients) and dicycloplatin-resistant (seven patients) groups.

TABLE 3 | Classifiers of several genes in combination to assess dicycloplatin resistance (Nos. 1–14) and dicycloplatin susceptibility (Nos. 15–18).

No.	Genes	TP	FN	FP	TN	Sensitivity	Specificity	*P* value
1	*SP8/HNRNPCL1*	6	1	0	9	85.7%	100.0%	0.001
2	*SP8/FRG1*	6	1	0	9	85.7%	100.0%	0.001
3	*FRG1/HNRNPCL1*	6	1	0	9	85.7%	100.0%	0.001
4	*SP8/CTAGE4*	6	1	0	9	85.7%	100.0%	0.001
5	*HNRNPCL1/CTAGE4*	6	1	0	9	85.7%	100.0%	0.001
6	*SP8/GAGE2C*	6	1	0	9	85.7%	100.0%	0.001
7	*HNRNPCL1/ASTE1*	7	0	1	8	100.0%	88.9%	0.001
8	*FRG1/ASTE1*	7	0	1	8	100.0%	88.9%	0.001
9	*ASTE1/CTAGE4*	7	0	1	8	100.0%	88.9%	0.001
10	*ASTE1/HORMAD1*	7	0	1	8	100.0%	88.9%	0.001
11	*MUC16/GAGE2C*	7	0	1	8	100.0%	88.9%	0.001
12	*HNRNPCL1/GAGE2C*	7	0	1	8	100.0%	88.9%	0.001
13	*SP8/HNRNPCL1/GAGE2C*	7	0	0	9	100.0%	100.0%	0
14	*SP8/FRG1/GAGE2C*	7	0	0	9	100.0%	100.0%	0
15	*TMSB4Y/HIST2H2AC*	9	0	2	5	100.0%	71.4%	0.005
16	*CDSN/TMSB4Y*	7	2	0	7	77.8%	100.0%	0.003
17	*MUC22/TMSB4Y*	7	2	0	7	77.8%	100.0%	0.003
18	*DPCR1/TMSB4Y*	8	1	0	7	88.9%	100.0%	0.001

TP, true positive; FN, false negative; FP, false positive; and TN, true negative.
The sensitivity in percentage is derived from the equation [TP/(TP + FN)]. The specificity in percentage is derived from the equation [TN/(TN + FP)]. p values are based on Fisher' exact test between the dicycloplatin-sensitive (nine patients) and dicycloplatin-resistant (seven patients) groups.

From a clinical point of view, when the specificity of a screen is <100%, the patients who are false-positively identified can be tested by complementary methods and further cleared with their drug response status. However, if the sensitivity of a screen is <100%, the patients who are false-negatively identified will be missed in the screen. For oncologists and patients, the cost of suboptimal specificity is lower than the cost of suboptimal sensitivity. Therefore, it is preferable to evaluate patients' drug response with a high-sensitivity screen. One way to increase the sensitivity is through the combination of multiple marker genes. We test several configurations and found the combinations of either *SP8/HNRNPCL1/GAGE2C* or *SP8/FRG1/GAGE2C* reached 100% in both sensitivity and specificity (**Table 3**). Although the sample size of this study is

small, the putative drug response marker genes identified should provide a preliminary but critical assessment of the clinical value of dicycloplatin in PCa.

COSMIC mutational signature 25 had the higher proportion in the sensitive group than in the resistant group (**Supplementary Figure 3**). It was found highly represented in Hodgkin's lymphoma, although its etiology remains unknown and needs further study.

In addition to the molecular markers, we also found that patients who had prior endocrine therapy were more likely to develop resistance to dicycloplatin (**Table 1**). It is possible that endocrine therapy and dicycloplatin resistance are mechanistically linked, probably due to clonal selection. However, another possibility is that the endocrine-treated patients were already in a more advanced stage when they received dicycloplatin, therefore had a worse clinical performance overall.

In conclusion, our study identifies that mutation or CNV in several genes are putatively predictive to dicycloplatin response in PCa. However, further tests in cell and animal models are necessary to search for and verify the possible action mechanism of these genes in platin drug resistance. The prediction method that we postulated should be valuable to screen patients suitable for dicycloplatin treatment, therefore reducing the suffering of PCa patients who are predicted as not good responders of the therapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA727718.

ETHICS STATEMENT

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

ML, CL, and GZ carried out carried on the study design and contributed to sample preparation. ML, XZ, and JZ analyzed the clinical data and wrote manuscript. JL and SJ revised the manuscript and approved final manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.669605/full#supplementary-material

Supplementary Figure 1 | Dynamics of PSA and free PSA (fPSA) of patients during treatment course. PSA **(A)** and fPSA **(B)** levels were decreased in 6 PR patients and 3 CR patients, were increased in 6 PD patients and 1 SD patient.

Supplementary Figure 2 | Mutational landscape of selected genes associated with tumor and platinum metabolism. **(A)** Oncoplot of the top 20 mutated gene in dicycloplatin-treated samples. The percentages and the bar plots on the right represent the fraction and the number of samples with variations in the corresponding gene. **(B)** Differentially somatic mutated genes between dicycloplatin-sensitive group and dicycloplatin-resistant group. The dots and horizontal bars denote the hazard rate and 5–95% CI. $*P$-value < 0.05. NS: not significant.

Supplementary Figure 3 | Comparison of 30 cosmic signatures between dicycloplatin-sensitive and dicycloplatin-resistant patients.

REFERENCES

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., et al. (2015). Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407. doi: 10.1038/ng.3441

Apps, M. G., Choi, E. H., and Wheate, N. J. (2015). The state-of-play and future of platinum drugs. *Endocr. Relat. Cancer* 22, R219–R233.

Bos, M. K., Angus, L., Nasserinejad, K., Jager, A., Jansen, M. P. H. M., Martens, J. W. M., et al. (2020). Whole exome sequencing of cell-free DNA – A systematic review and Bayesian individual patient data meta-analysis. *Cancer Treat. Rev.* 83:101951. doi: 10.1016/j.ctrv.2019.101951

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 9:34.

Chang, W., Ma, L., Lin, L., Gu, L., Liu, X., Cai, H., et al. (2009). Identification of novel hub genes associated with liver metastasis of gastric cancer. *Int. J. Cancer* 125, 2844–2853. doi: 10.1002/ijc.24699

Das, S., Rachagani, S., Torres-Gonzalez, M. P., Lakshmanan, I., Majhi, P. D., Smith, L. M., et al. (2015). Carboxyl-terminal domain of MUC16 imparts tumorigenic and metastatic functions through nuclear translocation of JAK2 to pancreatic cancer cells. *Oncotarget* 6, 5772–5787. doi: 10.18632/oncotarget.3308

Erinjeri, N. J., Nicolson, N. G., Deyholos, C., Korah, R., and Carling, T. (2018). Whole-exome sequencing identifies two discrete druggable signaling pathways in follicular thyroid cancer. *J. Am. Coll. Surg.* 226, 950–959 e955.

Fernandez-Zapico, M. E., Lomberk, G. A., Tsuji, S., DeMars, C. J., Bardsley, M. R., Lin, Y. H., et al. (2011). A functional family-wide screening of SP/KLF proteins identifies a subset of suppressors of KRAS-mediated cell growth. *Biochem. J.* 435, 529–537. doi: 10.1042/bj20100773

Fortin, S., Brasseur, K., Morin, N., Asselin, É, and Bérubé, G. (2013). New platinum(II) complexes conjugated at position 7alpha of 17beta-acetyl-testosterone as new combi-molecules against prostate cancer: design, synthesis, structure-activity relationships and biological evaluation. *Eur. J. Med. Chem.* 68, 433–443. doi: 10.1016/j.ejmech.2013.08.011

Ge, Y., Schuster, M. B., Pundhir, S., Rapin, N., Bagger, F. O., Sidiropoulos, N., et al. (2019). The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nat. Commun.* 10:172.

Gonzalez-Billalabeitia, E., Conteduca, V., Wetterskog, D., Jayaram, A., and Attard, G. (2019). Circulating tumor DNA in advanced prostate cancer: transitioning from discovery to a clinically implemented test. *Prostate Cancer Prostatic. Dis.* 22, 195–205. doi: 10.1038/s41391-018-0098-x

Kirby, M., Hirst, C., and Crawford, E. D. (2011). Characterising the castration-resistant prostate cancer population: a systematic review. *Int. J. Clin. Pract.* 65, 1180–1192. doi: 10.1111/j.1742-1241.2011.02799.x

Liu, K. J., Guan, Z. Z., Liang, Y., Yang, X. Q., Peng, J., Huang, H., et al. (2014). A double-blind, randomized phase II study of dicycloplatin plus paclitaxel versus carboplatin plus paclitaxel as first-line therapy for patients with advanced non-small-cell lung cancers. *Arch. Med. Sci.* 10, 717–724. doi: 10.5114/aoms.2014.44862

Lokich, J., and Anderson, N. (1998). Carboplatin versus cisplatin in solid tumors: an analysis of the literature. *Ann. Oncol.* 9, 13–21. doi: 10.1023/a:1008215213739

Maitra, A., Biswas, N. K., Amin, K., Kowtal, P., Kumar, S., Das, S., et al. (2013). Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat. Commun.* 4:2873. doi: 10.1038/ncomms3873

Mehrad, M., LaFramboise, W. A., Lyons, M. A., Trejo Bittar, H. E., and Yousem, S. A. (2018). Whole-exome sequencing identifies unique mutations and copy number losses in calcifying fibrous tumor of the pleura: report of 3 cases and review of the literature. *Hum. Pathol.* 78, 36–43. doi: 10.1016/j.humpath.2018.04.005

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54.

Ozols, R. F., Bundy, B. N., Greer, B. E., Fowler, J. M., Clarke-Pearson, D., Burger, R. A., et al. (2003). Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. *J. Clin. Oncol.* 21, 3194–3200. doi: 10.1200/jco.2003.02.153

Rossi, A., Di Maio, M., Chiodini, P., Rudd, R. M., Okamoto, H., Skarlos, D. V., et al. (2012). Carboplatin- or cisplatin-based chemotherapy in first-line treatment of small-cell lung cancer: the COCIS meta-analysis of individual patient data. *J. Clin. Oncol.* 30, 1692–1698. doi: 10.1200/jco.2011.40.4905

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30.

Tirkes, T., Hollar, M. A., Tann, M., Kohli, M. D., Akisik, F., and Sandrasegaran, K. (2013). Response criteria in oncologic imaging: review of traditional and new criteria. *Radiographics* 33, 1323–1341. doi: 10.1148/rg.335125214

Tiwari, A., Mukherjee, B., Hassan, M. K., Pattanaik, N., Jaiswal, A. M., and Dixit, M. (2019). Reduced FRG1 expression promotes prostate cancer progression and affects prostate cancer cell migration and invasion. *BMC Cancer* 19:346. doi: 10.1186/s12885-019-5509-4

Wang, C., and Yu, T. (2004). Cytotoxicity and mice hematological toxicology of cisplatin, carboplatin and dicycloplatin. *Med. J. Liaoning* 18, 273–274.

Yang, X., Jin, X., Song, Q., Tang, K., Yang, Z. Y., Zhang, X. F., et al. (2010). Structural studies of dicycloplatin, an antitumor supramolecule. *Sci. China Chem.* 53, 1346–1351. doi: 10.1007/s11426-010-3184-z

Yang, Z., Qu, C. B., Zhang, Y., Zhang, W. F., Wang, D. D., Gao, C. C., et al. (2019). Dysregulation of p53-RBM25-mediated circAMOTL1L biogenesis contributes to prostate cancer progression through the circAMOTL1L-miR-193a-5p-Pcdha pathway. *Oncogene* 38, 2516–2532. doi: 10.1038/s41388-018-0602-8

Yu, J. J., Yang, X., Song, Q., Mueller, M. D., and Remick, S. C. (2014). Dicycloplatin, a novel platinum analog in chemotherapy: synthesis of chinese pre-clinical and clinical profile and emerging mechanistic studies. *Anticancer Res.* 34, 455–463.

Zang, Y. S., Dai, C., Xu, X., Cai, X., Wang, G., Wei, J., et al. (2019). Comprehensive analysis of potential immunotherapy genomic biomarkers in 1000 Chinese patients with cancer. *Cancer Med.* 8, 4699–4708. doi: 10.1002/cam4.2381

Conflict of Interest: XZ, JL, and CL were employed by company GenomiCare Biotechnology Co., Ltd., Shanghai, China.

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 September 2021
doi: 10.3389/fonc.2021.725988

Gene Expression Profiling for Differential Diagnosis of Liver Metastases: A Multicenter, Retrospective Cohort Study

Qifeng Wang[1,2,3,4†], Fen Li[5†], Qingming Jiang[6†], Yifeng Sun[7†], Qiong Liao[4,8], Huimin An[4,9], Yunzhu Li[8], Zhenyu Li[6], Lifang Fan[10], Fang Guo[10], Qinghua Xu[4,7,11,12], Yixin Wo[7], Wanli Ren[7], Junqiu Yue[4,10], Bin Meng[4,13*], Weiping Liu[14*] and Xiaoyan Zhou[1,2,3,4*]*

[1] Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China, [2] Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, [3] Institute of Pathology, Fudan University, Shanghai, China, [4] The Cancer of Unknown Primary Group of Pathology Committee, Chinese Research Hospital Association, Shanghai, China, [5] Department of Pathology, Chengdu Second People's Hospital, Chengdu, China, [6] Department of Pathology, Chongqing University Cancer Hospital, Chongqing, China, [7] The Canhelp Genomics Research Center, Canhelp Genomics Co., Ltd., Hangzhou, China, [8] Department of Pathology, Sichuan Cancer Hospital, Chengdu, China, [9] Department of Pathology, Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University, Hangzhou, China, [10] Department of Pathology, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [11] The Institute of Machine Learning and Systems Biology, College of Electronics and Information Engineering, Tongji University, Shanghai, China, [12] Xuzhou Engineering Research Center of Medical Genetics and Transformation, Department of Genetics, Xuzhou Medical University, Xuzhou, China, [13] Department of Pathology, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [14] Department of Pathology, West China Hospital, Sichuan University, Chengdu, China

Background: Liver metastases (LM) are the most common tumors encountered in the liver and continue to be a significant cause of morbidity and mortality. Identification of the primary tumor of any LM is crucial for the implementation of effective and tailored treatment approaches, which still represents a difficult problem in clinical practice.

Methods: The resection or biopsy specimens and associated clinicopathologic data were archived from seven independent centers between January 2017 and December 2020. The primary tumor sites of liver tumors were verified through evaluation of available medical records, pathological and imaging information. The performance of a 90-gene expression assay for the determination of the site of tumor origin was assessed.

Result: A total of 130 LM covering 15 tumor types and 16 primary liver tumor specimens that met all quality control criteria were analyzed by the 90-gene expression assay. Among 130 LM cases, tumors were most frequently located in the colorectum, ovary and breast. Overall, the analysis of the 90-gene signature showed 93.1% and 100% agreement rates with the reference diagnosis in LM and primary liver tumor, respectively. For the common primary tumor types, the concordance rate was 100%, 95.7%, 100%, 93.8%, 87.5% for classifying the LM from the ovary, colorectum, breast, neuroendocrine, and pancreas, respectively.

OPEN ACCESS

Edited by:
Xiao Zhu, Guangdong Medical University, China

Reviewed by:
Paolo Cremaschi, Human Technopole, Italy
Gisela Ceballos, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico

***Correspondence:**
Xiaoyan Zhou
13524324387@163.com
Weiping Liu
liuweiping2001@vip.sina.com
Bin Meng
mbincn@163.com
Junqiu Yue
yuejunqiu@hotmail.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: *16 June 2021*
Accepted: *18 August 2021*
Published: *22 September 2021*

Citation:
Wang Q, Li F, Jiang Q, Sun Y, Liao Q, An H, Li Y, Li Z, Fan L, Guo F, Xu Q, Wo Y, Ren W, Yue J, Meng B, Liu W and Zhou X (2021) Gene Expression Profiling for Differential Diagnosis of Liver Metastases: A Multicenter, Retrospective Cohort Study. Front. Oncol. 11:725988. doi: 10.3389/fonc.2021.725988

Conclusion: The overall accuracy of 93.8% demonstrates encouraging performance of the 90-gene expression assay in identifying the primary sites of liver tumors. Future incorporation of the 90-gene expression assay in clinical diagnosis will aid oncologists in applying precise treatments, leading to improved care and outcomes for LM patients.

Keywords: liver metastasis, tissue of origin, gene expression profiling, real-time PCR, tumor classification

INTRODUCTION

Liver metastases (LM) are tumors that have propagated to the liver from tumors originating from other parts of the body. Due to the venous blood returning from the gastrointestinal system through portal vein circulation, gastrointestinal tract tumors are more likely to metastasize to the liver (1). Besides, the liver microenvironment also plays a significant role in the development of hepatic metastasis. Numerous studies have shown that both the acellular such as extracellular matrix proteins (i.e. collagen) and the cellular components of the liver such as Kupffer cells, hepatic stellate cells and liver sinusoidal endothelial cells contribute to the metastatic ability of tumors of different origins (2). According to the statistical data in the Surveillance, Epidemiology and End Results (SEER) database, 5.14% to 6.46% of cancer patients are diagnosed with synchronal LM at the time of primary cancer diagnosis (3, 4). Of note, during the course of the cancer disease, up to 50% of patients with various tumor types will either present with or develop LM (1). The most common tumor that spreads to the liver is breast cancer for younger women and colorectal cancers for younger men (3). In the current era, several studies investigated that the incidence rate of cancer of unknown primary (CUP) is currently decreasing and reaches 1-2% (5). Liver CUP is the most common CUP subgroup (30–40%) and has the most dismal prognosis with median overall survival (OS) of 1–2 months and one-year OS of 5–12% (6).

The prognosis of LM varies to tumor types. LM originated from small intestine cancer shows the best prognosis, followed by testis cancer and breast cancer (4). Traditionally, the treatment approaches were established according to the primary tumor of LM. For example, resection can be usually performed in patients with colorectal liver metastases (CLM) and neuroendocrine tumor liver metastases (NETLM), but it may be not appropriate for patients with LM from pancreatic cancer, esophageal cancer, melanoma and adrenocortical cancer (1). In addition, different tumor types carry specific genetic alterations, genomic feature analysis which could provide precise and pertinent clinical details for disease management. For CLM patients, information on the mutation status of oncogenes such as *BRAF, NRAS*, and *KRAS* as well as analysis of microsatellite instability (MSI) status have led to precise therapy and prognostic stratification (7). Therefore, identification of the primary tumor of any LM is pivotal for the implementation of valid and tailored treatment options, which still acts as a troublesome problem in the clinical setting. In most cases, metastatic tumors with representative histological features similar to the primary lesion can be correctly distinguished with hematoxylin-eosin (H&E) staining and immunohistochemistry (IHC) (8). However, the distinction between intrahepatic cholangiocarcinoma and metastatic adenocarcinoma is frequently challenging owing to the overlapping phenotypic profiles (9).

Over the last decades, molecular profiling has been under speed development for predicting tumor site of origins in CUP patients (10, 11). According to the tumor origin, specific gene expression profiling has been well recognized in most tumor types, which reflects the different expression profilings in their normal tissues of origin. Differences in gene expression pattern thus allow distinction between various solid tumors and provide a valuable method for diagnosis of the tissue of origin in CUP patients. Recently, our group has developed a 90-gene expression assay for the classification of 21 common tumor types which represent approximately 95% of the incident solid tumors that are known to produce distant metastases. In a retrospective cohort of 609 clinical specimens, the 90-gene expression assay illustrated an overall accuracy of 90.4% for primary tumors and 89.2% for metastatic tumors. Furthermore, in a real-world cohort of 141 CUP patients, the gene expression assay was able to provide instructive predictions of primary tumors in 71.6% of patients (101 of 141). These findings suggest that the 90-gene expression assay could efficiently identify the primary site for a broad spectrum of tumor types and support its diagnostic utility of molecular classification in difficult-to-diagnose metastatic tumors (12). Recently, Wang et al. performed the 90-gene expression assay for the differential diagnosis of metastatic triple-negative breast cancer (TNBC) (13). This assay correctly identified 97.6% of TNBC lymph node metastases (41 of 42) and 96.8% of distant metastatic tumors (30 of 31). Zheng et al. investigated the potential utility of the 90-gene expression assay in diagnosing the tumor origin of brain tumors (14). The molecular assay illustrated 100% accuracy for discriminating primary brain tumors from brain metastases and accurately predicted primary sites for 89% of brain metastases (39 of 44).

In the present study, we conducted a multi-center retrospective study based on seven cancer centers in China to assess the performance of the 90-gene expression assay and explore its potential diagnostic utility for LM.

MATERIAL AND METHODS

Patient Enrollment and Specimen Acquisition

The study protocol was approved by the institutional review board of Fudan University Shanghai Cancer Center (FUSCC, Shanghai, China), West China Hospital Sichuan University (WCHSU, Chengdu,

Sichuan, China), Sichuan Cancer Hospital (SCH, Chengdu, Sichuan, China), Chongqing Cancer Hospital (CCH, Chongqing, China), Tianjin Medical University Cancer Institute & Hospital (TMUCIH, Tianjin, China), Sir Run Run Shaw Hospital (SRRSH, Hangzhou, Zhejiang, China) and Hubei Cancer Hospital (HCH, Wuhan, Hubei, China). Between January 2017 and December 2020, a total of 156 surgical or biopsy specimens from the liver and associated clinicopathologic data were archived from seven independent centers, and 146 cases (130 LM and 16 primary liver tumors) that met all criteria were enrolled in the present study. Biopsy samples were obtained by needle core biopsy (NCB) or fine-needle aspiration (FNA), using either transabdominal ultrasound or computed tomography (CT) guidance. The inclusion criteria for all specimens were the following: (1) formalin-fixed, paraffin-embedded (FFPE) tumor tissues, (2) the primary tumor sites were verified through evaluation of available medical records, pathological and imaging information. The reference diagnosis of primary tumor was conform to the 21 tumor types of the assay (**Supplementary Table 1**) (12), (3) at least 60% tumor cell content, and (4) less than 40% necrosis based on the H&E staining evaluation.

Sample Preparation and RNA Isolation

Five to fifteen 5μm unstained sections were freshly cut for gene expression analysis. The FFPE tissue samples were centralized and the H&E-stained slide of each case had been reviewed for evaluation of the percentage of tumor cells and necrotic areas by two senior pathologists from FUSCC (QF W and XY Z). The regions of tumor tissue were marked on the H&E-stained slides and macro-dissected manually for tumor cells enrichment. Total RNA isolation and gene expression profiling were performed at the Canhelp Genomic Reference Laboratory (Hangzhou, China). Total RNA was extracted using a FFPE Total RNA Isolation Kit (Canhelp Genomics Co., Ltd., Hangzhou, China) according to the protocols. Briefly, FFPE tissue was deparaffinized, followed by digestion, DNase treatment and total RNA elution. The concentration of total RNA was measured by spectrophotometer at 260-nm absorbance, and the purity was quantified using A260/A280 ratio. RNA samples with A260/A280 ratios between 1.7 and 2.1 were enrolled in this study.

Gene Expression Profiling Analysis

After performing reverse transcription on isolated total RNA (2ug per specimen) using a High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems, Foster City, CA), the 90-gene real-time PCR (RT-PCR) assay (Canhelp Genomics, Hangzhou, China). was analyzed with a 7500 Real-Time PCR system (Applied Biosystems) to analyze cancer-specific gene expression profiles. The RT-PCR program was initiated at 95°C for 10 minutes, followed by 40 cycles at 95° C for 15 seconds and 60°C for 1 minute. To correct for input variation, for each sample, cycle threshold (Ct) measurements of target genes were normalized to multiple reference genes. For samples with the Ct values of reference genes greater than 38 were excluded. The 90-genes expression data of valid samples were provided in **Supplementary Table 2**.

90-Gene Classifier for Tumor Classification: Algorithm Development and Data Analysis

Initially, the cancer-specific gene markers were identified based on a pan-cancer transcriptome database comprising 5434 specimens representing 21 tumor types (15). The database included both primary and metastatic tumors and well-differentiated to undifferentiated tumors. The SVM-RFE (Support Vector Machine-Recursive Feature Elimination) machine learning algorithm was used to select the Top-10 most predictive genes for each of the 21 tumor types. After removing redundant genes, a list of 90 genes specific to 21 tumor types was identified. Details of the 90-gene list were provided in **Supplementary Table 3**. Then, an SVM linear model was trained using the whole pan-cancer transcriptome database to form a multiclass classification algorithm ("90-gene classifier").

Mathematically, the 90-gene classifier creates a hyperplane for each tumor type in a 90-dimensional space. For an unknown test sample, the algorithm calculates its 90 genes' expression values, projects it to the 90-dimensional space, and estimates the distance of the test sample to each of 21 hyperplanes. The position of the test sample relative to the hyperplane determines its membership in one or the other class (e.g., "breast cancer" *vs.* "not breast cancer"). Furthermore, the confidence of the test sample belongs to a tumor type is proportional to the distance of the test sample from the corresponding hyperplane. The far the distance, the higher the confidence. Then, the distances of the test sample from each of the 21 hyperplanes were compared and transformed to the similarity scores with the Platt Scaling formula (16). Intuitively, the similarity scores reflect how much the gene expression pattern of the test sample is similar to the global gene expression pattern of the indicated tumor type. The similarity scores were probability-based, with a reported range from 0 to 100, and all 21 scores sum to 100. The tumor type with the highest similarity score was defined as the predicted tumor type by the 90-gene classifier. An example was shown in **Supplementary Figure 1**. The primary site with the highest similarity score is gastroesophagus, thus indicating the most likely tissue of origin is gastroesophagus.

Statistical Analysis

All statistical analysis was performed using the R software (version 3.6.1) and packages from the Bioconductor project (version 3.9). The hierarchical clustering of clinical specimens based on the gene expression pattern was performed using "pheatmap" package (version 1.0.12). The average linkage hierarchical clustering method was performed where the metric of similarity was Pearson's correlation between every pair of samples. The receiver operating characteristic (ROC) curves were estimated using "multiROC" package (version 1.1.1). The gene expression assay performance was assessed by calculating the area under curve (AUC) for each tumor type and aggregation across all tumor types. For multi-class evaluation, the AUC for all tumor types was calculated through a micro-averaging approach, which stacked all tumor

types together, thus converting the multiclass classification into binary classification. The micro-averaging approach further considered the contributions of different tumor types and weight metrics toward the largest type when some tumor types have more instances than others. P-value was computed two-sided and considered as statistically significant if p-value < 0.05.

RESULTS

Patients and Samples

Initially, 156 FFPE specimens covering 16 primary tumor types were collected in the present study, 148 had successful histologic quality control and 146 of these samples passed the RT-PCR quality control. More specifically, seven specimens were excluded because of less than 60% tumor cell content, one because of more than 40% necrosis, and two because of RT-PCR quality control failures. Finally, 130 LM and 16 primary liver tumor specimens met all quality control criteria and were successfully analyzed by the 90-gene expression assay. The overall study design is presented in **Figure 1**. The sample enrollment of seven center hospitals was shown in **Supplementary Table 4**.

The demographics and clinical characteristics of the cohort are provided in **Table 1**. The cohort included 63 males and 83 females, with a median age of 57.5 years old (range 14-83). All specimens were taken from the liver, of which 36 were biopsy samples and 110 were resection samples. The origin of LM came from 15 primary sites and the most common tissue of origin were colorectum (n=23) and ovary (n=23), followed by breast (n=19), neuroendocrine (n=16), pancreas (n=16) and gastroesophagus (n=10). Other relatively rare tumor types comprising melanoma (n=4), cervix (n=4), lung (n=3), adrenal (n=3), germ cell (n=2), head&neck (n=2), sarcoma (n=2), kidney (n=2) and urinary (n=1) were also included. Among 146 samples, the degree of differentiation of 97 cases was defined, 26 (26.8%) cases were well-differentiated, whereas 71 (73.2%) cases were poorly differentiated.

Performance of the 90-Gene Expression Assay in Liver Tumors

For primary liver tumors, the 90-gene expression assay correctly classified all 16 samples showing a 100% accuracy. For 130 LM

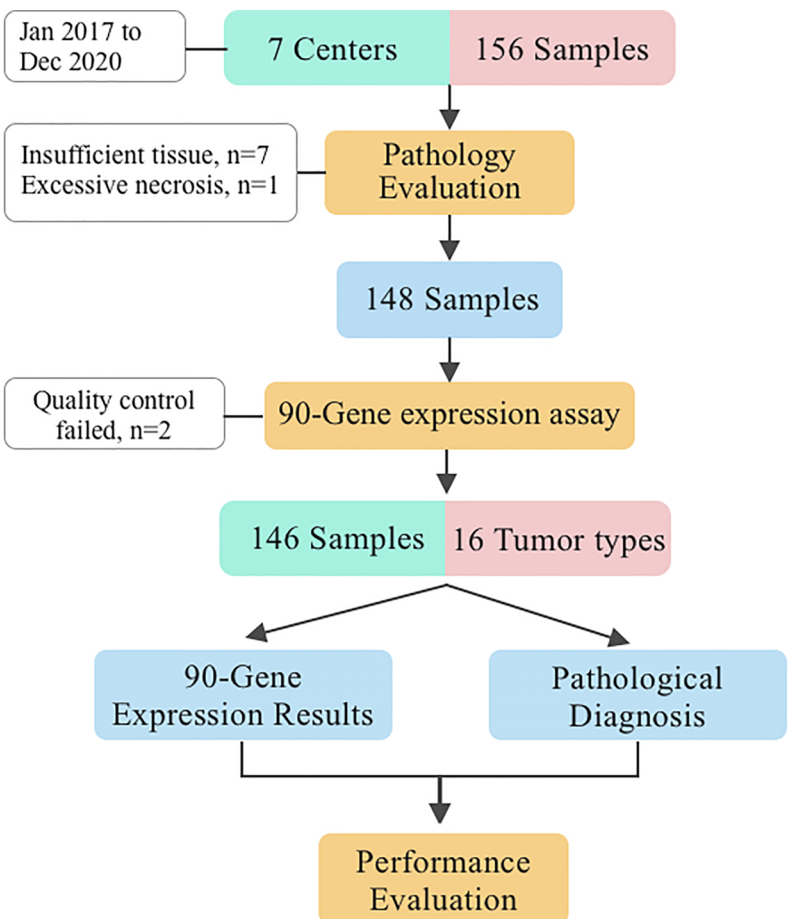


FIGURE 1 | Study design.

TABLE 1 | The demographics and clinical characteristics of the cohort.

Characteristic	Number of specimens (N = 146)	Percentage (%)
Gender		
Male	63	43.2
Female	83	56.8
Age		
Median	57.5	
Range	14-83	
Tumor types		
Liver	16	11.0
Colorectum	23	15.8
Ovary	23	15.8
Breast	19	13.0
Neuroendocrine	16	11.0
Pancreas	16	11.0
Gastroesophagus	10	6.8
Melanoma	4	2.7
Cervix	4	2.7
Lung	3	2.1
Adrenal	3	2.1
Germ cell	2	1.4
Head&neck	2	1.4
Sarcoma	2	1.4
Kidney	2	1.4
Urinary	1	0.7
Histological Subtype		
Adenocarcinoma	114	78.1
Neuroendocrine	16	11.0
Squamous cell carcinoma	10	6.8
Melanoma	4	2.7
Sarcoma	2	1.4
Degree of differentiation[1]		
Well-differentiated	26	26.8
Poorly differentiated	71	73.2

[1]*The degree of differentiation of 49 specimens was undefined.*

cases, the 90-gene expression assay achieved a 93.1% (121/130, 95% CI: 0.87-0.97) accuracy by comparing the predicted tumor types with the reference diagnosis. The AUC of the Top-5 common tumor types ranged from 0.945 to 1 (**Figures 2A–E**), and the weighted AUC for all tumor types reached 0.981 (**Figure 2F**). As shown in **Table 2**, the sensitivities of the 90-gene expression assay are variable, ranging from 50% (head&neck) to 100% (ovary, breast, melanoma, etc.).

Of the 146 specimens, 26 were well or moderately differentiated tumors, 71 were poorly or undifferentiated tumors, and 49 were not specified. More specifically, the classification accuracy was 96.2% (25 of 26) for well or moderately differentiated tumors and 88.7% (63 of 71) for poorly or undifferentiated tumors, with no statistically notable difference ($p = 0.47$). In addition, the present study enrolled 36 biopsy specimens and 110 resection specimens. The overall accuracy of 90-gene expression assay showed no significant difference between biopsy and resection groups, (88.9% and 95.4%, respectively, p value equals 0.31).

In subgroup analysis, the neuroendocrine tumors were originated from pancreas (n=5), gallbladder (n=2), thyroid (n=2), ovary (n=1), esophagus (n=1), lung (n=1) and undefined (n=4), with an overall accuracy of 93.8% (15/16). For ten cases of squamous cell carcinoma, their origins were composed of the cervix (n=4), gastroesophagus (n=3), head&neck (n=2), and lung (n=1). The 90-gene signature correctly classified the tissue of origin in 7 of 10 cases (70%).

To illustrate the similarity between clinical samples, we performed hierarchical clustering based on primary liver cancer and six main metastatic tumor types (n >5). As shown in **Figure 3**, the samples were clustered into distinct groups that followed the tumor types based on the 90-gene expression pattern. The primary liver tumor samples were clustered together and showed distinct patterns from six LM types. Among LM types, digestive system neoplasms including colorectal, gastroesophageal, and pancreatic tumors were more likely to share similar gene expression patterns. For example, most of gastroesophageal tumors were clustered together, whereas few samples were similar to colorectal, pancreatic tumors.

A total of nine LM cases had discordant predictions compared with reference diagnoses. The histological types of nine misclassified samples included gastroesophageal (n=3), pancreas (n=2), lung (n=1), colorectum (n=1), neuroendocrine (n=1) and head&neck (n=1). Among nine cases, five were adenocarcinoma, three were squamous cell carcinoma and one was a neuroendocrine tumor. Eight of nine cases were poorly differentiated. The detailed characteristics of the discordant cases were investigated in **Table 3**.

DISCUSSION

LM is the most common tumors encountered in the liver and continues to be a notable factor for morbidity and mortality. The identification of the primary tumor in the conditions of any LM is critical to define optimal management. In clinics, imaging modalities such as ultrasonography, CT, Magnetic resonance imaging (MRI), and positron emission tomography (PET) scans are typically most often applied for LM diagnosis (17). Hui et al. developed B-mode ultrasound radiomic models to distinguish the origin of liver metastatic lesions from the digestive tract tumor, lung tumor and breast tumor, with the sensitivity ranging from 70% to 75% (18). Moreover, serum tumor markers can potentially aid in the diagnosis of patients with LM. For instance, carcinoembryonic antigen (CEA) is one of the most crucial tumor markers for colorectal cancer. Other useful biomarkers for LM diagnosis include CA 19-9 (pancreaticobiliary cancer), chromogranin A (neuroendocrine tumor), CA 15-3 (breast cancer) and CA-125 (germinal tumor) (19). Although these serum markers are indicative for certain primary tumors, their specificities are still limited (19). For instance, the increase of serum CEA level may indicate the presence of colorectal cancer, but it can be also observed in 30-60% of pancreatic cancer patients (20).

Histological examinations including morphological and IHC analyses are the gold standard for tumor origin diagnosis. However, most of the LM originated from adenocarcinoma, which shares overlapping histological features with primary liver tumors or between each other (1). Thus, additional

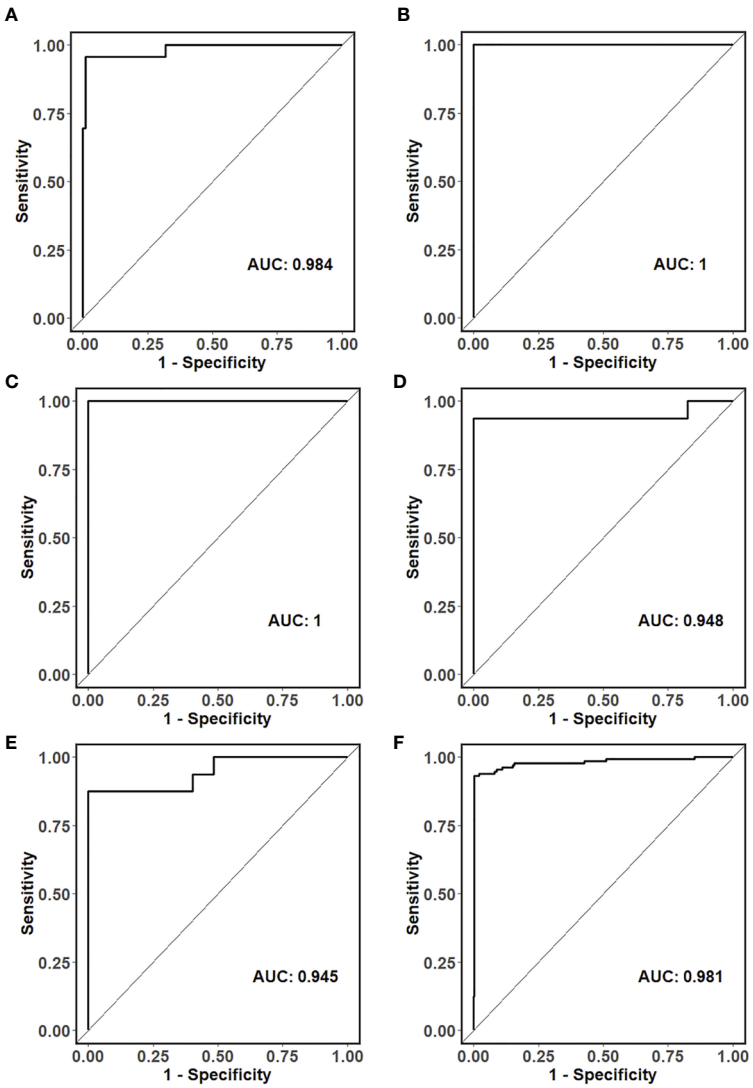


FIGURE 2 | ROC curves for the classification of the tissue of origin in **(A)** colorectal, **(B)** ovarian, **(C)** breast, **(D)** neuroendocrine, **(E)** pancreatic and **(F)** all liver metastatic tumors.

organ-specific IHC panels are crucial to characterize the tumor origin. A combination cytokeratin (CK) panel CK7/CK20 is recommended for initial evaluation (1, 8, 21). For example, CK7(-)/CK20(+) tumors may originate from colorectum, CK7 (+)/CK20(+) tumors may originate from pancreas, biliary tract and gastroesophageal, etc., CK7(+)/CK20(-) tumors may originate from breast and ovary, etc., and CK7(-)/CK20(-) tumors may originate from hepatocellular carcinoma and squamous cell carcinoma (8, 21). However, many tumors express more than one phenotype, especially in gastrointestinal carcinoma. In a recent meta-analysis, IHC analysis correctly distinguished the primary site in 77.7% of metastatic liver cancers with the average usage of 6.9 ± 4.1 markers (8).

In recent years, several gene expression profiling-based assays were developed to identify the primary site of metastatic tumors. This technique is based on the theory that tumors share distinct gene expression patterns specific to their sites of origin (22). A commercial assay called CancerTYPE ID (Biotheranostics, San Diego, CA, USA), which is a RT-PCR assay involves 92 genes, allowing the identification of 28 common tumor types (23). A multisite validation study done by Sarah et al. demonstrates an overall sensitivity of 87% in primary site identification (24). Another assay named Tissue of Origin (TOO) test (Vyant Bio, New Jersey, USA) is microarray-based and measured the gene expression pattern of 1550 genes that related to 15 tumor types. In a multicenter cohort of 547 specimens, the TOO assay accurately classified 87.8% of cases (25). Over the past decades, DNA methylation profiling have been developed rapidly, which could be a useful approach to unmask the primary site of CUP. Sebastian et al. reported a DNA-methylation-based assay termed

TABLE 2 | The performance of the 90-gene expression assay in liver metastases.

Tumor type	Number of samples	Correctly classified samples by the gene expression assay	Sensitivity (%)
Ovary	23	23	100
Colorectum	23	22	95.7
Breast	19	19	100
Neuroendocrine	16	15	93.8
Pancreas	16	14	87.5
Gastroesophagus	10	7	70.0
Melanoma	4	4	100
Cervix	4	4	100
Lung	3	2	66.7
Adrenal	3	3	100
Germ Cell	2	2	100
Head&neck	2	1	50.0
Sarcoma	2	2	100
Kidney	2	2	100
Urinary	1	1	100
Total	130	121	93.1

"EPICUP" for predicting primary sites of CUP (26). In a clinical validation set, EPICUP predicted a primary tumor of origin in 87% of CUP patients. More interestingly, patients with EPICUP diagnoses who received a tumor type-specific therapy showed improved overall survival compared with that in patients who received empiric therapy. However, neither of these assays has been validated in a large cohort of liver biopsy samples. Recently, only Katharina et al. reported a microRNA classifier showing an overall classification accuracy of 74.5% for primary site identification of liver biopsy specimens (27). This result was unsatisfactory for solving the urgent need of LM diagnosis in the clinic.

In the present study, the 90-gene expression assay achieved a precise classification of the tumor origin in 146 liver tumors with an overall accuracy of 93.8%, which was comparable to the EPICUP with 94% (501 of 534) accuracy in metastatic tumors (26). Moreover, the performance of the 90-gene expression assay was significantly better than the accuracy of the gold standard histopathology (77.7%) (8). In practice, the turnaround time of the 90-gene expression assay from archived FFPE samples to tumor type prediction was less than one day, which might greatly shorten patients' waiting time compared with the conventional histopathological evaluation. These results indicated that the 90-gene expression assay might serve as a useful tool for accurately identifying the tissue of origin for liver tumors. In the daily diagnostic routine, FFPE liver biopsy specimens are widely used

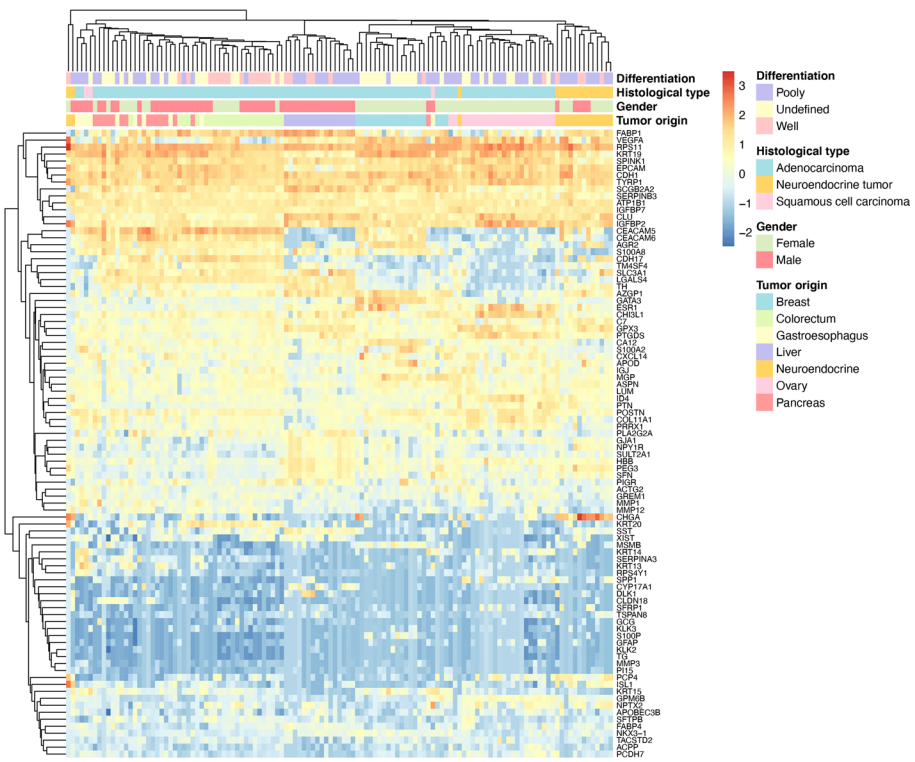


FIGURE 3 | Hierarchical clustering analysis of 90 genes in 123 specimens. The average linkage hierarchical clustering method was performed where the metric of similarity was Pearson's correlation between every pair of samples. The left panel shows a dendrogram of hierarchical clustering of 90 genes. Colored pixels capture the magnitude of the gene expression intensities, where shades of blue and red represent under-expression and over-expression, respectively, relative to the mean for each gene. The upper panel shows a dendrogram of hierarchical clustering of samples. The clinical features such as degree of differentiation, histological types, gender and tumor types of each sample are indicated in the upper panel. The number of tumor types less than five are not shown.

TABLE 3 | Investigation of nine cases misclassified by the 90-gene expression assay.

ID	Gender	Age	Sample type	Pathological diagnosis	90-gene expression assay predictions	Histological Subtype	Degree of differentiation
43	Male	69	Surgery	Gastroesophagus	Urinary	AC	Poorly
54	Male	61	Surgery	Colorectum	Gastroesophagus	AC	Well
57	Male	70	Biopsy	Pancreas	Colorectum	AC	Poorly
59	Male	74	Biopsy	Gastroesophagus	Liver	AC	Poorly
70	Male	53	Biopsy	Pancreas	Liver	AC	Poorly
75	Male	63	Biopsy	Lung	Head&neck	SCC	Poorly
114	Male	59	Surgery	Gastroesophagus	Liver	SCC	Poorly
119	Female	68	Surgery	Neuroendocrine	Endometrium	NET	Poorly
134	Female	63	Surgery	Head&neck	Liver	SCC	Poorly

AC, Adenocarcinoma; SCC, Squamous cell carcinoma; NET, Neuroendocrine tumor.

with limited amounts of tumor tissue and relatively high amounts of normal liver tissue than resections (27). Herein, this study enrolled 36 liver biopsy specimens, which were obtained by FNA or NCB. The analytic agreement reached 88.9% in biopsy specimens, which showed no statistically significant difference ($p = 0.31$) with an accuracy of 95.5% in resection specimens. Therefore, this assay could be compatible with FFPE biopsy specimens, which allows widespread access and applications in clinical practice. However, we still noticed that nine cases were misclassified. As shown in **Table 3**, the most obvious of these cases relate to poorly differentiated tumors, which are likely more susceptible to deterioration of gene expression with increasing dedifferentiation. Given six of nine misclassified cases were gastrointestinal tumors (gastroesophageal, colorectum and pancreatic tumors), it could be argued that gastrointestinal tumors indeed shared more homogenous gene expression patterns compared with other tumor types, which was also observed in the unsupervised hierarchical clustering illustration (**Figure 3**).

Previous studies have shown that the gene expression patterns were sustained in LM compared to corresponding primary tumors, but normal liver tissue contamination in the surrounding must be considered as a potential cause of misclassification in gene expression analysis (27, 28). Katharina et al. reported a microRNA-based trained without contamination consideration showing a disappointing classification accuracy of 38.2%. By adjusting for liver contamination, the classifier's accuracy was significantly improved to 67.3% (27). In the present study, we set high tumor content criteria above 60% to minimize possible confounding factors. Indeed, the overall accuracy of the 90-gene expression assay showed significant improvement compared with Katharina et al's results. However, we still observed four cases (two gastroesophageal cancer, one pancreatic cancer, and one head&neck tumor) were misclassified as primary liver cancer, suggesting the consistent source of variation induced by various tumor contents and infiltrating immune cells in the tumor environment would impact the gene expression analysis results. It's worth noting that although selecting high tumor content samples would be helpful to reduce the normal liver tissue contamination, this might significantly restrict the utilization of the gene expression assay in real clinical setting.

Besides, the present study also had several other limitations. First, the performance of the 90-gene expression assay is variable across different tumor types due to small enrolled number of certain tumor types. For example, the sensitivity is ranged from 50% (head&neck) to 100% (ovary, breast, melanoma, etc.). Further validation of the 90-gene expression assay on larger numbers of head&neck origin LM, gastroesophageal origin LM, rare LM types, and poorly differentiated LM are warranted. Second, although the 90-gene expression assay demonstrated an accuracy of 100% in classifying the neuroendocrine tumors from various origins, however, it was unable to evaluate the discriminating performance of the panel to distinguish the tumor origins of neuroendocrine tumors.

CONCLUSION

In conclusion, the results of the present study demonstrate encouraging performance of the 90-gene expression assay for distinguishing primary liver tumor from LM and identifying the primary sites of LM. In cases that morphology and IHC analyses cannot confirm the tissue of origin, the 90-gene expression assay maybe serves as a helpful instrument for discriminating the primary tumor. Future incorporation of the 90-gene expression assay in clinical diagnosis will aid oncologists in applying precise treatments, leading to improved care and outcomes for LM patients. In future studies, additional effort needs to be done for the distinguishing of head&neck origin LM, gastroesophageal origin, rare LM types, or poorly differentiated LM.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of Fudan University

Shanghai Cancer Center (FUSCC, Shanghai, China), West China Hospital Sichuan University (WCHSU, Chengdu, Sichuan, China), Sichuan Cancer Hospital (SCH, Chengdu, Sichuan, China), Chongqing Cancer Hospital (CCH, Chongqing, China), Tianjin Medical University Cancer Institute & Hospital (TMUCIH, Tianjin, China), Sir Run Run Shaw Hospital (SRRSH, Hangzhou, Zhejiang, China) and Hubei Cancer Hospital (HCH, Wuhan, Hubei, China). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

XZ, WL, BM, and JY designed the study. QW, FL, QJ, QL, HA, YL, ZL, LF, FG, and BM provided the specimens and collected clinical information. YS, YW, and WR performed the experiments. QW, YS, and QX analyzed all data. QW and YS wrote the initial manuscript draft. XZ, WL, BM, JY, and QX critically revised the manuscript and gave valuable insight to the study concept. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by research funding from Innovation Program of Shanghai Science and Technology Committee grants 20Z11900300 (XZ), Innovation Group Project of Shanghai Municipal Health Commission grants 2019CXJQ03 (XZ), Shanghai Science and Technology Development Fund grants 19MC1911000 (XZ, QW) and Shanghai Municipal Key Clinical Specialty grants shslczdzk01301 (XZ, QW), the National Natural Science Foundation of China grants 81,401,963 (QW) and 81,972,728 (QW), Fudan University Shanghai Cancer Center grant YJYQ201603 (QW), Health commission of Hubei Province scientific research project grant WJ2019H124 (JY), Hubei science and technology program grant 2020CFB869 (FG) and Canhelp Genomics Co., Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.725988/full#supplementary-material

REFERENCES

1. Tsilimigras DI, Brodt P, Clavien P-A, Muschel RJ, D'Angelica MI, Endo I, et al. Liver Metastases. *Nat Rev Dis Primers* (2021) 7(1):1–23. doi: 10.1038/s41572-021-00261-6
2. Brodt P. Role of the Microenvironment in Liver Metastasis: From Pre- to Prometastatic Niches. *Clin Cancer Res* (2016) 22(24):5971–82. doi: 10.1158/1078-0432.CCR-16-0460
3. Horn SR, Stoltzfus KC, Lehrer EJ, Dawson LA, Tchelebi L, Gusani NJ, et al. Epidemiology of Liver Metastases. *Cancer Epidemiol* (2020) 67:101760. doi: 10.1016/j.canep.2020.101760
4. Wang S, Feng Y, Swinnen J, Oyen R, Li Y, Ni Y. Incidence and Prognosis of Liver Metastasis at Diagnosis: A Pan-Cancer Population-Based Study. *Am J Cancer Res* (2020) 10(5):1477–517.
5. Rassy E, Pavlidis N. The Currently Declining Incidence of Cancer of Unknown Primary. *Cancer Epidemiol* (2019) 61:139–41. doi: 10.1016/j.canep.2019.06.006
6. Rassy E, Pavlidis N. Progress in Refining the Clinical Management of Cancer of Unknown Primary in the Molecular Era. *Nat Publ Group* (2020) 17(9):541–54. doi: 10.1038/s41571-020-0359-1
7. Tsilimigras DI, Ntanasis-Stathopoulos I, Bagante F, Moris D, Cloyd J, Spartails E, et al. Clinical Significance and Prognostic Relevance of KRAS, BRAF, PI3K and TP53 Genetic Mutation Analysis for Resectable and Unresectable Colorectal Liver Metastases: A Systematic Review of the Current Evidence. *Surg Oncol* (2018) 27(2):280–8. doi: 10.1016/j.suronc.2018.05.012
8. Amin K, El-Rayes D, Snover D, Mettler T, Vogel RI, Khalifa MA. Patterns of Immunohistochemistry Utilization in Metastases to the Liver. *Immunohistochem Mol Morphol* (2019) 27(6):441–7. doi: 10.1097/PAI.0000000000000643
9. Swaid F, Downs D, Rosemurgy AS. A Practical Approach to Liver Metastasis From Unknown Primary Cancer: What Surgeons Need to Know. *Cancer Genet* (2016) 209(12):559–66. doi: 10.1016/j.cancergen.2016.08.004
10. Varadhachary G. New Strategies for Carcinoma of Unknown Primary: The Role of Tissue-of-Origin Molecular Profiling. *Clin Cancer Res* (2013) 19 (15):4027–33. doi: 10.1158/1078-0432.CCR-12-3030
11. Greco FA, Lennington WJ. Molecular Profiling Diagnosis in Unknown Primary Cancer: Accuracy and Ability to Complement Standard Pathology. *DSJJO* (2013) 105(11):782–90. doi: 10.1093/jnci/djt099
12. Ye Q, Wang Q, Qi P, Chen J, Sun Y, Jin S, et al. Development and Clinical Validation of a 90-Gene Expression Assay for Identifying Tumor Tissue Origin. *J Mol Diagnost* (2020) 22(9):1139–50. doi: 10.1016/j.jmoldx.2020.06.005
13. Wang Q, Xu M, Sun Y, Chen J, Chen C, Qian C, et al. Gene Expression Profiling for Diagnosis of Triple-Negative Breast Cancer: A Multicenter, Retrospective Cohort Study. *Front Oncol* (2019) 9:354. doi: 10.3389/fonc.2019.00354
14. Zheng Y, Ding Y, Wang Q, Sun Y, Teng X, Gao Q, et al. 90-Gene Signature Assay for Tissue Origin Diagnosis of Brain Metastases. *J Transl Med* (2019) 17 (1):1–9. doi: 10.1186/s12967-019-2082-1
15. Xu Q, Chen J, Ni S, Tan C, Xu M, Dong L, et al. Pan-Cancer Transcriptome Analysis Reveals a Gene Expression Signature for the Identification of Tumor Tissue Origin. *Mod Pathol* (2016) 29(6):546–56. doi: 10.1038/modpathol.2016.60
16. Platt J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *JPAILM* (1999) 10(3):61–74. doi: 10.1.1.41.1639
17. Sahani DV, Bajwa MA, Andrabi Y, Bajpai S, Cusack JC. Current Status of Imaging and Emerging Techniques to Evaluate Liver Metastases From Colorectal Carcinoma. *Ann Surg* (2014) 259(5):861. doi: 10.1097/SLA.0000000000000525
18. Qin H, Wu Y-Q, Lin P, Gao RZ, Li X, Wang XR, et al. Ultrasound Image–Based Radiomics. *J Ultrasound Med* (2021) 40(6):1229–44. doi: 10.1002/jum.15506
19. Locker GY, Hamilton S. ASCO 2006 Update of Recommendations for the Use of Tumor Markers in Gastrointestinal Cancer. *Oncol JHOC* (2006) 24 (33):5313–27. doi: 10.1200/JCO.2006.08.2644
20. Meng Q, Shi S, Liang C, Liang D, Xu W, Ji S, et al. Diagnostic and Prognostic Value of Carcinoembryonic Antigen in Pancreatic Cancer: A Systematic Review and Meta-Analysis. *Oncol Targets Ther* (2017) 10:4591–8. doi: 10.2147/OTT.S145708
21. Selves J, Long-Mira E, Mathieu M-C, Rochaix P, Ilié M. Immunohistochemistry for Diagnosis of Metastatic Carcinomas of Unknown Primary Site. *Cancers* (2018) 10(4):108. doi: 10.3390/cancers10040108
22. Su AI, Welsh JB, Sapinoso LM, Kem SG, Dimitrov P, Lapp H, et al. Molecular Classification of Human Carcinomas by Use of Gene Expression Signatures. *Cancer Res* (2001) 61(20):7388–93.
23. Ma X-J, Patel R, Wang X, Salunga R, Murage J, Desal R, et al. Molecular Classification of Human Cancers Using a 92-Gene Real-Time Quantitative Polymerase Chain Reaction Assay. *Arch Pathol Lab Med* (2006) 130(4):465–73. doi: 10.1043/1543-2165(2006)130[465:MCOHCU]2.0.CO;2

24. Kerr SE, Schnabel CA, Sullivan PS, Zhang Y, Singh V, Carey B, et al. Multisite Validation Study to Determine Performance Characteristics of a 92-Gene Molecular Cancer Classifier. *Clin Cancer Res* (2012) 18(14):3952–60. doi: 10.1158/1078-0432.CCR-12-0920
25. Monzon FA, Lyons-Weiler M, Buturovic LJ, Rigi CT, Henner WD, Sciulli C, et al. Multicenter Validation of a 1,550-Gene Expression Profile for Identification of Tumor Tissue of Origin. *Am Soc Clin Oncol Educ Book* (2009) 27(15):2503–8. doi: 10.1200/JCO.2008.17.9762
26. Moran S, Martinez-Cardús A, Sayols S, Musuien E, Balana C, Estival-Gonzalez A, et al. Epigenetic Profiling to Classify Cancer of Unknown Primary: A Multicentre, Retrospective Analysis. *Lancet Oncol* (2016) 17 (10):1386–95. doi: 10.1016/S1470-2045(16)30297-2
27. Perell K, Vincent M, Vainer B, Petersen BL, Federspiel B, Moller AK, et al. Development and Validation of a microRNA Based Diagnostic Assay for Primary Tumor Site Classification of Liver Core Biopsies. *Mol Oncol* (2015) 9 (1):68–77. doi: 10.1016/j.molonc.2014.07.015
28. Elloumi F, Hu Z, Li Y, Parker JS, Gulley ML, Amos K, et al. Systematic Bias in Genomic Classification Due to Contaminating Non-Neoplastic Tissue in Breast Tumor Samples. *BMC Med Genomics* (2011) 4(1):1–12. doi: 10.1186/1755-8794-4-54

Conflict of Interest: Author YS, YW, WR, and QX were employed by the company Canhelp 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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fonc.2021.722808

90-Gene Expression Profiling for Tissue Origin Diagnosis of Cancer of Unknown Primary

Yi Zhang[1†], Lei Xia[1†], Dawei Ma[1†], Jing Wu[2], Xinyu Xu[1] and Youtao Xu[3*]*

[1] Department of Pathology, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China, [2] Department of Radiation Oncology, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China, [3] Department of Thoracic Surgery, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, China

OPEN ACCESS

Edited by:
Zhengfei Zhu,
Fudan University, China

Reviewed by:
Sabahattin Cömertpay,
Kahramanmaras Sütçü Imam University, Turkey
Giulia Maria Stella,
Fondazione Ospedale San Matteo (IRCCS), Italy

***Correspondence:**
Xinyu Xu
xxyblk@163.com
Youtao Xu
rafasmile@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

Received: *18 June 2021*
Accepted: *21 September 2021*
Published: *07 October 2021*

Citation:
Zhang Y, Xia L, Ma D, Wu J, Xu X and Xu Y (2021) 90-Gene Expression Profiling for Tissue Origin Diagnosis of Cancer of Unknown Primary. Front. Oncol. 11:722808. doi: 10.3389/fonc.2021.722808

Cancer of unknown primary (CUP), in which metastatic diseases exist without an identifiable primary location, accounts for about 3–5% of all cancer diagnoses. Successful diagnosis and treatment of such patients are difficult. This study aimed to assess the expression characteristics of 90 genes as a method of identifying the primary site from CUP samples. We validated a 90-gene expression assay and explored its potential diagnostic utility in 44 patients at Jiangsu Cancer Hospital. For each specimen, the expression of 90 tumor-specific genes in malignant tumors was analyzed, and similarity scores were obtained. The types of malignant tumors predicted were compared with the reference diagnosis to calculate the accuracy. In addition, we verified the consistency of the expression profiles of the 90 genes in CUP secondary malignancies and metastatic malignancies in The Cancer Genome Atlas. We also reported a detailed description of the next-generation coding sequences for CUP patients. For each clinical medical specimen collected, the type of malignant tumor predicted and analyzed by the 90-gene expression assay was compared with its reference diagnosis, and the overall accuracy was 95.4%. In addition, the 90-gene expression profile generally accurately classified CUP into the cluster of its primary tumor. Sequencing of the exome transcriptome containing 556 high-frequency gene mutation oncogenes was not significantly related to the 90 genes analysis. Our results demonstrate that the expression characteristics of these 90 genes can be used as a powerful tool to accurately identify the primary sites of CUP. In the future, the inclusion of the 90-gene expression assay in pathological diagnosis will help oncologists use precise treatments, thereby improving the care and outcomes of CUP patients.

Keywords: cancer of unknown primary, exome panel sequencing, expression characteristics, FFPE, RT-PCR

INTRODUCTION

Cancer of unknown primary (CUP) is a term applied to a group of heterogeneous metastatic malignancies whose primary location cannot be detected when the tumor migrates (1). In such cases, general investigations are unable to clarify the primary location at the time of diagnosis. CUP is the seventh to eighth most common cancer and is the fourth most common cause of cancer-related deaths (2). Although several advances have been made with specialized tools for cancer diagnosis, and the incidence of CUP has steadily reduced from 5% to approximately 2% of all newly diagnosed invasive cancers, the prognosis of CUP patients is still poor. Therefore, a precise diagnosis is essential for CUP patients to allow site-specific treatment and to improve their outcomes (3).

In those diagnosed with CUP, only about 20% share clinicopathological characteristics with particular known metastatic cancers. Median survival in this group may be as long as 24 months with treatment directed at the likely primary site, usually under the supervision of the relevant site-specific treatment (4). The remaining 80% response to systemic therapy is often limited, with a median survival of 6–9 months (5). In addition, the lack of primary tumor definition, prevent most patients to be treated in clinical practice with a novel, very effective treatment such as immunotherapy or molecular targeted therapies for which currently registered indications are mostly disease-oriented. Thus, an accurate diagnosis is urgently needed to identify the most probably site-of-origin or an approach based on personalized medicine. It is useful to assist in the selection of the best treatment options and potentially improve CUPs prognosis and survival.

Generally, the diagnosis of CUP is based on the European Society of Medical Oncology guidelines (6). Clinical manifestations, tumor markers, and imaging diagnostic analyses are used to clarify the origin of metastatic cancer. In clinical practice, histopathologic and immunohistochemistry (IHC) analyses are still particularly important for identifying the anatomical origin of CUP patients. However, these traditional approaches become difficult when the hypothetical primary malignant tumor grows too large to be identified before it migrates (7).

As an alternative, molecular structure analysis of malignant tumors is a promising technology that can improve the diagnosis of the origin in CUP patients (8). Currently, there are several available testing methods that use reverse transcription-polymerase chain reaction (RT-PCR) or genetic microarray technology (9). In a previous study, a microarray-based 1550-gene expression profile was used to distinguish the primary site in 13 specimens known to originate from brain metastases, and excellent results were obtained with an accuracy of 92.3% (10). Another study revealed the expression characteristics of 154 genes, which could accurately classify 21 common types of malignant tumors (11). It is hoped that this gene panel will become an effective tool for identifying the origin of malignant tumors (12). The RT-PCR technology is more convenient and accurate than the microarray technology. It can also be used for formalin-fixed paraffin-embedded (FFPE) specimen collection and is widely used in clinical medicine (13).

In this study, we investigated the accuracy of a 90-gene expression assay for classifying 21 types of malignant tumors, in comparison with the actual disease diagnosis. In a multisite study, the performance of the 90-gene expression assay was illustrated in 609 tumor samples of known primary origin, with an accuracy of 89.8%. More specifically, the classification accuracy reached 90.4% in primary tumors and 89.2% in metastatic tumors (14).In addition, a full exome analysis was carried out to find the mutation spectrum that is likely to be beneficial for revealing the vulnerabilities of CUP patients.

Abbreviations: CUP, cancer of unknown primary; FFPE, formalin-fixed paraffin-embedded; IHC, immunohistochemistry; RT-PCR, reverse transcription-polymerase chain reaction; ASPN, Asporin; GATA3, GATA binding protein 3; VEGF-A, vascular endothelial growth factor A; AE1/AE3, pancytokeratin; CK5/6, Cytokeratin 5/6; P40, polypeptide 40; CK7, Cytokeratin7; TTF-1, Thyroid transcription factor-1; CK20, Cytokeratin 20; Hep-1, hepatocyte paraffin 1; Arg-1, Arginase-1; PAX-8, The paired box transcription factor; CDX-2, Caudal-related homeobox transcription factor 2; Syn, synaptophysin; CgA, chromogranin A; NapsinA, cytoplasmic aspartic protease.

MATERIALS AND METHODS

Patient and FFPE Specimen Collection

CUP samples of 44 patients archived from May 2018 to December 2020 were used. All hematoxylin and eosin-stained slides were assessed by two pathologists to ensure consistency with the reference diagnosis and certify the percentage of malignant cells. The inclusion criteria were as follows: i) availability of FFPE tumor tissue samples, ii) disease diagnosis included in the 21 types of tumors for the 90-gene panel, and iii) at least 60% tumor cell content based on hematoxylin and eosin-stained slides. Clinical data, physical examination information, and results of imaging, light microscopy, and IHC tests were obtained from medical records.

Details of the 90 Tumor-Specific Genes

Initially, the tumor-specific genes were identified on basis of a pan-cancer transcriptome database comprising 5434 samples representing 21 tumor types (14). Next, the Top-10 most predictive genes for each of the 21 tumor types were screened by using the Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm. A list of 90 genes corresponding to 21 tumor types was identified after removing redundant genes (**Additional Table 1**). Finally, an SVM linear model was trained using the whole pan-cancer transcriptome database to establish a multiclass classification algorithm termed "90-gene classifier". The details of the 90 specific genes were list in the **Additional Table 2**.

Intuitively, the similarity scores for each of the 21 tumor types were calculated by the 90-gene classifier, which reflect how much the gene expression pattern of the test specimen is similar to the global gene expression pattern with known tumor type. The similarity scores were probability-based, with a reported range from 0 to 100, and all 21 similarity scores sum to 100. The tumor

type with the highest similarity score was considered as the predicted tumor type by the 90-gene classifier.

Sample Preparation and RNA Isolation

An FFPE Total RNA Isolation Test Kit (Canhelp Genomics, Hangzhou, China) was used to isolate total RNA from FFPE samples, as described previously. Briefly, tumor sections were placed in a small 1.5 mL centrifuge tube, deparaffinized with xylene at 50°C for 3 minutes, and then washed twice with 100% alcohol. The samples were incubated in a trypsin K aqueous solution at 56°C for 15 minutes and subsequently at 80°C for another 15 minutes. DNase was then used to digest and absorb the protein. We used 40 μL of RNase-free water to obtain the total RNA. The concentration of total RNA was measured with a 260 nm photometer, and the purity was measured by the A260/A280 ratio. The RT-PCR (Applied Biosystems 7500) analysis was only carried out on RNA samples with A260/A280 ratios between 1.2 and 2.4.

Expression Profiling of 90 Tumor-Specific Genes

For each sample, cDNA was generated from total isolated RNA using a High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems, Foster City, CA, United States). RT-PCR was used to analyze the expression profiles of 90 specific genes in malignant tumors on a 96-well plate.

Downloading Public Data and Analyses

RNA expression profiles (workflow type: HT Seq-Counts) and the patients' clinical information were downloaded from The Cancer Genome Atlas website using the "TGCA biolinks" R package (Version 2.14.1). In order to acquire the relative expression of each mRNA, we conducted normalize each counts *via* a standard pipeline of "DESeq2" R package (Version 4.1). "Combat" function of "sva" R package (Version 3.0) was used to removed batch effects between our panel profiling and TCGA RNA-seq.

Library Preparation and Sequencing

For targeted therapeutic transcriptome sequencing, genomic DNA from FFPE sections or biopsy samples and the whole blood control samples were extracted with QIAamp DNA FFPE Tissue Kit and DNeasy Blood and tissue kit (Qiagen), respectively, while cfDNAs from whole blood which were collected with Streck Cell-Free DNA BCT were extracted with QIAamp Circulating Nucleic Acid Kit (Qiagen) and then quantified with Qubit 3.0 using the dsDNA HS Assay Kit (Thermo-Fisher Scientific). Library preparations were performed using the KAPA HyperPlus Prep Kit (KAPA Biosystems). For the targeted panel, customized xGen lockdown probes (Integrated DNA Technologies) targeting 556 cancer-relevant genes were used for hybridization enrichment. The capture reaction was performed with Dynabeads M-270 (Life Technologies) and the xGen Lockdown Hybridization and Wash Kit (Integrated DNA Technologies), according to the manufacturers' protocols. Captured libraries were subjected to on-beads PCR amplification with Illumina p5 (5' AAT GAT ACG GCG ACC ACC GA 3') and p7 primers (5' CAA GCA GAA GAC GGC ATA CGA GAT 3') using the KAPA HiFi HotStart ReadyMix (KAPA Biosystems), followed by purification using Agencourt AMPure XP beads. Libraries were quantified by qPCR using the KAPA Library Quantification Kit (KAPA Biosystems). The library fragment size was determined using the Bioanalyzer 4200 (Agilent Technologies). The target-enriched library was then sequenced on a NovaSeq 6000 system (Illumina) according to the manufacturer's instructions, with an average coverage depth of 2000× for tumors and 8000× for cfDNA (500× in normal blood controls) using a panel. The average mask size of the panel was 2.8 Mb.

Statistical Analysis

Gene expression data analysis was performed using R software and packages from the Bioconductor project. The gene expression mode of each specimen was compared with 21 specific types of malignant tumors based on the expression characteristics of 90 genes (**Additional Tables 1, 2**). For each of the 21 types of malignant tumors, similarity scores were calculated, which indicated the similarities of the intermediate gene expression patterns between the template and the specific malignant tumor type. The similarity score ranged from 0 (very low similarity) to 100 (very high similarity). The type of malignant tumor with the highest similarity score was considered to indicate the origin. For each specimen, the primary location of the malignant tumor was predicted and analyzed in comparison with the clinical reference diagnosis.

RESULTS

Patients and Samples

The flow diagram for the study participants is shown in **Figure 1**. From May 2018 to December 2020, 44 patients were recruited from Jiangsu Oncology Hospital for this study. As shown in **Table 1**, the study included 23 men and 21 women with a mean age of 59 years (range 32–89 years). The 44 specimens were classified into 12 types based on the location of malignant tumor invasion, including the cervical lymph nodes, axillary lymph nodes, groin lymph nodes, head and neck, lungs, liver, female genitalia, omentum, abdominal cavity, retroperitoneum, bone, and sternum. The most common erosion sites for the first, second, and third malignant tumors were the cervical lymph nodes (20.4%), head and neck (18.2%), and lungs (13.6%), respectively. Among the 44 specimens collected, 15 (34%) were well-differentiated tumors and 29 (66%) were poorly differentiated. In accordance with the European Society of

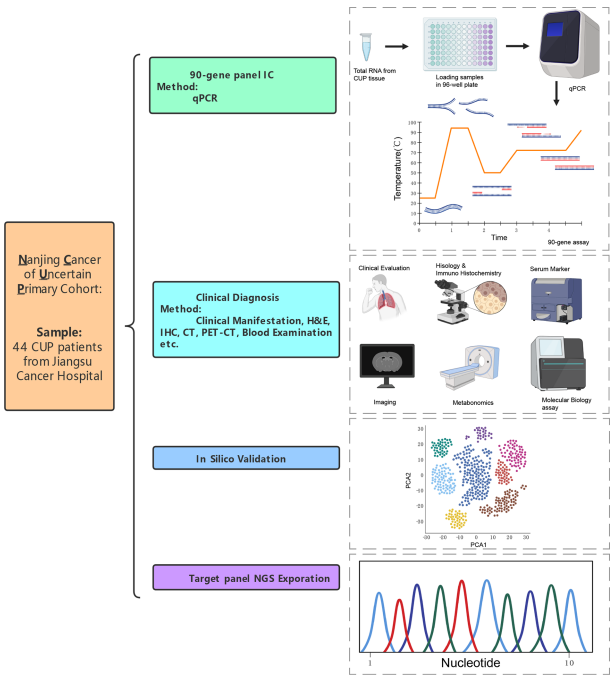


FIGURE 1 | An overview of research design that verification of 90-gene expression signature to identify the origin of CUP.

TABLE 1 | Patients and clinical characteristics.

Characteristic	No. of specimens	Percentage (%)
Gender		
Male	23	53
Female	21	47
Age at diagnosis		
Mean	59	
Range	32-89	
Diagnostic Method		
Biopsy	24	54.5
Surgery	20	45.5
Invasion site		
Cervical Lymph nodes	9	20.4
Axillary Lymph nodes	3	6.8
Inguinal Lymph nodes	3	6.8
Neck	8	18.2
Lung	6	13.6
Liver	4	9
Female reproductive	4	9
Omentum	2	4.5
Abdomen	2	4.5
Peritoneum	1	2.2
Bone	1	2.2
Chest wall	1	2.2
Histology		
Well-differentiated	15	34
Poorly-differentiated	29	66

Medical Oncology manual, a comprehensive CUP multi-process exercise was performed to eliminate the presence of primary malignant tumors (**Figure 2**).

The Characteristics of the Expression Characteristics of 90 Genes in CUP

Total RNA was isolated from tissue sections of the 44 samples. The concentration range was 2.59–653 ng/μL, with an average of 106 ng/μL. The A260/A280 ratio ranged from 1.2 to 2.4. In **Figure 3**, the red grid represents the non-compliance with the reference diagnosis, and the blue grid represents the compliance with the clinically predicted diagnosis. After analyzing the clinical data of these forty-four patients, it was found that the diagnosis coincidence rate was 81.2% based on immunohistochemistry analysis; the diagnosis coincidence rate based on morphological analysis was 88.6%; the diagnosis coincidence rate based on serological examination was 40.9% and the coincidence rate based on imaging diagnosis alone was 34%. Of the 44 samples, 95.4% (42/44) showed agreement between the prediction of the 90-gene expression assay and the reference diagnosis.

The molecular structure classifications were inconsistent in two of the specimens, as shown in **Table 2**. In one case, a malignant tumor of the cervical lymph nodes was predicted to be liver adenocarcinoma, but the IHC marker for liver adenocarcinoma was negative. The reason for the case is the limitation of IHC and the tumor heterogeneity that affects antibody expression. The other inconsistency was from a patient whose malignant tumor was diagnosed by pathophysiologic examination as a poorly differentiated cancer that migrated to the lung, but was assessed as sarcoma by the 90-gene expression analysis. The excuse can be due to the abundant necrosis decreased the amount of entity of tumor, and meanwhile the scant components of tumor affected the accuracy of 90-gene analysis. This phenomenon of necrosis is the reflected response to radiotherapy and chemotherapy treatment and always be the consequence of insufficient blood supply.

In-Silico Validation of the 90 Gene Expression Signature in CUP

To further validate the performance of the 90-gene expression signature in CUP, we evaluated the consistency of the 90-gene expression profile between the CUP primary tumor and metastatic tumor. We downloaded the transcriptome data of lung adenocarcinoma, lung squamous cell carcinoma, gastric adenocarcinoma, head and neck squamous cell carcinoma, breast cancer, and pancreatic cancer from The Cancer Genome Atlas database and extracted the 90-gene expression signatures. Satisfactorily, the 90-gene expression profile generally accurately classifies CUP into the cluster of its primary tumor (**Figure 4**), suggesting that the 90-gene expression signal can cover the transcriptome characteristics of the patient's primary tumor.

Exploration of Exome Panel Sequencing in CUP Patients

Moreover, we conducted exome panel sequencing in nine patients with cervical lymph node metastatic tumors. We performed a high-throughput exome sequencing technology containing 556 cancer-related genes. It was a pan-solid tumor-

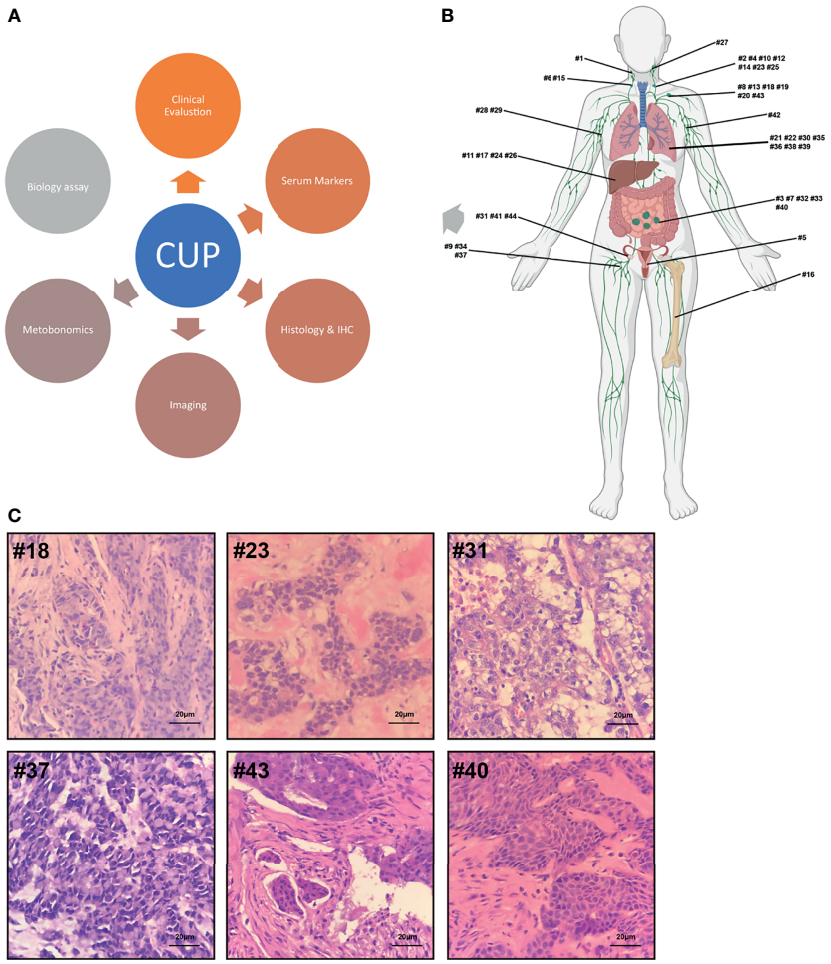


FIGURE 2 | **(A)** Multistep origin of CUP diagnostic workflow. **(B)** Metastases distribution: The 44 metastases were retrieved from surgery and puncture biopsy specimens from 44 CUP patients. **(C)** Histology: Most metastatic lesions are composed of poorly differentiated.

associated big panel that includes comprehensive mutational information to guide targeted therapy and immunotherapy. Moreover, gene associated with sensitivity or tolerance to chemoradiotherapy was also included to help optimize treatment options (**Figure 5**). Unfortunately, the single nucleotide polymorphisms and copy number variations of the 556 genes did not correlate significantly with the expression tag of the 90 genes, nor with tumor mutation burden or *PD-L1* expression. This also reflects the irreplaceable expression profile of the 90-gene expression signature.

DISCUSSION

In the present study, the expression characteristics of 90 genes were validated in CUP patients, and the overall agreement with the reference diagnosis was 95.4%. Although the expression characteristics of 90 genes mainly showed accurate aspect ratios for classifying CUP malignant tumors, we noticed that two medical records were incorrectly classified. In one case, a negative result was obtained from an IHC analysis that used markers for malignant tumors, indicating that the pathological diagnosis was limited.

When addressing the potential diagnosis of CUP, clinical practice suggests a complete diagnosis based on the European Society of Medical Oncology guidelines (15). The diagnostic test is based on the precise location of the migration foci, the patient's clinical symptoms and sex, magnetic resonance imaging, endoscopy, positron emission tomography, or evaluation of special serum protein tumor markers. The final step is IHC testing, which is still the most important diagnostic tool for identifying the origin (16). In 69% of cases, histological or IHC analysis clarified the primary location of a series of poorly differentiated tumors (17). Many molecular structure tests based on gene expression profiles have shown useful value in identifying the type of primary malignant tumors in patients with unknown or uncertain diagnoses (18).

A previous meta-analysis showed that IHC provided appropriate origin identification in only 65.6% of metastatic

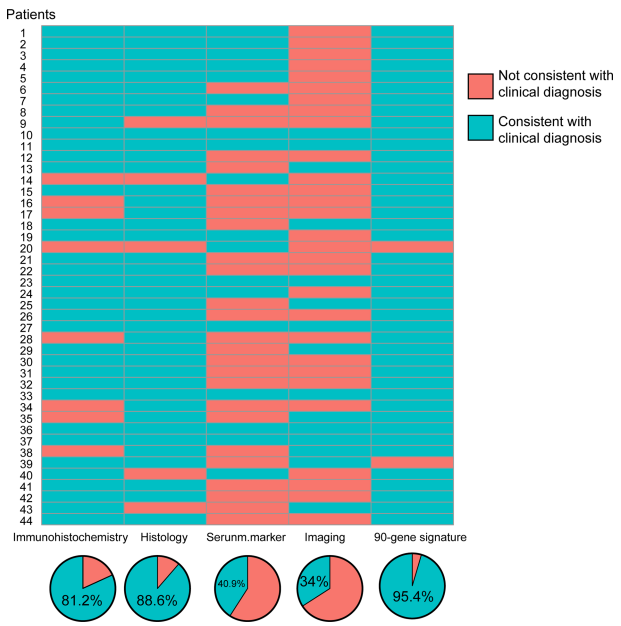


FIGURE 3 | Consistency and coincidence rate of CUP origin multidisciplinary diagnosis results with 90-gene expression signature.

cancers. Recent studies have also shown that molecular structure profiling is better than IHC classification, especially in cases of poorly differentiated malignant tumors (19). The second incorrectly classified case was a metastatic malignant tumor with a wide area of necrosis and was classified as sarcoma based on gene expression. A large amount of necrosis reduced the total number of solid lines of malignant tumors, and the lack of malignant tumor components affected the accuracy of the 90-gene expression assay. This type of necrosis reflects radiotherapy and chemotherapy and is usually the result of insufficient blood supply (20). Therefore, 90-gene expression assay can be used to completely classify these undiagnosed malignancies.

In the molecular era, gene expression profiling and next-generation sequencing have been proposed to identify the site of origin and to replace the standard pathological examinations based on histologic examination and IHC (21). Next-generation sequencing panels are still the most common way to identify the primary location of CUP and establish targeted drug therapy (22). However, our study showed that the origin of CUP cannot be confirmed based on the results of a common next-generation sequencing panel with transcriptome sequencing (including gene mutation, copy number variation, microsatellite instability, and tumor mutational burden). This may be due to the heterogeneity of intermediate genes between primary and migrating malignancies (23). For example, case #44 in our study involved liver metastasis from a mutation in the *KRAS p.G12V* gene, which is a common mutation in pancreatic tumors. Comprehensive clinical symptoms, pathologic examination, imaging diagnosis, and serological examination all indicated that the primary malignant tumor originated from the colon. In addition, the composition of *BRAF* and *MEK* inhibitors is reasonable for melanoma patients with *BRAF V600* mutations but has limited efficacy for patients with rectal cancer (24).

Although research on predicting the origin of CUP through transcriptome data is gaining popularity, there is a widespread lack of clinical research certification and clinical medical applications, especially in China (25). Since RNA sequencing requires a lot of resources for specimen collection, the results are unstable, and a large financial burden is placed on the patient. A simple, cheap, and stable PCR method must be applied to assess the expression levels of 90 genes in patients. In a previous study, a 92-gene control panel developed by Ma et al. was able to distinguish the origin of CUP with an overall sensitivity of 87% (26). In another study, a 10-gene control panel could distinguish between six types of multiple malignancies. These reflect the feasibility of PCR to assess the origin of CUP (27). Moran et al. developed and designed a DNA methylation profile for CUP patients with an overall accuracy of 90% (28). This inspired us to further improve the accuracy of CUP origin identification from an epigenetic perspective in the future.

In our study, the 90-gene expression profile was similar for metastatic malignant tumors and primary malignant tumors. This shows that secondary malignant tumors and metastatic malignant tumors have molecular structures similar to those of malignant progression. This provides an identification method for CUP and the possible vulnerabilities of CUP patients, including ASPN (29), GATA3 (30), and VEGF-A (31). Meanwhile, some scientific research groups have reported many genetic changes in CUP malignant tumor origins or liquid biopsies (32). In addition, studies investigating the presence of driver mutations and molecular aberrations in CUP provide conflicting evidence on whether these changes are "potentially druggable" (33).

In conclusion, this 90-gene expression assay can identify the source of malignant tumors with good accuracy, and the expression data can cover the transcriptome characteristics of the patient's primary malignant tumor. In addition, we carried

TABLE 2 | Investigation of cases with discordant 90-gene expression.

ID	Reference diagnosis	History	Immunohistochemical staining	90-gene expression results	Similarity score
#20	Poorly differentiated cancer	A 48-year-old man with tumor in cervical lymph nodes	AE1/AE3++; CK5/6-; P40-; CK7+; TTF-1-; CK20-; Hep-1-; Arg-1-; PAX-8-; GATA-3-; CDX-2-.	Hepatobiliary tumors	54.3
#39	Lung epithelial tumor	A 49-year-old man with a tumor in the lung	AE1/AE3+; Syn-; CgA-; NapsinA+; PAX-8-; GATA-3-; CDX-2-.	sarcoma	48.8

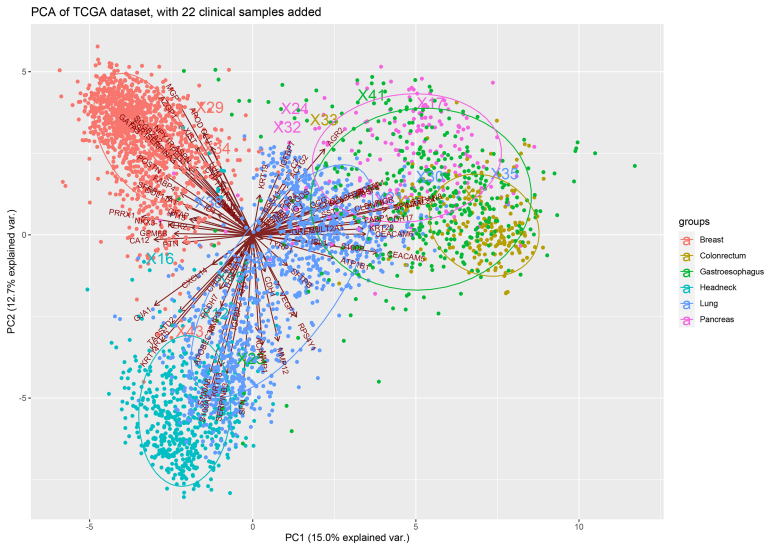


FIGURE 4 | Principal component analysis of 90-gene expression between CUP and primary tumor from TCGA database.

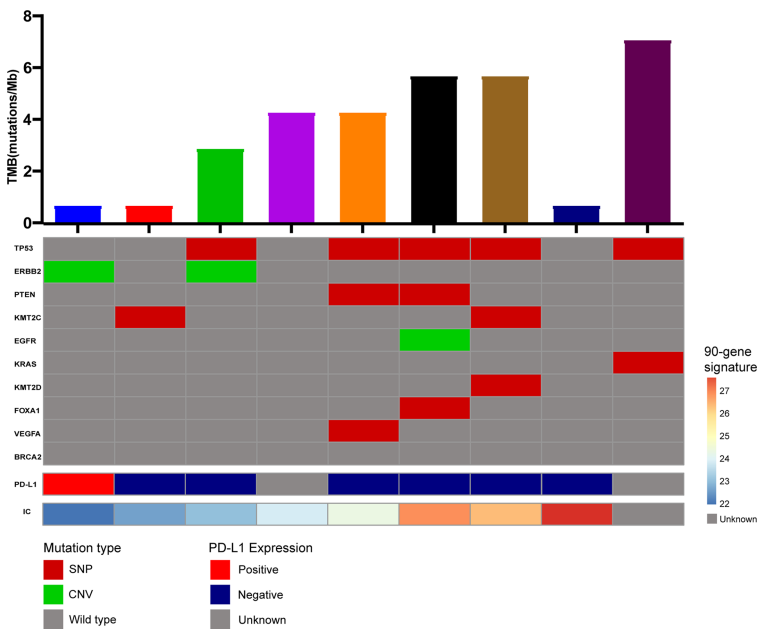


FIGURE 5 | NGS panel sequencing of metastatic tissues from 9 CUP patients.

out exome transcriptome sequencing, including 556 high-frequency gene mutation oncogenes, to explore the gene expression of CUP patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Genomics Data Center, HRA001047.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nanjing Medical University. 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

YZ, LX, and DM participated in the study design and data analysis and draft the manuscript. JW and YX carried out the immunoassays and help to draft the manuscript. XX conceived the study, participated in its design and coordination, and help to draft the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by grants from Jiangsu Cancer Hospital (numbers ZM202006), the National Natural Science Foundation of China (number 81702692), the Jiangsu Province "Six Talent Peaks Project" (WSW-019).

ACKNOWLEDGMENTS

The authors thank the patients for their willingness to cooperate with our study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.722808/full#supplementary-material.

REFERENCES

1. Lombardo R, Tosi F, Nocerino A, Bencardino K, Gambi V, Ricotta R, et al. The Quest for Improving Treatment of Cancer of Unknown Primary (CUP) Through Molecularly-Driven Treatments: A Systematic Review. *Front Oncol* (2020) 10:533. doi: 10.3389/fonc.2020.00533
2. Loef C, Merkx MAW. Cancer of Unknown Primary Origin; Know the Enemy. *Eur J Cancer Care (Engl)* (2021) 00:e13460. doi: 10.1111/ecc.13460
3. Yan N, Zhang Y, Guo X, Yuan D, Tian G, Yang J. A Review on Cancer of Unknown Primary Origin: The Role of Molecular Biomarkers in the Identification of Unknown Primary Origin. *Methods Mol Biol* (2020) 2204:109–19. doi: 10.1007/978-1-0716-0904-0_10
4. Pavlidis N, Pentheroudakis G. Cancer of Unknown Primary Site. *Lancet* (2012) 379(9824):1428–35. doi: 10.1016/S0140-6736(11)61178-1
5. Jones W, Allardice G, Scott I, Oien K, Brewster D, Morrison DS. Cancers of Unknown Primary Diagnosed During Hospitalization: A Population-Based Study. *BMC Cancer* (2017) 17(1):85. doi: 10.1186/s12885-017-3083-1
6. Benvenuti S, Milan M, Geuna E, Pisacane A, Senetta R, Gambardella G, et al. Cancer of Unknown Primary (CUP): Genetic Evidence for a Novel Nosological Entity? *A Case Rep EMBO Mol Med* (2020) 12(7):e11756. doi: 10.15252/emmm.201911756
7. Lee MS, Sanoff HK. Cancer of Unknown Primary. *BMJ* (2020) 371:m4050. doi: 10.1136/bmj.m4050
8. Rassy E, Pavlidis N. Progress in Refining the Clinical Management of Cancer of Unknown Primary in the Molecular Era. *Nat Rev Clin Oncol* (2020) 17 (9):541–54. doi: 10.1038/s41571-020-0359-1
9. Dermawan JK, Rubin BP. The Role of Molecular Profiling in the Diagnosis and Management of Metastatic Undifferentiated Cancer of Unknown Primary(): Molecular Profiling of Metastatic Cancer of Unknown Primary. *Semin Diagn Pathol* (2020). doi: 10.1053/j.semdp.2020.12.001
10. Wu AH, Drees JC, Wang H, VandenBerg SR, Lal A, Henner WD, et al. Gene Expression Profiles Help Identify the Tissue of Origin for Metastatic Brain Cancers. *Diagn Pathol* (2010) 5:26. doi: 10.1186/1746-1596-5-26
11. Xu Q, Chen J, Ni S, Tan C, Xu M, Dong L, et al. Pan-Cancer Transcriptome Analysis Reveals a Gene Expression Signature for the Identification of Tumor Tissue Origin. *Mod Pathol* (2016) 29(6):546–56. doi: 10.1038/modpathol.2016.60
12. Rassy E, Assi T, Pavlidis N. Exploring the Biological Hallmarks of Cancer of Unknown Primary: Where do We Stand Today? *Br J Cancer* (2020) 122 (8):1124–32. doi: 10.1038/s41416-019-0723-z
13. Awadallah SF, Bowling MR, Sharma N, Mohan A. Malignant Pleural Effusion and Cancer of Unknown Primary Site: A Review of Literature. *Ann Transl Med* (2019) 7(15):353. doi: 10.21037/atm.2019.06.33
14. Ye Q, Wang Q, Qi P, Chen J, Sun Y, Jin S, et al. Development and Clinical Validation of a 90-Gene Expression Assay for Identifying Tumor Tissue Origin. *J Mol Diagn* (2020) 22(9):1139–50. doi: 10.1016/j.jmoldx.2020.06.005
15. Davalos V, Esteller M. Insights From the Genetic and Transcriptional Characterization of a Cancer of Unknown Primary (CUP). *EMBO Mol Med* (2020) 12(7):e12685. doi: 10.15252/emmm.202012685
16. Timar J. [Pathological Diagnostics of CUP (Cancer of Unknown Primary)]. *Magy Onkol* (2019) 63(2):67–74. doi: 10.1007/978-3-319-56424-1_1
17. Weiss LM, Chu P, Schroeder BE, Singh V, Zhang Y, Erlander MG, et al. Blinded Comparator Study of Immunohistochemical Analysis Versus a 92-Gene Cancer Classifier in the Diagnosis of the Primary Site in Metastatic Tumors. *J Mol Diagn* (2013) 15(2):263–9. doi: 10.1016/j.jmoldx.2012.10.001
18. Moran S, Martinez-Cardus A, Boussios S, Esteller M. Precision Medicine Based on Epigenomics: The Paradigm of Carcinoma of Unknown Primary. *Nat Rev Clin Oncol* (2017) 14(11):682–94. doi: 10.1038/nrclinonc.2017.97
19. Anderson GG, Weiss LM. Determining Tissue of Origin for Metastatic Cancers: Meta-Analysis and Literature Review of Immunohistochemistry Performance. *Appl Immunohistochem Mol Morphol* (2010) 18(1):3–8. doi: 10.1097/PAI.0b013e3181a75e6d
20. Zhang X, Li Z, Sun Y, Liu G, Liu X, Zhou Y. Overexpression of Tumour Necrosis Factor-Alpha-Induced Protein 8 Is Associated With Prognosis in Colon Cancer. *Cancer Manag Res* (2021) 13:4055–65. doi: 10.2147/CMAR.S297451
21. Laprovitera N, Riefolo M, Ambrosini E, Klec C, Pichler M, Ferracin M. Cancer of Unknown Primary: Challenges and Progress in Clinical Management. *Cancers (Basel)* (2021) 13(3):451. doi: 10.3390/cancers13030451
22. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, et al. Site-Specific and Targeted Therapy Based on Molecular Profiling by Next-Generation Sequencing for Cancer of Unknown Primary Site: A Nonrandomized Phase 2 Clinical Trial. *JAMA Oncol* (2020) 6(12):1931–8. doi: 10.1001/jamaoncol.2020.4643
23. Wei Q, Ye Z, Zhong X, Li L, Wang C, Myers RE, et al. Multiregion Whole-Exome Sequencing of Matched Primary and Metastatic Tumors Revealed Genomic Heterogeneity and Suggested Polyclonal Seeding in Colorectal Cancer Metastasis. *Ann Oncol* (2017) 28(9):2135–41. doi: 10.1093/annonc/mdx278
24. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved Survival With Vemurafenib in Melanoma With BRAF V600E Mutation. *N Engl J Med* (2011) 364(26):2507–16. doi: 10.1056/NEJMoa1103782
25. Li R, Liao B, Wang B, Dai C, Liang X, Tian G, et al. Identification of Tumor Tissue of Origin With RNA-Seq Data and Using Gradient Boosting Strategy. *BioMed Res Int* (2021) 2021:6653793. doi: 10.1155/2021/6653793
26. Kerr SE, Schnabel CA, Sullivan PS, Zhang Y, Singh V, Carey B, et al. Multisite Validation Study to Determine Performance Characteristics of a 92-Gene Molecular Cancer Classifier. *Clin Cancer Res* (2012) 18(14):3952–60. doi: 10.1158/1078-0432.CCR-12-0920
27. Talantov D, Baden J, Jatkoe T, Hahn K, Yu J, Rajpurohit Y, et al. A Quantitative Reverse Transcriptase-Polymerase Chain Reaction Assay to Identify Metastatic Carcinoma Tissue of Origin. *J Mol Diagn* (2006) 8 (3):320–9. doi: 10.2353/jmoldx.2006.050136

28. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A Microrna Expression Signature of Human Solid Tumors Defines Cancer Gene Targets. *Proc Natl Acad Sci USA* (2006) 103(7):2257–61. doi: 10.1073/pnas.0510565103
29. Hughes RM, Simons BW, Khan H, Miller R, Kugler V, Torquato S, et al. Asporin Restricts Mesenchymal Stromal Cell Differentiation, Alters the Tumor Microenvironment, and Drives Metastatic Progression. *Cancer Res* (2019) 79(14):3636–50. doi: 10.1158/0008-5472.CAN-18-2931
30. Lin MC, Lin JJ, Hsu CL, Juan HF, Lou PJ, Huang MC. GATA3 Interacts With and Stabilizes HIF-1alpha to Enhance Cancer Cell Invasiveness. *Oncogene* (2017) 36(30):4380. doi: 10.1038/onc.2017.196
31. Claesson-Welsh L, Welsh M. VEGFA and Tumour Angiogenesis. *J Intern Med* (2013) 273(2):114–27. doi: 10.1111/joim.12019
32. Gatalica Z, Millis SZ, Vranic S, Bender R, Basu GD, Voss A, et al. Comprehensive Tumor Profiling Identifies Numerous Biomarkers of Drug Response in Cancers of Unknown Primary Site: Analysis of 1806 Cases. *Oncotarget* (2014) 5(23):12440–7. doi: 10.18632/oncotarget.2574
33. Conway AM, Mitchell C, Kilgour E, Brady G, Dive C, Cook N. Molecular Characterisation and Liquid Biomarkers in Carcinoma of Unknown Primary (CUP): Taking the 'U' Out of 'Cup'. *Br J Cancer* (2019) 120(2):141–53. doi: 10.1038/s41416-018-0332-2

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 in Oncology

CASE REPORT
published: 11 November 2021
doi: 10.3389/fonc.2021.702887

Case Report: Tissue Origin Identification for Cancer of Unknown Primary: Gene Expression Profiling Approach

Xingxiang Pu[1†], Sa Yang[2†], Yan Xu[1], Bolin Chen[1], Qianzhi Wang[1], Qian Gong[3] and Lin Wu[1]*

[1] Department of Thoracic Medical Oncology, Hunan Cancer Hospital/The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China, [2] Department of Pharmacy, Xiangxiang People's Hospital, Xiangxiang, China, [3] Department of Pharmacy, Hunan Cancer Hospital/The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Yi Ba,
Tianjin Medical University, China

Reviewed by:
Jie Xu,
Harvard Medical School, United States
Vassilis Georgoulias,
University of Crete, Greece
Haiyang Zhang,
Tianjin Medical University, China

***Correspondence:**
Lin Wu
wulin-calf@vip.163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

Received: *30 April 2021*
Accepted: *25 October 2021*
Published: *11 November 2021*

Citation:
Pu X, Yang S, Xu Y, Chen B, Wang Q, Gong Q and Wu L (2021) Case Report: Tissue Origin Identification for Cancer of Unknown Primary: Gene Expression Profiling Approach. Front. Oncol. 11:702887. doi: 10.3389/fonc.2021.702887

The treatment of cancer of unknown primary (CUP) is a huge challenge for clinicians. Gene expression profiling can help identify the tissue origin of tumors by detecting the expression levels of specific genes in tumor tissues. Herein, we report four CUP cases. All of them have been successfully identified with the corresponding primary tumor sites through gene expression profiling analysis. Then all patients received accurate treatment, providing reference to guide therapeutic decisions to treat CUP tumors in the future.

Keywords: cancer of unknown primary, gene expression profiling, tissue of origin, real-time PCR, immunohistochemistry

INTRODUCTION

Cancer of unknown primary (CUP) is a well-recognized clinical syndrome, in which clinical, radiographic, and pathologic evidence of a primary site of origin is lacking (1). Adenocarcinoma and poorly differentiated carcinoma are the most common tissue types of CUP. As reported to be the malignant tumor with the seventh or eighth highest incidence, CUP accounts for 3–5% of total cancer diagnoses worldwide; and with its short course and rapid progression (2, 3), it is also the fourth most common cause of cancer death (4). Life expectancy statistics largely vary, but one meta-analysis indicated a median survival of only 4.5 months, with a 1-year survival rate of 20% and a 5-year survival rate of 4.7% (5). Therefore, an accurate diagnosis is essential for treatment and prognosis.

Recently, researchers have found that the gene expression profiling (GEP) of metastatic tumors differ from those of tissues at the metastatic site but are similar to those of tissues at the primary site, suggesting that during the process of tumorigenesis, development, and metastasis, the metastatic tumor retains the gene expression characteristics of its tissue of origin. Based on this assumption, Ye et al. developed a real-time PCR (RT-PCR) based assay (the 90-gene expression assay) for identifying the tissue of origin based on GEP analysis (6). Initially, a pan-cancer transcriptome database comprising 5434 samples representing 21 tumor types was established to identify the tumor-specific genes. Next, the Top-10 most predictive genes for each of the 21 tumor types were selected by using the Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm. Finally, a list of

90 genes was identified after removing redundant genes and the 90 genes were used for establishing an SVM linear model termed "90-gene classifier" to identify tumor tissue of origin. Intuitively, the similarity scores for each of the 21 tumor types were calculated by the 90-gene classifier, which reflects how much the gene expression pattern of the test specimen is similar to the global gene expression pattern with known tumor type. The similarity scores were probability-based, with a reported range from 0 to 100, and all 21 similarity scores sum to 100. The tumor type with the highest similarity score was considered as the predicted tumor type by the 90-gene classifier. In a multisite study, the performance of the 90-gene expression assay was illustrated in 609 tumor samples of known primary origin, with an accuracy of 89.8%. More specifically, the classification accuracy reached 90.4% in primary tumors and 89.2% in metastatic tumors. Moreover, in a real-world cohort comprising 141 CUP patients, the 90-gene expression assay was able to provide helpful predictions of tumor origin in 71.6% of patients (101 of 141) (6). In the present study, we employed the 90-gene expression assay to identify the primary sites of four CUP cases and evaluated the therapeutic outcomes in these patients after receiving treatment guided by GEP-based source-tracing, aiming to provide referencing information to guide the diagnosis and treatment decisions in other CUP cases.

CASE REPORT

Case 1: A 45-year-old female was admitted to a tertiary hospital in Xinjiang Province due to vaginal bleeding following three months of sexual activity and had just been diagnosed with cervical cancer on March 10, 2014. Computed tomography (CT) examination revealed cervical cancer and multiple enlarged lymph nodes in the pelvic cavity and bilateral groin area. Histopathological analysis from an outside hospital suggested moderately differentiated cervical squamous cell carcinoma, and interventional treatment and chemotherapy (paclitaxel combined with cisplatin) were administered. Then, laparoscopic extensive uterus, bilateral salpingectomy, pelvic lymph node dissection, and ovarian suspension (bilateral) were implemented under general anesthesia. On May 12, 2014, the postoperative pathological result showed cervical invasive moderately differentiated squamous cell carcinoma with the tumor-infiltrating in nearly all layers of the cervical wall, and tumor cell metastasizing to the lymph node of the left inner iliac (1/1). On June 13, 2014, she was given three-dimensional intensity-modulated radiotherapy in the external pelvic area (DT 4500cGY/25 times) and in abdominal enlarged lymph nodes of the external pelvic area (DT6000cGY/25 times), followed by three cycles of chemotherapy (paclitaxel combined with cisplatin). No obvious signs of recurrence or metastasis were found in regular review after discharge. Reexamination of CT on March 27, 2019, revealed progression in the lower lobes of the lung. A lung biopsy was performed by an outside provider under the guidance of CT, and the pathological findings (**Figure 1A**) were consistent with the previous conclusion of squamous cell carcinoma. Thus, the practitioners considered the possibility of metastatic cervical cancer. The immunohistochemistry (IHC) results suggested P40 (-), Ki-67 (90%+), TTF1 (-), CK7 (-), and P16 (-). Next, she was admitted to Hunan Cancer hospital on April 1, 2019. CT results showed the occupancy of the right lower lung dorsal segment and no obvious abnormality in the abdominal and pelvic cavity. We considered the possibility of primary lung cancer after evaluating the pathological results and related inspections. To further clarify the primary tumor, the 90-gene expression assay was performed on the biopsy tissue of the right lung tumor, and the test results (**Figure 1B**) indicated that the right lung mass was primary lung cancer. Finally, the patient was diagnosed with primary bronchial lung cancer, right lower lobe squamous cell carcinoma and underwent video-assisted thoracic surgery (VATS) radical resection of the right lower lobe cancer under general anesthesia

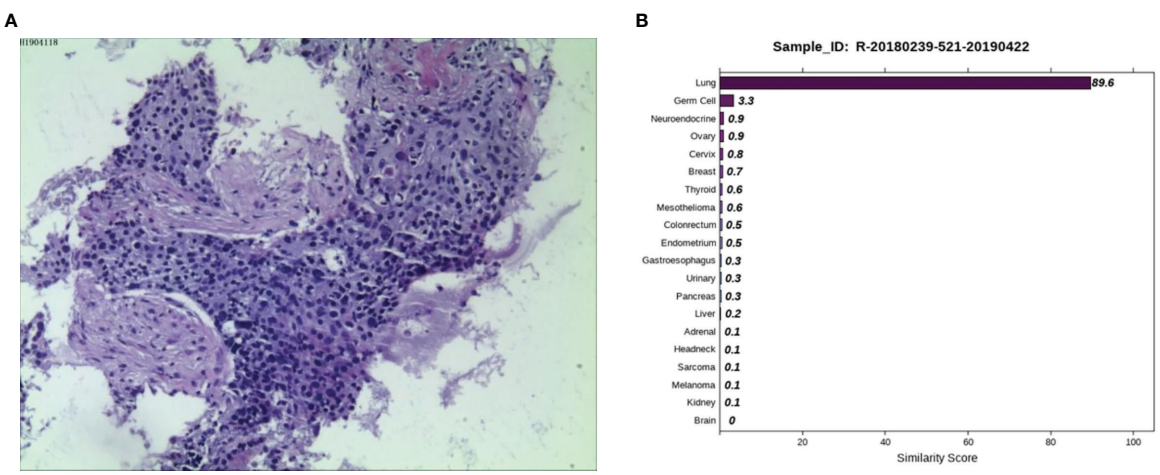


FIGURE 1 | Pathological and 90-gene expression assay results of case 1. **(A)** The pathological diagnosis of the resected specimens was squamous cell carcinoma. Immunohistochemistry staining: P40 (-), Ki67 (90%+), TTF1 (-), CK7 (-), P16 (-). **(B)** The 90-gene expression assay showed that the similarity score of the right lung mass coming from the lung tissue is 89.6.

on April 30, 2019 (pT2aN0M0, stage IB). The patient is in a disease-free state after regular reexamination so far without further adjuvant chemotherapy.

Case 2: A 59-year-old male appeared with lumbosacral pain accompanied by obvious traction pain of the right lower extremity, which was aggravated after sitting and activity, without symptoms such as coughing, sputum, chest tightness, night sweats, or fatigue. Positron emission tomography/computer tomography (PET/CT) performed in a local hospital showed small nodules in the posterior right upper lung with low glucose metabolism and multiple bone destruction throughout the body with increased glucose metabolism. Pathological examination of high-metabolism sites showed that the spindle cells had proliferated diffusely, the epithelioid cells were aberrant and distributed in a sheet, and necrosis and sequestrum formation were visible. Next, he was admitted to Hunan Cancer Hospital for further diagnosis and treatment. After examining the pathological section (**Figure 2A**), we considered it to be a poorly differentiated metastatic malignant tumor. Epithelioid differentiated sarcoma was not excluded because this sarcoma can appear in the form of sarcomatoid carcinoma with transparent cell cytoplasm. The IHC results showed CK (+), LCK (+), P40 (-), P63 (-), Hep (+, focally), ERG (-), CD31 (-), Syn (-), CgA (-), NapsinA (-), GATA3 (-), TTF1 (-), PSA (-), TG (-), and Urop3 nucleus (+). The enhanced CT scans of the neck, chest, abdomen, and pelvis showed multiple damaged areas of cervical, thoracic, lumbar vertebral, and right iliac bone, and no other malignant lesions in the lung, liver, kidney, and prostate. Brain magnetic resonance imaging (MRI) revealed the abnormal bone quality of the right frontal bone and C2 vertebra. To eliminate digestive tract tumors, gastroscopy and colonoscopy were performed and results showed no obvious abnormalities. To further identify the primary site, the 90-gene expression assay was performed on the biopsy tissue of the right iliac bone and the test results indicated that the right iliac bone had metastatic kidney cancer (**Figure 2B**). The patient was diagnosed with primary renal clear cell carcinoma (TxN0M1, stage IV). Then the patient returned to the local hospital for receiving palliative radiotherapy and chemotherapy.

Case 3: A 53-year-old male, appeared with numbness of the left upper limb without radiation pain and other discomforts such as dizziness, headache, nausea and vomiting. Lumbar MRI in a local hospital showed lumbar disc herniation. No therapeutic outcome was achieved in the patient after being firstly treated with unknown medication. The patient reported the worsened situation and then surgery was planned in a hospital in Hunan Province on May 7, 2019. The lumbar MRI before surgery showed a mass in the left front of S1, S2 vertebral with bone marrow edema, and the left nerve root was violated. They are highly suspicious to be tumors and a sacral puncture biopsy was performed on May 20, 2019. The histopathological findings (**Figure 3A**) demonstrated metastatic poorly differentiated carcinoma. The IHC results showed CK-pan (+), CK7 (-), CK20 (-), Villin (-), Ki-67 (70%+), P63 (+), P40 (+), Hepatocyte (-), Glypican-3 (-), PSA (-), PSAP (-), D2-40 (-), CR (-), and RCC (-). To find a primary tumor, PET/CT was performed in a hospital in Changsha city on May 30, 2019, found that the bone on the left side of S1 and S2 vertebrae was damaged with increased glucose metabolism, no obvious abnormality in the head, neck, chest and abdomen organs. The patient underwent colonoscopy, gastroscopy, and prostate color doppler ultrasound without any obvious abnormalities in our hospital on June 3, 2019, and the pelvic MRI revealed bone destruction and mass formation on the left side of the sacrum. Pathological diagnosis of the nasopharyngeal biopsy showed a large amount of inflammatory cell infiltrate in the mucosa, and atypical small squamous epithelial cells were seen in the interstitium. Epstein-Barr encoding region (EBER) *in-situ* hybridization showed negative results. To further clarify the primary tumor, the 90-gene expression assay was performed

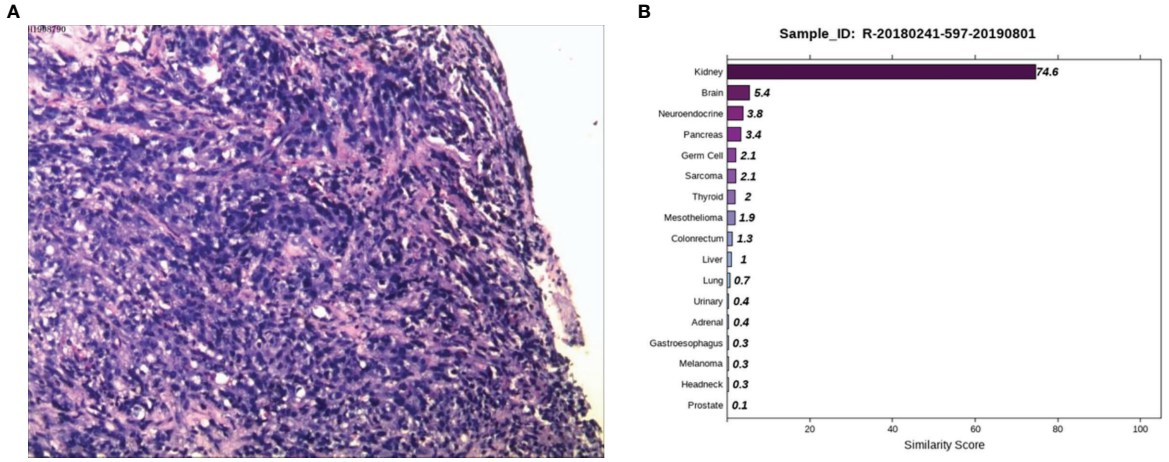


FIGURE 2 | Pathological and 90-gene expression assay results of case 2. **(A)** The pathological results showed diffuse proliferation of spindle cells, atypical epithelioid cells in a sheet-like distribution, and necrosis and bone formation were seen. Immunohistochemical staining confirmed CK, LCK, Hep (focally), Urop3 nucleus as positive. **(B)** The 90-gene expression assay indicated that there is a 74.6 similarity score of metastatic renal cancer on the right side of the bone.

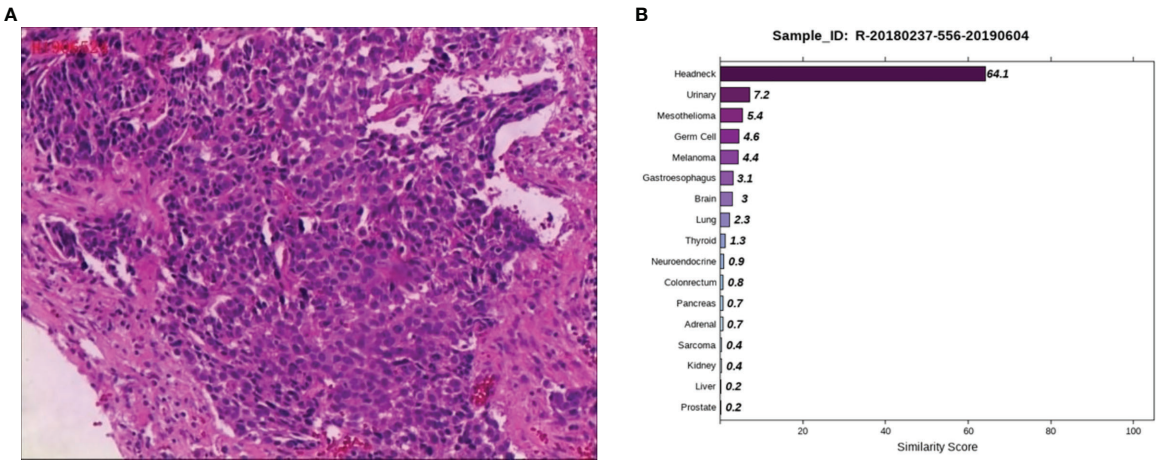


FIGURE 3 | Pathological and 90-gene expression assay results of case 3. **(A)** The pathological diagnosis of the resected specimens was metastatic poorly differentiated carcinoma. Immunohistochemistry staining: CK-pan, Ki-67, P63, P40 were positive respectively. **(B)** The 90-gene expression assay indicated that the similarity score of bone biopsy tissue is head and neck tumor metastasis was 64.1.

on the sacral biopsy tissue, and the test results indicated that the sacral biopsy tissue was head and neck tumor metastasis (**Figure 3B**). The patient's head and face related MRI showed abnormalities in bilateral tonsils but he refused to undergo surgical biopsy of bilateral tonsils. We considered the possibility that the primary tumor is tonsil cancer and give systemic palliative chemotherapy (albumin paclitaxel and carboplatin) combined with local radiotherapy. The patient is stable now and has survived for more than two years.

Case 4: A 50-year-old male presented with abdominal pain and discomfort without obvious inducement in May 2019. The pain was paroxysmal and dull, which accompanied by lower abdomen swelling. The pain was obvious when fasting, and it was relieved after eating. In June 2020, abdominal pain significantly worsened than before. CT examination at a local hospital showed multiple retroperitoneal lymphadenopathies, and B-ultrasonography showed a lymph node cyst in the left neck and supraclavicular fossa. A biopsy of the left cervical lymph node was performed on June 28, 2020. The postoperative examination revealed a malignant tumor in the left neck, and poorly differentiated squamous cell carcinoma was considered (**Figure 4A**). The patient visited Hunan Cancer Hospital on July 8, 2020 for further treatment. CT examination revealed multiple lymphadenopathies in the left supraclavicular area, mediastinum, hilum, retroperitoneum, and right iliac vessels. Head MRI examination showed no obvious occupying in the nasopharyngeal region. We consulted the local hospital regarding the pathological sections and combined IHC, and considered the diagnosis of malignant tumors, which tend to be poorly differentiated metastasis. Because IHC provides no clear indication, it is difficult to determine the tumor type. It is recommended to check the kidney, lung, liver, prostate, etc. in detail. IHC demonstrated a result of CK (+++), CK7 (-), CK20 (-), Villin (-), TTF1 (-), SATB-2 (+), MelanA (+/-), NapsinA (-), P40 (-), CR (-), CK5/6 (-), CDX-2 (-), S-100 (-), Des (-), SOX-10 (-), and HMB45 (-). PET-CT was performed to understand the general condition, results showed multiple swollen lymph nodes on the double supraclavicular, mediastinum, double lung hilum, right posterior of the right diaphragm, retroperitoneal area, and right iliac vessels, and abnormal radioactive concentration was seen in the corresponding positions, no obvious abnormality in the head, neck, chest and abdomen organs. To further clarify the source of the tumor, the 90-gene expression assay was performed, and the results suggested that the tumors on the left cervical lymph node maybe a metastasis from kidney tumors (**Figure 4B**). We then performed a renal MRI examination and found a nodule at the lower pole of the right kidney with a diameter of approximately 1.4 cm, considering the possibility of renal cancer. For further confirmation, a CT-guided kidney biopsy was performed on August 11, 2020. The cytological results showed cancer cells in the kidney puncture smear (**Figure 4C**), but no clear malignant tumor was found in the pathological section. Because the kidney mass was small (approximately 1 cm), the pathological diagnosis of renal biopsy still could not be further confirmed. Combined with the results of genetic testing and cytology, we considered the possibility of renal cancer, and the patient was treated with the targeted drug pazopanib. On September 18, 2020, the patient returned to the hospital for review. CT of the chest and abdomen showed a nodule with a density of 1.3×1.1 cm in the lower pole of the right kidney, which was smaller than before. At present, he is in stable condition and is undergoing immunotherapy.

DISCUSSION

Case 1 was a squamous cell carcinoma lesion in the lung found five years after surgery of cervical squamous cell carcinoma. The pathological test was not able to distinguish whether it was cervical cancer lung metastasis or second primary lung cancer. The 90-gene expression assay results suggested second primary cancer. Notably, treatment plans and prognoses differ completely

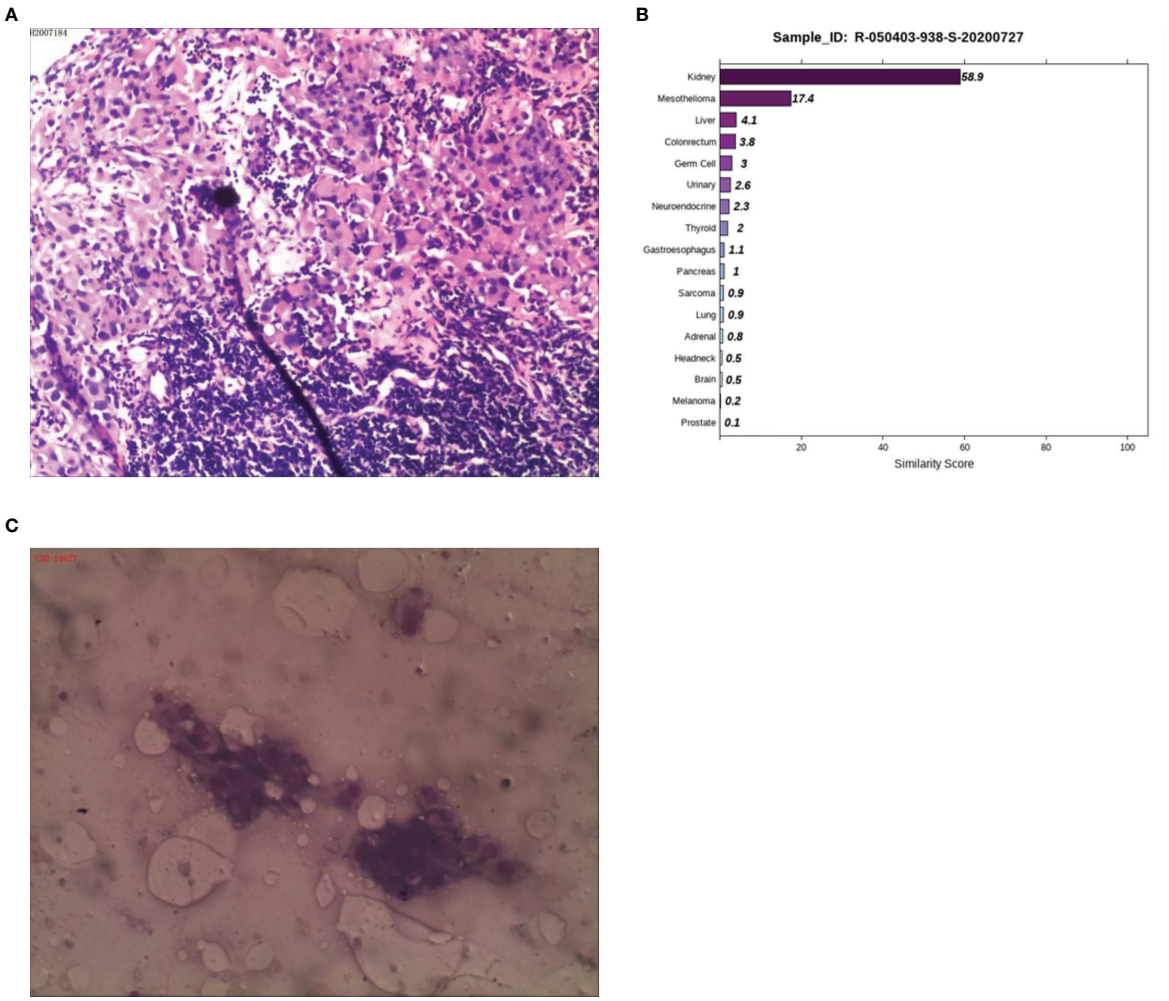


FIGURE 4 | Pathological, 90-gene expression assay and cytology results of case 4. **(A)** The pathological diagnosis of the resected specimens was malignant tumors which tend to be metastatic poorly differentiated. Immunohistochemistry staining: CK (+++), CK7 (-), CK20 (-), Villin (-), TTF1 (-), SATB-2 (+), MelanA (+/-), NapsinA (-), P40 (-), CR (-), CK5/6 (-), CDX-2 (-), S-100 (-), Des (-), SOX-10 (-), HMB45 (-). **(B)** The 90-gene expression assay indicated that the similarity score of metastatic renal cancer and liver cancer on the left cervical lymph node was 58.9 and 17.4, respectively. **(C)** The cytology results showed cancer cells in the kidney puncture smear.

between metastasis and second primary cancer. There is still the possibility of a cure through surgery if it is second primary cancer, while metastasis can only be treated with palliative care. The 90-gene expression assay helped distinguish second primary cancer from recurrent metastasis.

Case 2 involved bone metastases with an unknown primary site. There were no obvious lesions observed from the imaging examination. The 90-gene expression assay results suggested that the source of the tumor was the kidney. The 90-gene expression assay can facilitate the discovery of occult lesions and has guided the follow-up treatment.

Case 3 was a case of bone metastasis with no primary tumor. The imaging examination was normal. The 90-gene expression assay suggested that it was a head-and-neck tumor. Based on the 90-gene expression assay result, we performed an MRI of the head and face, with results suggesting the tumor site was tonsil. The 90-gene expression assay can quickly identify the primary site, thereby shortening the inspection time and reducing the opportunity cost of medical interventions.

Case 4 was a malignant tumor with multiple lymph node metastases; however, the immunohistochemistry results of left neck lymph node biopsy showed no clear source of tumor. The 90-gene expression assay results suggested that the source of the tumor was likely to be the kidney. Kidney biopsy cytology was performed to find cancer cells, but the pathological section could not identify malignant tumors. On August 1, 2020, the kidney MRI showed a nodule with a diameter of approximately 1.4 cm at the lower edge of the right kidney. After oral treatment with the kidney cancer-targeted drug pazopanib for a month, CT was performed, and the size of the right renal nodule was 1.3 x 1.1 cm, having diminished in size, suggesting that the treatment was effective. In the absence of clear and solid evidence from

traditional diagnostic methods, the 90-gene expression assay can help to quickly identify suspicious target lesions, enabling patients to receive early and accurate treatment.

A so-called "primary tumor of unknown origin" is diagnosed partly because the primary tumor is too small to be found from imaging. The possible reasons are as follows: 1) the primary focus grows slowly because of the body's strong immunity, 2) some cells are poorly differentiated, and distant metastasis appears at an early stage when the primary tumor is too small to be detected by traditional diagnostic methods, and 3) in a few patients, the primary tumor had been removed many years ago; the primary tumor disappeared, but distant metastasis appeared. Over the past decades, genomic profiles including gene expression profiling, DNA methylation and genomic alteration have been developed rapidly to identify tumor tissue of origin (6–8). The prognosis of a CUP patient largely depends on the biological characteristics of the primary tumor. Hainsworth et al. demonstrated that CUP patients who received a tumor type-specific therapy based on the results of gene expression profiling analysis had better median survival than CUP patients who received empiric therapy (12.5 *vs.* 9.1 months) (4). Therefore, clarifying the tissue origin of the tumor and adopting targeted treatment are of great significance for improving the prognosis of patients. Herein, the 90-gene expression assay can help elucidate the primary tumor source, and it may become a helpful molecular diagnostic method for CUP patients in the future.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from the relevant individual(s), and/or minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XP: Data analysis and participation in the writing of the manuscript. SY: Data analysis and interpretation and participation in the writing of the manuscript. YX: Data analysis and interpretation. BC: Data analysis and interpretation. QW: Data analysis and interpretation. QG: Collection and assembly of data. LW: Study design and direction and writing of the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Hunan Provincial Natural Science Foundation of China (No: 2019jj80018 to XP), Changsha Science and Technology Project of China (No: KJ1901073 to XP).

ACKNOWLEDGMENTS

We would like to thank the patients who enrolled in this project.

REFERENCES

1. Hasegawa H, Ando M, Yatabe Y, Mitani S, Honda K, Masuishi T, et al. Site-Specific Chemotherapy Based on Predicted Primary Site by Pathological Profile for Carcinoma of Unknown Primary Site. *Clin Oncol* (2018) 30:667–73. doi: 10.1016/j.clon.2018.06.012
2. Pavlidis N, Fizazi K. Cancer of Unknown Primary (CUP). *Crit Rev Oncol/Hematol* (2005) 54:243–50. doi: 10.1016/j.critrevonc.2004.10.002
3. Varadhachary GR, Raber MN. Cancer of Unknown Primary Site. *N Engl J Med* (2014) 371:757–65. doi: 10.1056/nejmra1303917
4. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, et al. Molecular Gene Expression Profiling to Predict the Tissue of Origin and Direct Site-Specific Therapy in Patients With Carcinoma of Unknown Primary Site: A Prospective Trial of the Sarah Cannon Research Institute. *J Clin Oncol* (2013) 31:217–23. doi: 10.1200/jco.2012.43.3755
5. Richardson A, Wagland R, Foster R, Symons J, Davis C, Boyland L, et al. Uncertainty and Anxiety in the Cancer of Unknown Primary Patient Journey: A Multiperspective Qualitative Study. *BMJ Support Palliat Care* (2013) 5:366–72. doi: 10.1136/bmjspcare-2013-000482
6. Ye Q, Wang Q, Qi P, Chen J, Sun Y, Jin S, et al. Development and Clinical Validation of a 90-Gene Expression Assay for Identifying Tumor Tissue Origin. *J Mol Diagn* (2020) 22:1139–50. doi: 10.1016/j.jmoldx.2020.06.005
7. Moran S, Martinez-Cardús A, Sayols S, Musulén E, Balañá C, Estival-Gonzalez A, et al. Epigenetic Profiling to Classify Cancer of Unknown Primary: A Multicentre, Retrospective Analysis. *Lancet Oncol* (2016) 17(10):1386–95. doi: 10.1016/S1470-2045(16)30297-2
8. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, et al. Site-Specific and Targeted Therapy Based on Molecular Profiling by Next-Generation Sequencing for Cancer of Unknown Primary Site: A Nonrandomized Phase 2 Clinical Trial. *JAMA Oncol* (2020) 6(12):1931–8. doi: 10.1001/jamaoncol.2020.4643

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

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

frontiers in Cell and Developmental Biology

CASE REPORT
published: 10 January 2022
doi: 10.3389/fcell.2021.755135

Case Report: Partial Response Following Nivolumab Plus Docetaxel in a Patient With *EGFR* Exon 20 Deletion/Insertion (p.N771delinsGF) Mutant Lung Adenocarcinoma Transdifferentiated From Squamous Cell Carcinoma

Lingling Zhu[1†], Yanyang Liu[1†], Honglin Gao[2], Jiewei Liu[1], Qinghua Zhou[1] and Feng Luo[1*]*

[1]*Lung Cancer Center, West China School of Medicine, West China Hospital of Sichuan University, Chengdu, China,* [2]*Key Laboratory of Drug Targeting and Drug Delivery System of the Education Ministry, Sichuan Engineering Laboratory for Plant-Sourced Drug, Sichuan Research Center for Drug Precision Industrial Technology, College of Polymer Science and Engineering, West China School of Pharmacy, Sichuan University, Chengdu, China*

The histological transformation from lung squamous cell carcinoma (LUSC) to lung adenocarcinoma (LUAD) and p. N771delinsGF mutations in *EGFR* exon 20 (ex20) are exceedingly rare in non–small cell lung carcinoma (NSCLC). *EGFR* ex20 mutations are insensitive to EGFR tyrosine kinase inhibitors in NSCLC. Here, we present a 76-year-old male smoker harboring LUAD with a novel p. N771delinsGF deletion/insertion mutation in *EGFR* ex20 transdifferentiating from advanced LUSC after chemoradiotherapy. The patient presented reduced hydrothorax and relieved tightness with the treatment of nivolumab plus docetaxel and carboplatin after the failure of second-line chemotherapy. The case highlights the importance of rebiopsy and molecular retesting after the progression of lung cancer and supports the idea that the combination of immune checkpoint blockade and chemotherapy may be an attractive option for patients with *EGFR* ex20 mutations associated with LUSC–LUAD transformation.

Keywords: non-small cell lung cancer, immune checkpoint blockade, EGFR exon 20 mutation, histological transformation, immune therapy

OPEN ACCESS

Edited by:
Qinghua Xu,
The Canhelp Genomics Research Center, China

Reviewed by:
Luis Corrales,
Cancer Research and Management Center (CIMCA), Costa Rica
Hushan Zhang,
Fudan University, China

****Correspondence:***
Qinghua Zhou
prof_qh_zhou@126.com
Feng Luo
luofeng66666@sina.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Cell and Developmental Biology

Received: *08 August 2021*
Accepted: *03 December 2021*
Published: *10 January 2022*

Citation:
Zhu L, Liu Y, Gao H, Liu J, Zhou Q and Luo F (2022) Case Report: Partial Response Following Nivolumab Plus Docetaxel in a Patient With EGFR *Exon 20 Deletion/Insertion (p.N771delinsGF) Mutant Lung Adenocarcinoma Transdifferentiated From Squamous Cell Carcinoma. Front. Cell Dev. Biol. 9:755135. doi: 10.3389/fcell.2021.755135*

INTRODUCTION

Lung cancer has the highest lethality rate among all cancers. Non–small cell lung carcinoma (NSCLC) is a major subtype of lung cancer, accounting for 85% of lung cancer cases, including lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and large cell carcinoma histologic subtypes. LUAD and LUSC harbor different features (Krause et al., 2020; Wang et al., 2020). Accumulating evidence from animal experiments and clinical observations indicates phenotypic plasticity in lung cancer cells, such as transdifferentiation of LUAD to LUSC or to small cell lung cancer. This phenotypic transition is a novel cellular mechanism for drug resistance in chemotherapy and in *EGFR*-targeted therapy (Hou et al., 2017). However, information regarding the transformation from LUSC to LUAD remains limited.

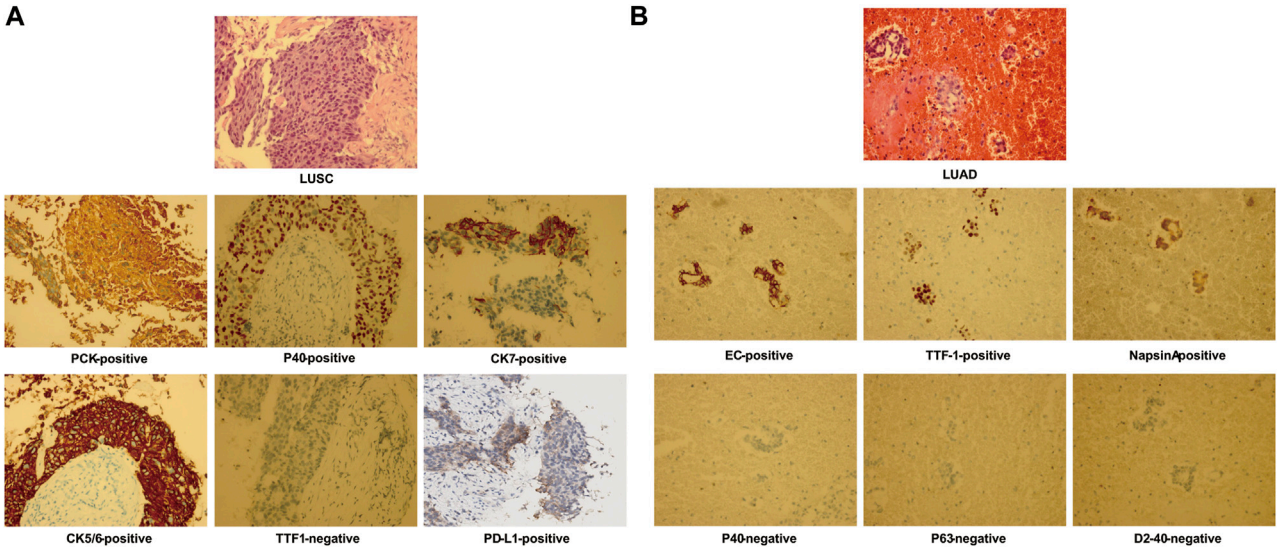


FIGURE 1 | Hematoxylin–eosin staining and immunohistochemistry staining from pulmonary biopsy shows lung squamous cell carcinoma (LUSC) **(A)** and of pleural fluid–exfoliated cells shows lung adenocarcinoma (LUAD) **(B)**.

Studies have shown that approximately 12% of patients with *EGFR*-mutant NSCLC harbor insertion mutations in exon 20 (ex20ins), which are insensitive to first- and second-generation EGFR tyrosine kinase inhibitors (TKIs) (Riess et al., 2018). LUAD and LUSC of adenosquamous carcinomas (ASCs) also share a monoclonal origin, with *EGFR* mutations being the most common oncogenic driver in the Asian population (Lin et al., 2020). However, no information is available regarding the relationship between *EGFR* ex20ins mutations and the molecular mechanisms of driving LUSC transdifferentiating to LUAD.

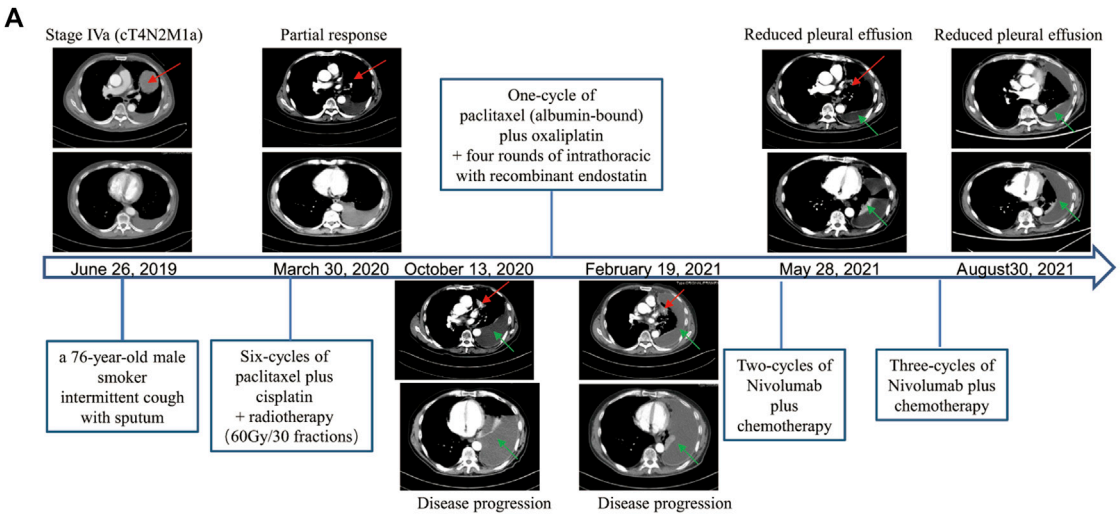
Here, we describe a case of a patient with LUAD harboring a novel p. N771delinsGF deletion/insertion (delins) mutation in *EGFR* ex20 who experienced transdifferentiation from LUSC after failure of chemotherapy and subsequently responded favorably to immune checkpoint blockade (ICB) plus chemotherapy.

CASE PRESENTATION

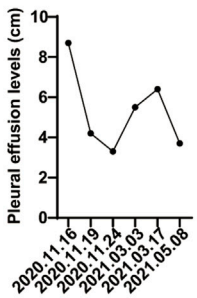
In June 2019, a 76-year-old male smoker weighing 50 kg with no notable medical, family, or psychosocial history was admitted to the hospital with a complaint of intermittent cough with sputum for >1 month. Chest computed tomography (CT) revealed a left upper lobe lung mass with multiple metastases involving the left pleura, left pleural effusion, and left mediastinal and hilar lymph node enlargement. Lung fine-needle aspiration biopsy confirmed the diagnosis of low-differentiated LUSC at stage IVA (cT4N2M1a) (**Figures 1A**, **2A**). The immunohistochemistry of pulmonary biopsy samples was positive for PCK, P40, CK7, CK5/6, and PD-L1 (40%) and was negative for TTF1. Genetic abnormalities such as amplification of *CCDN1*, *CCDN2*, *FGF19*, *FGF3*, *FGF4*, *FGFR1*, *KRAS*, and *PIK3CA*; ex2 missense mutation of *CDKN2A*; and ex5 code-shift mutation of *TP53* were observed in the primary lung cancer lesion by using next-generation sequencing (NGS; 520-gene panel, Oncoscreen plus, Burning Rock Biotech, Guangzhou, China). Plasma NGS demonstrated amplification of *CCDN1*, *FGF19*, *FGF3*, and *FGF4*; ex2 missense mutation of *CDKN2A*; ex5 shift mutation of *TP53*; and 11.1 mutations/Mb (**Table 1**).

From July to December 2019, the patient was treated with six cycles of paclitaxel plus cisplatin chemotherapy and subsequent radiotherapy (60 Gy/30 fractions at 2 Gy per day). He experienced an initial partial response with a decrease in tumor size (from 8.5 × 9.6 × 10.4 cm to unmeasurable). However, he was readmitted to our hospital with breathlessness and tightness of the chest. Chest CT in October 2020 showed that the amount of pleural effusion in his left chest had notably increased. LUAD cells (EC-positive, TTF-1–positive, napsin A–positive, P40-negative, P63-negative, D2-40–negative) were found by cytological examination of pleural effusion (**Figure 1B**). Unfortunately, the patient showed poor compliance resulting from the side effects of kidney function and gastrointestinal symptoms of cisplatin during first-line chemotherapy. We have demonstrated the safety and efficacy of the combination of paclitaxel plus oxaliplatin in pretreated advanced NSCLC (Xie et al., 2004); oxaliplatin was administered in the second-line treatment instead of cisplatin.

However, after one cycle of second-line treatment (albumin-bound paclitaxel plus oxaliplatin) and four rounds of recombinant endostatin (an antiangiogenesis agent), the disease continued to progress. Chest CT and B-mode ultrasound revealed an increase in the amount of pleural effusion (from 4.2 cm in November 2020 to 6.4 cm in February 2021; **Figures 2A, B**). A novel p. N771delinsGF indel mutation in *EGFR* ex20, a p. E17k mutation in *AKT1* ex3 and 6.72 mutations/Mb were identified in the pleural fluid



B



C

Time	Therapeutic schedule	Curative effect	Progression of lung lesion or pleural effusion
Jun. 2019 to Dec. 2019	(Paclitaxel + Cisplatin) * 4 cycles and subsequent chest radiotherapy (60Gy/ 30 fractions)	Partial response	Decrease of pleural effusion
Nov.2020 to Feb. 2021	Thoracic perfusion of recombinant endostatin * 4 cycle and (Albicin paclitaxel + Oxaliplatin)* 1 cycle	Disease progression	Increase of pleural effusion
Mar. 2021 to Apr. 2021	Nivoluma plus Docetaxel + Carboplatin * 1 cycle and Thoracic perfusion of carboplatin * 2 cycles and Thoracic perfusion of recombinant endostatin * 5 cycles	Stable	Decrease of pleural effusion
May. 2021 to Aug. 2021	Nivoluma plus Docetaxel + Carboplatin * 5 cycles	Stable	Well-controlled pleural effusion

FIGURE 2 | Summary of treatment schedule and associated therapeutic effects. **(A)** Computed tomography (CT) from June 26, 2019, shows a lesion in the left lung and enlarged left mediastinal and hilar lymph nodes (red arrows). CT from March 30, 2020, shows partial response after six cycles of paclitaxel plus cisplatin chemotherapy and subsequent radiotherapy (red arrows). CT from October 13, 2020 shows progressive spread of the tumor within the chest associated with left pleural effusion (green arrows). CT from February 19, 2021, shows increased pleural effusion after one cycle of second-line chemotherapy (green arrows). CT from May 28, 2021, to August 30, 2021, shows well-controlled pleural effusion after five courses of nivolumab plus chemotherapy and four rounds of intrathoracic chemotherapy with recombinant endostatin followed by two rounds of that with recombinant endostatin plus carboplatin (green arrows). **(B)** Pleural effusion levels before and after treatment with nivolumab plus chemotherapy by B-mode ultrasound. **(C)** Treatment course and results.

ctDNA by a hybrid-capture NGS method using a 1,021-gene panel (**Table 1**).

Five cycles of nivolumab plus docetaxel and carboplatin therapy as third-line therapy were initiated on March 19, 2021. The amount of pleural effusion then decreased (from 6.4 cm in March to 3.7 cm in May by B-mode ultrasound, **Figure 2B**) with well-controlled symptoms of breathlessness and chest tightness, and the lung lesion appeared stable. Moreover, an improved Eastern Cooperative Oncology Group performance status score (from 2 to 0) was observed in the patient after treatment with nivolumab plus chemotherapy. Therapeutic response was stable until the last follow-up on August 31, 2021, according to the well-controlled and unmeasurable lung lesion, decreased pleural effusion tested by chest CT, and well-controlled symptoms (**Figure 2A**). During the whole process of treatment with immunochemotherapy, the patient experienced no severe side effects.

DISCUSSION

We report a patient who was initially diagnosed with LUSC and was then found to have LUAD cells in the pleural effusion after chemotherapy resistance. One explanation for this is that the LUAD cells may have arisen from an original LUAD component of ASCs in our case. However, the patient had no shared mutations between the two genomic profiles of LUAD and LUSC detected by NGS. This is not in line with previous studies suggesting that LUAD and LUSC of ASCs share a monoclonal origin (Lin et al., 2020). Such a case may present more like two different pathological types. Recently, emerging data have provided convincing evidence supporting a phenotypic transition in lung cancer, such as adenocarcinoma–to–squamous cell transdifferentiation (AST) (Hou et al., 2017). Animal studies suggest that the mechanisms of AST are associated with *Kras* activation and *Lkb1* deletion. Moreover, AST is observed during

TABLE 1 | Next-generation sequencing results of pulmonary biopsy and plasma before treatments and of pleural fluid–exfoliated cells after chemoradiotherapy.

Mutant genes	Primary lung cancer before treatment (lung squamous cell carcinoma)			Plasma before treatment (lung squamous cell carcinoma)			Pleural effusion ctDNA after chemo-radiotherapy (lung adenocarcinoma)		
	Mutation type	Mutation location	Mutation abundance	Mutation type	Mutation location	Mutation abundance	Mutation type	Mutation location	Mutation abundance
CCDN1	Amplification	11q13.3	CN: 8.9	Amplification	11q13.3	CN: 3.1	—	—	—
CCDN2	Amplification	12p13.32	CN: 3.6	—	—	—	—	—	—
FGF19	Amplification	11q13.3	CN: 8.3	Amplification	11q13.3	CN: 3.3	—	—	—
FGF3	Amplification	11q13.3	CN: 7.7	Amplification	11q13.3	CN: 3.2	—	—	—
FGF4	Amplification	11q13.3	CN: 7.6	Amplification	11q13.3	CN: 2.9	—	—	—
FGFR1	Amplification	8p11.22	CN: 4.9	—	—	—	—	—	—
KRAS	Amplification	12p12.1	CN: 4.0	—	—	—	—	—	—
PIK3CA	Amplification	3q26.32	CN: 3.9	—	—	—	—	—	—
CDKN2A	Exon 2 missense mutation	c.242c>T p.Pro81Leu	65.05%	Exon 2 missense mutation	c.242c>T p.Pro81Leu	6.14%	—	—	—
TP53	Exon 5 code-shift mutation	c.545_552del p.Cys182fs	53.07%	Exon 5 code-shift mutation	c.545_552del p.Cys182fs	7.19%	—	—	—
TMB	—	—	—	—	—	11.1 Muts/Mb	—	—	6.72 Muts/Mb
EGFR	—	—	—	—	—	—	p.N771delinsGF deletion insertion mutation	Exon 20	0.5%
AKT1	—	—	—	—	—	—	p.E17k	Exon 3	0.8%

the development of chemotherapy resistance, suggesting that the transdifferentiation of lung cancer represents a novel mechanism of drug resistance (Han et al., 2014). Thus, we suggest that there may have been a transdifferentiation from LUSC to LUAD which led to chemotherapy resistance in our case. Although this hypothesis needs to be further confirmed, the high plasticity of lung cancer indicates this possibility.

Notably, LUSC and LUAD in our report were diagnosed using bronchoscopy biopsy specimens and cytological examination of exfoliated cells in pleural effusion, respectively. This highlights the importance of repetitive and comprehensive sampling and molecular testing, including that of tissues (biopsy specimens or surgery), plasma, and body fluid, both before and after treatments. The overall genomic profiles of both LUAD and LUSC in our case were detected by NGS, which allows sequencing of a high number of nucleotides (Mosele et al., 2020) and facilitated identification of the novel *EGFR* ex20 p. N771delinsGF mutation in our case. However, it has been reported that each fluid type (blood, urine, or saliva) still retains its own characteristic signature based on its unique molecules (El-Mogy et al., 2018). For instance, cell-free DNA from body fluid supernatants has a higher detection rate and sensitivity for tumor-specific mutations than cell-free DNA from sedimented tumor cells and plasma from body fluid (Guo et al., 2019; Zhang et al., 2019). LUAD was detected by pleural effusion in our case, indicating that more information may be revealed from a pleural effusion sample compared with LUSC diagnosed by biopsy specimens and plasma before treatment.

LUAD with a novel p. N771delinsGF mutation occurring in *EGFR* ex20 was found in our case. The *EGFR* ex20ins mutations are diverse, accounting for approximately 12% of patients with EGFR mutations. The V769_D770insASV is the most common in *EGFR* ex20ins mutation (23.0%) (Tsiambas et al., 2016; Riess et al., 2018). Similar to other *EGFR* mutation subtypes, *EGFR* ex20in functions as an oncogenic driver. However, compared with those with *EGFR* ex19 deletion or ex21 L858R mutation, NSCLC patients with *EGFR* ex20ins mutations have reduced affinity for EGFR-TKIs. They experienced a shorter median survival than patients with EGFR-TKI–sensitive mutations (16.5 vs. 33.0 months) during EGFR-TKI therapy (Naidoo et al., 2015; Yang et al., 2020). Nowadays, targeted therapeutics such as amivantamab, mobocertinib, and poziotinib have shown initial efficacy in NSCLC patients with *EGFR* ex20ins (Köhler and Jänne, 2021); however, no targeted therapeutics have been approved for *EGFR* ex20ins-driven NSCLC by the US Food and Drug Administration. Currently, chemotherapy remains the suitable treatment for patients harboring these mutations in China (Naidoo et al., 2015). One study showed that metastatic NSCLC patients with *EGFR* ex20ins treated with platinum chemotherapy presented a more favorable overall survival (OS; median 20 vs. 12 months) compared with NSCLC without targetable alterations (Choudhury et al., 2021). In a single-institution retrospective study (n = 29) of NSCLC with *EGFR* ex20ins, the median progression-free survival (PFS) with platinum-based chemotherapy was 7.1 months (95% confidence interval [CI], 6.3–13.7), and the median OS was 3.2 years (95% CI, 1.92–not reached) (Shah et al., 2021). These data suggest that LUAD patients with EGFR 20ins mutation may respond to chemotherapy. However, the present patient's disease remained unresponsive to second-line chemotherapy agents alone in our case, indicating an urgent need for the development of new therapeutic strategies.

Immunotherapy, such as ICB, has changed the treatment landscape in metastatic and recurrent NSCLC. Chemotherapy plus immunotherapy has shown particularly promising efficacy

in advanced NSCLC patients (Leonetti et al., 2019). Although compared with *EGFR* wide-type LUAD, cases of NSCLC with *EGFR* 19del or 21exon L858R mutations have a lower clinical response rate to ICB combined with chemotherapy. NSCLC patients with *EGFR* ex20ins mutation achieved both better PFS and OS than NSCLC patients with no targetable oncogenes after ICB and chemotherapy treatment (Remon et al., 2020). Therefore, chemotherapy in combination with nivolumab was administered to our patient, resulting in a significant reduction of pleural effusion and relieved dyspnea for more than 5 months. Likewise, in a retrospective single-center case series of patients with NSCLC and *EGFR* ex20ins mutation, two patients received platinum-doublet chemotherapy in combination with pembrolizumab, and both had stable disease (Shah et al., 2021). Thus, according to these reports and our case, ICB combined with chemotherapy can be considered an effective therapeutic strategy in NSCLC patients with *EGFR* ex20 p. N771delinsGF mutation. However, there is a lack of clinical trials to validate this hypothesis, and future clinical trials could be conducted to clarify the efficacy of ICB in combination with chemotherapy in NSCLC patients harboring *EGFR* ex20ins mutation.

In conclusion, our case suggests that phenotypic transition between LUSC and LUAD may occur in lung cancer during treatment. When lung cancer progresses, rebiopsy is helpful in accurately understanding and treating the disease. Moreover, NSCLC with an *EGFR* ex20 p. N771delinsGF mutation can benefit from nivolumab plus docetaxel and carboplatin. ICB plus chemotherapy may be a favorable strategy for NSCLC with *EGFR* ex20 mutation.

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 West China Hospital of Sichuan University. 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

LZ, YL, HG, and JL collected the clinical information, diagnostic information, therapeutic information, and images of the patients. LZ wrote the manuscript. LZ and YL identified the case and submitted the manuscript. FL, QZ, and HG revised the manuscript. FL and QZ proofread the manuscript. LZ and YL contributed equally to this work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Wu Jieping Medical Foundation, China (312150082), Post-Doctor Research Project of West China Hospital and Sichuan University (LZ).

ACKNOWLEDGMENTS

We sincerely appreciate Yan Wang, Ting Cao, and Hongying Chen from the Core Facility, West China Hospital of Sichuan University for their assistance and guidance. We also thank Yang Yang and Xiaoting Chen from the Animal Experimental Center of West China Hospital for their suggestions.

REFERENCES

Choudhury, N. J., Schoenfeld, A. J., Flynn, J., Falcon, C. J., Rizvi, H., Rudin, C. M., et al. (2021). Response to Standard Therapies and Comprehensive Genomic Analysis for Patients with Lung Adenocarcinoma with EGFR Exon 20 Insertions. *Clin. Cancer Res.* 27 (10), 2920–2927. doi:10.1158/1078-0432.CCR-20-4650

El-Mogy, M., Lam, B., Haj-Ahmad, T. A., McGowan, S., Yu, D., Nosal, L., et al. (2018). Diversity and Signature of Small RNA in Different Bodily Fluids Using Next Generation Sequencing. *BMC Genomics* 19 (1), 408. doi:10.1186/s12864-018-4785-8

Guo, Z., Xie, Z., Shi, H., Du, W., Peng, L., Han, W., et al. (2019). Malignant Pleural Effusion Supernatant Is an Alternative Liquid Biopsy Specimen for Comprehensive Mutational Profiling. *Thorac. Cancer* 10 (4), 823–831. doi:10.1111/1759-7714.13006

Han, X., Li, F., Fang, Z., Gao, Y., Li, F., Fang, R., et al. (2014). Transdifferentiation of Lung Adenocarcinoma in Mice with Lkb1 Deficiency to Squamous Cell Carcinoma. *Nat. Commun.* 5, 3261. doi:10.1038/ncomms4261

Hou, S., Zhou, S., Qin, Z., Yang, L., Han, X., Yao, S., et al. (2017). Evidence, Mechanism, and Clinical Relevance of the Transdifferentiation from Lung Adenocarcinoma to Squamous Cell Carcinoma. *Am. J. Pathol.* 187 (5), 954–962. doi:10.1016/j.ajpath.2017.01.009

Köhler, J., and Jänne, P. A. (2021). Amivantamab: Treating EGFR Exon 20-Mutant Cancers with Bispecific Antibody-Mediated Receptor Degradation. *Jco* 39 (30), 3403–3406. doi:10.1200/JCO.21.01494

Krause, A., Roma, L., Lorber, T., Habicht, J., Lardinois, D., De Filippo, M. R., et al. (2020). Deciphering the Clonal Relationship between Glandular and Squamous Components in Adenosquamous Carcinoma of the Lung Using Whole Exome Sequencing. *Lung Cancer* 150, 132–138. doi:10.1016/j.lungcan.2020.10.013

Leonetti, A., Wever, B., Mazzaschi, G., Assaraf, Y. G., Rolfo, C., Quaini, F., et al. (2019). Molecular Basis and Rationale for Combining Immune Checkpoint Inhibitors with Chemotherapy in Non-small Cell Lung Cancer. *Drug Resist. Updates* 46, 100644. doi:10.1016/j.drup.2019.100644

Lin, G., Li, C., Li, P. S., Fang, W. Z., Xu, H. P., Gong, Y. H., et al. (2020). Genomic Origin and EGFR-TKI Treatments of Pulmonary Adenosquamous Carcinoma. *Ann. Oncol.* 31 (4), 517–524. doi:10.1016/j.annonc.2020.01.014

Mosele, F., Remon, J., Mateo, J., Westphalen, C. B., Barlesi, F., Lolkema, M. P., et al. (2020). Recommendations for the Use of Next-Generation Sequencing (NGS) for Patients with Metastatic Cancers: a Report from the ESMO Precision Medicine Working Group. *Ann. Oncol.* 31 (11), 1491–1505. doi:10.1016/j.annonc.2020.07.014

Naidoo, J., Sima, C. S., Rodriguez, K., Busby, N., Nafa, K., Ladanyi, M., et al. (2015). Epidermal Growth Factor Receptor Exon 20 Insertions in Advanced Lung Adenocarcinomas: Clinical Outcomes and Response to Erlotinib. *Cancer* 121 (18), 3212–3220. doi:10.1002/cncr.29493

Remon, J., Hendriks, L. E. L., Cardona, A. F., and Besse, B. (2020). EGFR Exon 20 Insertions in Advanced Non-small Cell Lung Cancer: A New History Begins. *Cancer Treat. Rev.* 90, 102105. doi:10.1016/j.ctrv.2020.102105

Riess, J. W., Gandara, D. R., Frampton, G. M., Madison, R., Peled, N., Bufill, J. A., et al. (2018). Diverse EGFR Exon 20 Insertions and Co-occurring Molecular Alterations Identified by Comprehensive Genomic Profiling of NSCLC. *J. Thorac. Oncol.* 13 (10), 1560–1568. doi:10.1016/j.jtho.2018.06.019

Shah, M. P., Aredo, J. V., Padda, S. K., Ramchandran, K. J., Wakelee, H. A., Das, M. S., et al. (2021). EGFR Exon 20 Insertion NSCLC and Response to Platinum-Based Chemotherapy. *Clin. Lung Cancer* 162, 140. doi:10.1016/j.cllc.2021.07.001

Tsiambas, E., Lefas, A. Y., Georgiannos, S. N., Ragos, V., Fotiades, P. P., Grapsa, D., et al. (2016). EGFR Gene Deregulation Mechanisms in Lung Adenocarcinoma: A Molecular Review. *Pathol. - Res. Pract.* 212 (8), 672–677. doi:10.1016/j.prp.2016.06.005

Wang, B.-Y., Huang, J.-Y., Chen, H.-C., Lin, C.-H., Lin, S.-H., Hung, W.-H., et al. (2020). The Comparison between Adenocarcinoma and Squamous Cell Carcinoma in Lung Cancer Patients. *J. Cancer Res. Clin. Oncol.* 146 (1), 43–52. doi:10.1007/s00432-019-03079-8

Xie, C. Y., Wu, S. X., and Zhang, P. (2004). Efficacy of Combined Paclitaxel and Oxaliplatin Therapy in Patients with Pretreated Advanced Non-small Cell Lung Cancer. *Ai Zheng* 23 (8), 947–950.

Yang, G., Li, J., Xu, H., Yang, Y., Yang, L., Xu, F., et al. (2020). EGFR Exon 20 Insertion Mutations in Chinese Advanced Non-small Cell Lung Cancer Patients: Molecular Heterogeneity and Treatment Outcome from Nationwide Real-World Study. *Lung Cancer* 145, 186–194. doi:10.1016/j.lungcan.2020.03.014

Zhang, P., Wu, X., Tang, M., Nie, X., and Li, L. (2019). Detection of EGFR Gene Mutation Status from Pleural Effusions and Other Body Fluid Specimens in Patients with Lung Adenocarcinoma. *Thorac. Cancer* 10 (12), 2218–2224. doi:10.1111/1759-7714.13201

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, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

CASE REPORT
published: 30 March 2022
doi: 10.3389/fonc.2022.723140

Case Report: Molecular Profiling Assists in the Diagnosis and Treatment of Cancer of Unknown Primary

Bo Yu[1,2†], Qifeng Wang[2,3†], Xin Liu[1,2], Silong Hu[2,4], Liangping Zhou[2,5], Qinghua Xu[6,7], Yifeng Sun[6], Xichun Hu[1,2], Zhiguo Luo[1,2] and Xiaowei Zhang[1,2*]*

[1] Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai, China, [2] Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, [3] Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China, [4] Department of Nuclear Medicine, Fudan University Shanghai Cancer Center, Shanghai, China, [5] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [6] Canhelp Genomics Research Center, Canhelp Genomics, Hangzhou, China, [7] Institute of Machine Learning and Systems Biology, College of Electronics and Information Engineering, Tongji University, Shanghai, China

OPEN ACCESS

Edited by:
Mantang Qiu, Peking University People's Hospital, China

Reviewed by:
Qing Chun Zhao, Shenyang Pharmaceutical University, China
Antonio Lopez-Beltran, Cordoba University, United States
Nazanin Hosseinkhan, Iran University of Medical Science, Iran

***Correspondence:**
Xiaowei Zhang
dongfangzhizizhxw@aliyun.com
Zhiguo Luo
luozhiguo88@aliyun.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

Received: *29 June 2021*
Accepted: *24 February 2022*
Published: *30 March 2022*

Citation:
Yu B, Wang Q, Liu X, Hu S, Zhou L, Xu Q, Sun Y, Hu X, Luo Z and Zhang X (2022) Case Report: Molecular Profiling Assists in the Diagnosis and Treatment of Cancer of Unknown Primary. Front. Oncol. 12:723140. doi: 10.3389/fonc.2022.723140

Background: For cancer of unknown primary (CUP), non-selective empiric chemotherapy is usually used. However, patients suffering from CUP are generally assumed to have a dismal prognosis with median overall survival of less than 1 year. Therefore, clinicians eagerly await the establishment of effective strategies for diagnosis and treatment. In recent years, the remarkable advances in next-generation sequencing (NGS) technology have enabled the wide usage of DNA/RNA sequencing to comprehensively analyze the molecular information of individual tumors and identify potential targets for patients' diagnosis and treatment. Here, we describe a patient of CUP who was successfully diagnosed and treated with targeted therapy directed by comprehensive molecular profiling.

Case Presentation: A 61-year-old Asian woman with a painless, slow-growing mass lesion in the mesosternum underwent fluorodeoxyglucose-positron emission tomography/computed tomography and was found to have malignant metastatic tumors in the mesosternum. Conventional pathological examination of metastatic lesions could not conclude the primary origin of the tumors. The patient was diagnosed with CUP at first. Then, comprehensive molecular profiling was employed to identify the tumor origin and genetic alterations. A gene expression-based tissue origin assay was performed using a tissue biopsy sample. The test result suggested that the lesion tumors might be breast cancer metastasis. Furthermore, liquid biopsy-based circulating tumor DNA profiling detected an *ERBB2* copy number amplification. Subsequent surgery and additional postoperative pathology analysis confirmed that the primary tumor site was indeed located in the right outer upper quadrant of the breast. After local surgical resection, the patient received 8 cycles of Docetaxel + Carboplatin + Trastuzumab + Pertuzumab (TCbHP) chemotherapy with subsequent

human epidermal growth factor receptor 2 (*HER2*)-targeted maintenance therapy. Currently, the patient is on regular follow-up and has achieved disease control for up to 6 months.

Conclusion: Our findings suggest that molecular identification of the tumor origin and the detection of actionable molecular alterations may offer promise for improved diagnostic accuracy and important therapeutic implications for patients with the CUP syndrome.

Keywords: gene expression profiling, next-generation sequencing, cancer of unknown primary, HER2 amplification, tumor of origin

INTRODUCTION

Cancer of unknown primary (CUP) is a tumor that has metastasized from one part of the body to another part of the body. The place where it began, also called the primary site, is unknown. These cases make up about 3%–5% of all malignancies. The disease is characterized by late clinical presentation, early metastases, and poor prognosis, and the chance of 1-year survival is about 10%–20% (1). CUP often is challenging because it tends to be aggressive and has often spread to many parts of the body before it is found. CUPs are often found in the lymph nodes, liver, lung, peritoneum (lining of the bowel), or bone (2). In addition, because the origin of cancer is unknown, as once it has spread, it is difficult to locate its origin and so treatment and patient's survival rate tend to be hindered. Thus, it may be more challenging to choose the best treatment (3).

With the advent of next-generation sequencing (NGS) techniques, massively parallel sequencing of tumor DNA offers great opportunity to identify actionable mutations and enables targeted therapy in oncologic practice (4–7). However, mutation profiling alone is not thought to be sufficient to guide the personalized treatment given that targeted therapies against a particular driver mutation can act differently in different tumor types. For example, BRAF V600E mutations are observed in cancers arising from numerous tissue sites, and the likelihood of response to BRAF inhibitors varies widely as a function of tumor type (8). Given the heterogeneous nature of the disease, the accurate diagnosis of primary tumor has an even greater role to optimize targeted therapy for patients with CUP.

Molecular tumor profiling has been under development over the last decade for predicting the tumor site of origins in patients with CUP (9, 10). Specific gene expression profiles have been well recognized in most cancers according to their site of origin, which reflects the different expression profiles present in their normal tissues of origin. Differences in gene expression thus allow distinction between various solid tumors and provide a valuable method for diagnosis of the tissue of origin in patients with CUP. Recently, a real-time PCR assay termed "the 90-gene expression assay" (Canhelp Genomics Co., Ltd., Hangzhou, China) was developed for the classification of 21 common tumor types based on gene expression profiling. Firstly, researchers established a pan-cancer transcriptome database with 5,434 specimens encompassing 21 tumor types to screen tumor-specific genes. Then, the Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm was used to select the top 10 most predictive genes for each of the 21 tumor types. After removing redundant genes, a list of 90 genes was obtained and used to develop an SVM linear model named "90-gene classifier" for primary tumor identification. The 90-gene classifier was used to calculate the similarity scores of each tumor type, which reflect how much the gene expression pattern of the test sample is similar to the global gene expression pattern of a known tumor type. The similarity scores varied from 0 to 100, with the aggregate of all 21 similarity scores equaling 100. The tumor type with the highest similarity score was considered as the predicted tumor type by the 90-gene expression assay. In a retrospective cohort of 609 clinical samples, the gene expression assay demonstrated an overall accuracy of 90.4% for primary tumors and 89.2% for metastatic tumors. Furthermore, in a real-life cohort of 141 CUP patients, the gene expression assay was able to provide instructive predictions of primary tumors in 82.3% of patients (116/141). These findings suggest that the 90-gene expression assay could efficiently predict the primary origin for a broad spectrum of tumor types and support its diagnostic utility of molecular classification in difficult-to-diagnose metastatic cancer (11).

Here, we report a case of CUP that successfully achieved precise diagnosis and personalized treatment with targeted therapy of the disease based on comprehensive genomic analysis of the tumor and discuss the potential of using genomic tests to improve the diagnosis and management of CUP patients.

CASE REPORT

A 61-year-old Asian woman, previously healthy, presented with a painless slow-growing mass lesion in the mesosternum. Her medical and family histories were unremarkable. Physical examination revealed no generalized lymphadenopathy. Chest computed tomography (CT) scan revealed space-occupying lesions in the mesosternum, a greater likelihood of malignancy, and multiple rounds or round-like nodules with variable sizes scattered throughout both lungs. Fluorodeoxyglucose (FDG)-positron

Abbreviations: HER2, human epidermal growth factor receptor 2; CUP, cancer of unknown primary; FDG, fluorodeoxyglucose; NGS, next-generation sequencing; CT, computed tomography; MRI, magnetic resonance imaging; FFPE, formalin-fixed, paraffin-embedded; CNV, copy number variation; IHC, immunohistochemistry; TNBC, triple-negative breast cancer; TMB, tumor mutational burden.

emission tomography/computed tomography (PET-CT) detected two areas of bone destruction with soft tissue density shadow formation in the mesosternum with FDG uptake, malignant lesions, and possible metastases. Besides, low-density lesions in the right medial lobe of the liver showed poorly defined boundaries and slightly increased FDG metabolism; malignant lesion metastases were also considered. Bilateral thyroid gland showed heterogeneous density, a slightly low-density nodule in the right lobe, and poorly defined boundaries and slightly increased FDG metabolism; the middle segment wall of the esophagus was slightly thickened with slightly increased FDG metabolism, thus a combined gastroscopy test was recommended. Further liver magnetic resonance imaging (MRI) showed focal signal abnormalities in hepatic segment 8, inflammatory lesions were considered and malignant lesions were to be excluded, and liver puncture should be checked if necessary, also multiple hepatic hemangiomas (**Figure 1**). The thyroid ultrasound test revealed bilateral thyroid lesions with formation of multiple nodules (Thyroid Imaging Reporting and Data System, TI-RADS). The gastroscopic examination showed non-atrophic gastritis with erosion, fundic gland polyps (the greater curvature), HP (-), and chronic inflammation of antrum mucosa (the lesser curvature). The colonoscopy showed that the surface of the colorectal mucosa was smooth, and no significant abnormality was observed. However, these imaging evaluations plus endoscopy could not identify the primary location of the tumor. Significant laboratory investigations showed elevated tumor markers [Carbohydrate antigen19-9 (CA19-9): 38.80 U/ml↑, CA50: 32.02 IU/ml↑, Carcinoembryonic antigen (CEA): 10.00 ng/ml↑, Cytokeratin fragment antigen 21-1 (CYFRA21-1): 4.08 ng/ml↑]. The endoscopic ultrasonography showed multiple nodules in the liver (etiology to be determined), no obvious endobronchial space-occupying lesions were found in the pancreatic parenchyma, and no obvious abnormalities were observed in the gallbladder. Pathological examination of tissue from an ultrasound-guided fine-needle aspiration biopsy of the mesosternum lesions suggested a poorly differentiated metastatic

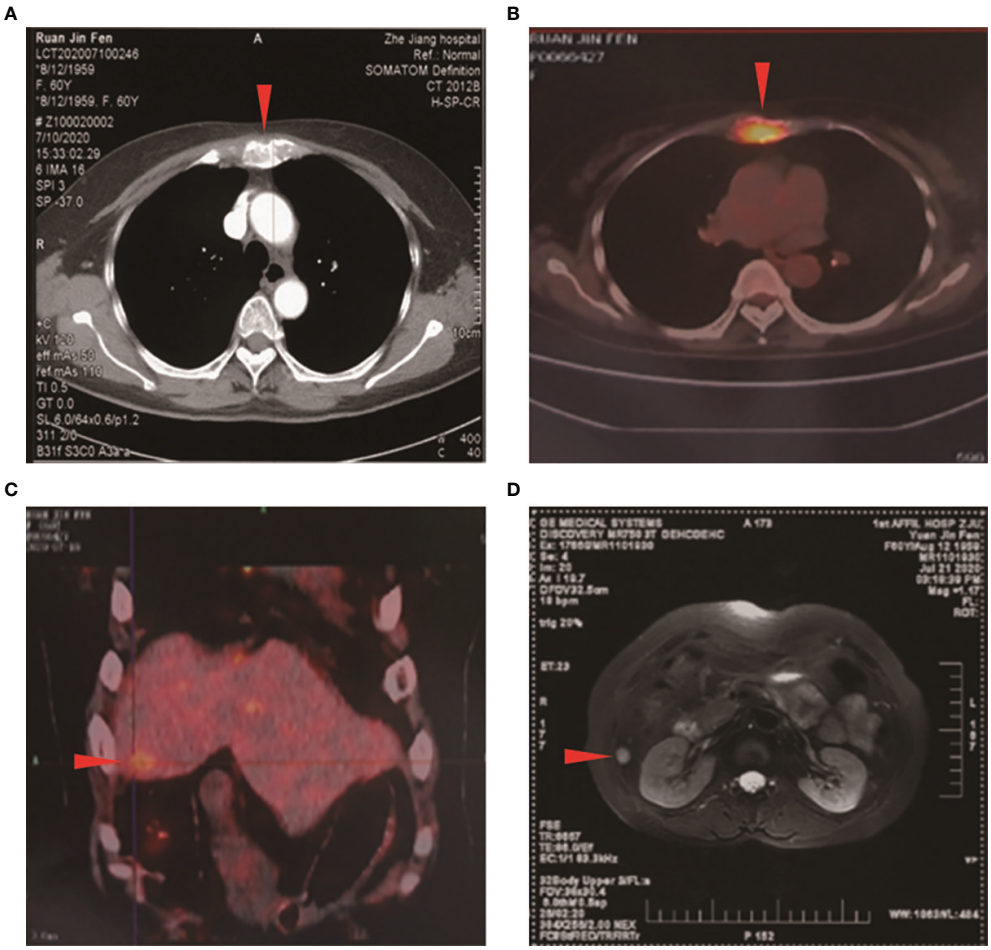


FIGURE 1 | Initial computed tomography (CT), positron emission tomography/computed tomography, and magnetic resonance imaging (MRI) scans. **(A)** CT scan showed a space-occupying lesion in the mesosternum. PET-CT demonstrated high fluorodeoxyglucose (FDG) uptake in the **(B)** mesosternum and in the **(C)** right medial lobe of the liver (arrowhead). **(D)** MRI showed focal signal abnormalities in hepatic segment 8 (arrowhead).

carcinoma, with a bias toward classical adenocarcinoma. Immunohistochemical (IHC) detection failed to determine the exact tissue origin. The mammography detection showed that there were focal calcifications in the upper outer quadrant of the breast. Ultrasound examinations revealed that the bilateral breast had lobular hyperplasia, with formation of several nodules Breast Imaging Reporting and Data System (BI-RADS) in the upper outer quadrant of the breast, and no axillary lymphadenopathy was detected. In addition, MRI revealed that the upper outer quadrant of the breast exhibited focal mass with non-mass-like enhancement (BI-RADS 3). There was sternal metastasis with adjacent soft tissue thickening. Because the primary site could not be identified, we decided to treat the patient with local radiation therapy (6MV-X, 50GY/10) to the metastatic lesion of the sternum first.

Comprehensive genomic analysis using the multiple-genes panel NGS that was designed to detect variants of cancer-related genes was performed. The genomic DNA was extracted from patient's peripheral blood. NGS was performed to an average depth of >5,000x on the Novaseq 6000 platform (Illumina, USA). This assay demonstrated the presence of an *ERBB2* copy number variation (CNV), amplification. This amplification is well known to be a positive biomarker for trastuzumab, a recombinant humanized monoclonal antibody that specifically targets the extracellular site of *human epidermal growth factor receptor 2* (*HER2*). The genomic analysis also identified additional *CCND1* amplification and multiple gene mutations, including *TP53* (37.59%) mutation, c.738G>A, p.M246I; *SPRED1* (29.20%) mutation, c.1150G>C, p.E384Q; *MLH3* (23.11%) mutation, c.616G>C, p.D206H; *LRP1B* (17.39%) mutation, c.5772C>G, p.I1924M; *ERBB2* (17.22%) mutation, c.2264T>C, p.L755S; *XRCC2* (15.41%) mutation, c.37G>A, p.E13K; *SETD2* (14.71%) mutation, c.5531C>G, p.S1844C; *NCOA3* (13.66%) mutation, c.2429C>T, p.S810F; *BRAF* (15.50%) mutation, c.1739A>T, p.N580I; *ERBB4* (11.02%) mutation, c.1573C>T, p.R525C; and *KMT2D* (0.89%) mutation, c.4843C>T, p.R1615. Tumor mutational burden (TMB): 7.9 mutations/Mb. Microsatellitestable (MSS), Programmed cell death-Ligand 1 (PD-L1) (-) (**Figures 2A, B**).

Furthermore, a 90-gene expression assay for identifying tumor tissue origin was performed based on a formalin-fixed paraffin-embedded (FFPE) biopsy sample of the patient's mesosternum

A

Mutation type	Gene	Exon	nucleic acid mutation	Amino acid mutation	abundance
Copy number variation	CCND1	CNV(Amplification)			5
	ERBB2	CNV(Amplification)			7
Single nucleotide variant or insertion/deletion	BRAF	exon14	c.1739A>T	p.N580I	12.50%
	ERBB2	exon19	c.2264T>C	p.L755S	17.22%
	ERBB4	exon13	c.1573C>T	p.R525C	11.02%
	KMT2D	exon19	c.4843C>T	p.R1615*	0.89%
	LRP1B	exon36	c.5772C>G	p.I1924M	17.39%
	MLH3	exon2	c.616G>C	p.D206H	23.11%
	NCOA3	exon13	c.2429C>T	p.S810F	13.66%
	SETD2	exon12	c.5531C>G	p.S1844C	14.71%
	SPRED1	exon7	c.1150G>C	p.E384Q	29.20%
	TP53	exon7	c.738G>A	p.M246I	37.59%
	XRCC2	exon1	c.37G>A	p.E13K	15.41%

B

Content	Result	Conclusion
TMB	TMB value: 7.9 mutations/Mb Reference value: 4.5 mutations/Mb	Immune checkpoint inhibitor therapy may be sensitive
MSI	MSS	Immune checkpoint inhibitor therapy may not be sensitive
PD-L1 expression	Negative	Immune checkpoint inhibitor therapy may not be sensitive

C

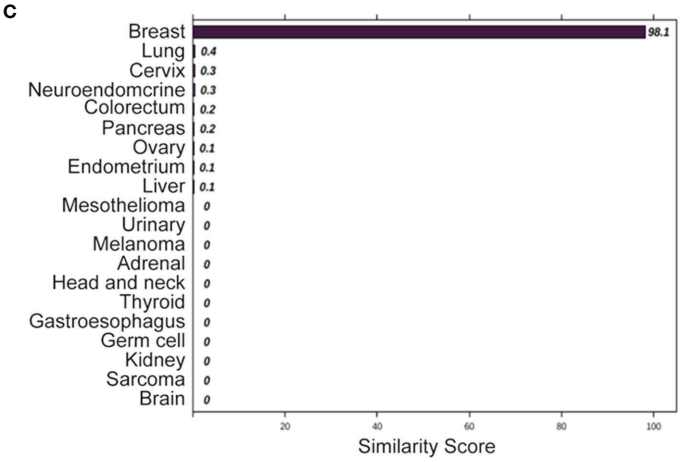


FIGURE 2 | The next-generation sequencing (NGS)-based multiplex assay and the 90-gene expression assay results. **(A, B)** The gene expression profiling was analyzed by the NGS-based multiplex assay. **(C)** The 90-gene expression assay result with one similarity score for each of the 20 tumor types based on the formalin-fixed, paraffin-embedded tissues collected from the lesion in the mesosternum. The top three predictions were breast (98.1), lung (0.4), and cervix (0.3). Therefore, the most likely site is breast (98.1).

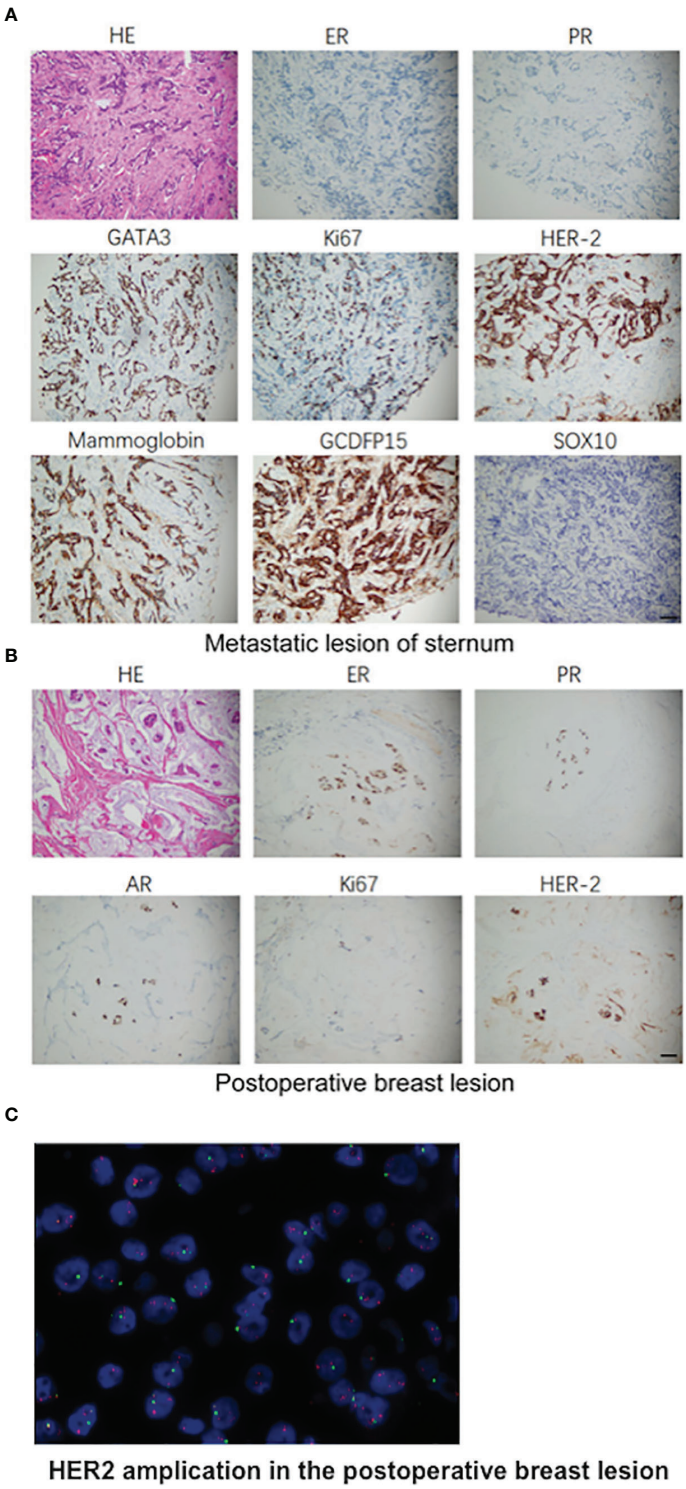


FIGURE 3 | Pathologic findings. **(A)** Histopathological examination of the mesosternum lesions and the immunohistochemical examinations of the cells stained for ER (-), PR (-), HER2 (3+), Ki-67 (+50%), Mammaglobin (+), GATA3 (+), GCDFP15 (+), SOX10 (-). **(B)** H&E staining of the surgical specimens of total regional adenomammectomy and the immunohistochemical examinations of the cells stained positive for ER (+, 90%), PR (+, 70%), HER2 (2+), Ki-67 (+, 10%), and AR (+, 10%). **(C)** Representative image of HER2 amplification in the postoperative breast lesion using FISH analysis. Scale bar, 50 μm.

lesions. As shown in **Figure 2C**, these test results strongly supported a breast cancer as the cell of origin with a score of 98.1. Then, the patient underwent careful examination after the tissue-of-origin test with attention paid to the breast region. Histopathological examination of the mesosternum lesions revealed a poorly differentiated metastatic adenocarcinoma; combined with immunohistochemical findings, the cells stained for Estrogen receptor (ER) (-), Progesterone receptor (PR) (-), HER2 (3+), Ki-67 (+50%), Mammaglobin (+), GATA3 (+), GCDFP15 (+), and SOX10 (-) and tend to be breast cancer metastasis (**Figure 3A**). These findings suggested adenocarcinoma of breast origin. MRI and mammography-based tumor localization were subsequently utilized to perform the calcification focal resection in the right outer upper quadrant of the breast. Postoperative pathology confirmed it as invasive ductal carcinoma of the right breast, grade II with mucin secretion, and foci of invasive papillary carcinoma, with 0.7 cm in maximum dimension. Immunohistochemical analysis indicated that the cells stained for ER (+, 90%), PR (+, 70%), HER2 (2+), Ki-67 (+, 10%), and Androgen receptor (AR) (+, 10%) (**Figure 3B**). Fluorescence in situ hybridisation (FISH) detection revealed HER2 gene amplification (**Figure 3C**). After a local surgical resection, the patient has received monthly zoledronic acid treatment and 8 cycles of TCbHP regimen (Docetaxel + Carboplatin + Herceptin + Pertuzumab), with subsequent HP (Herceptin + Pertuzumab) maintenance therapy intravenously at a standard dose every 3 weeks for up to 6 months. These regimens were well-tolerated. The patient was regularly reviewed at monthly intervals. Six months after the operation, she was found to have stable disease, with the obvious osteogenic changes in metastatic lesions of the sternum (**Figure 4A**), and no change in the solitary liver lesion, which further confirmed that atypical hemangioma may be involved (**Figure 4B**). Tumor markers (before treatment: CA19-9: 38.80 U/ml↑, CA50: 32.02 IU/ml↑, CEA: 10.00 ng/ml↑, CYFRA21-1: 4.08 ng/ml↑) all decreased to normal levels.

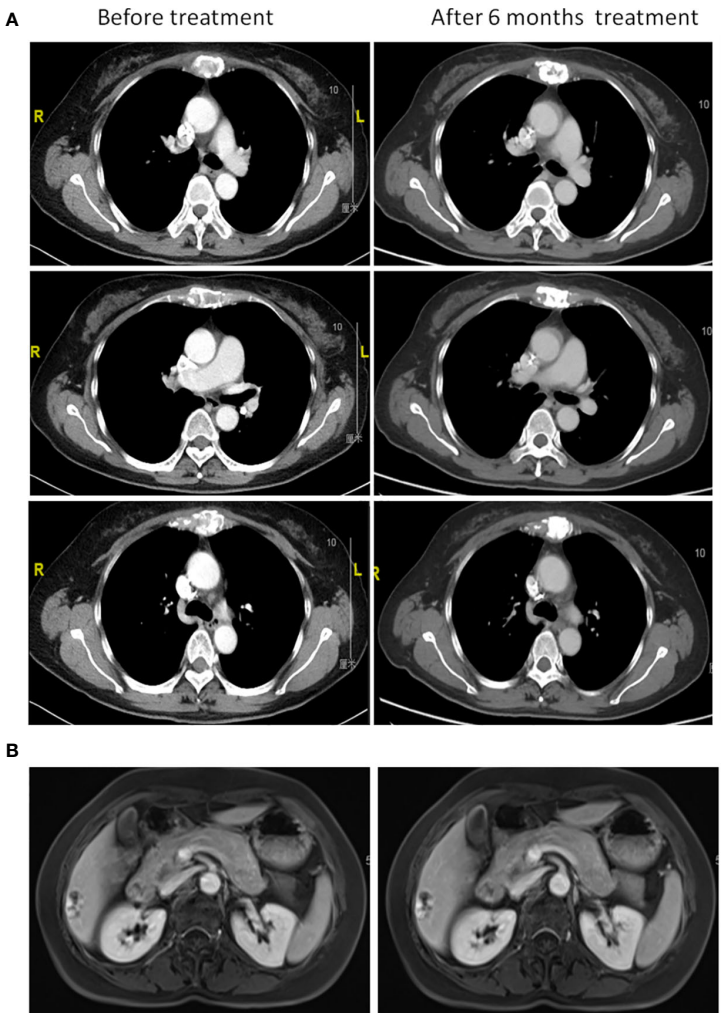


FIGURE 4 | Clinical course after initiation of standard chemotherapy. **(A)** Chest computed tomography and **(B)** liver magnetic resonance imaging before chemotherapy treatment and after 6 months.

DISCUSSION

CUP is defined as metastatic cancer in the absence of a clinically detectable anatomically defined primary tumor site after an adequate diagnostic evaluation. It constitutes 3%–5% of all newly diagnosed cancers per year worldwide. Including cancers of uncertain primary origin, the total number increases to 12%–15% of all newly diagnosed malignancies (12, 13). In the past decades, CUP patients commonly have poor prognoses due to treatment with a non-selective empirical therapy using cytotoxic agents (2). The era of precision medicine in oncology offers promise for improved diagnosis and better therapy for patients with the CUP syndrome. Gene expression-based tissue-of-origin molecular profiling has played an important role in the diagnostic armamentarium of CUP cancers. Besides, genomic characterization of CUP cancers using NGS techniques may also reveal actionable biomarkers for targeted therapies. Recently, research (14) has conducted a phase 2 clinical trial to evaluate the clinical benefit of comprehensive genomic profiling for CUP patients. Their findings demonstrate that site-specific treatment including molecularly targeted therapy based on profiling gene expression and gene alterations indeed contributes to treating patients with CUP.

Here, we present a patient who suffered from tumors with metastasis to the mesosternum, including a high likelihood of developing metastases in the liver, which usually indicates a poor prognosis (15). Her surgical specimen from mesosternum lesions first underwent conventional pathologic evaluations, which failed to accurately identify the primary site. Thus, we performed a novel 90-gene expression test based on an FFPE biopsy sample of the patient's mesosternum lesions. The test results strongly suggested that the tissue of origin is breast cancer. Interestingly, the patient then underwent careful examination after the tissue-of-origin test with attention paid to the breast region. The MRI and mammography-based tumor localization indeed identified the primary tumor in the right breast. In the present case, considering that the remaining tissue biopsy sample was limited, the liquid biopsy-based circulating tumor DNA (ctDNA) profiling was performed afterward. The ctDNA profiling successfully identified oncogenic amplifications (*ERBB2, CCND1*) and mutations (*TP53, BRAF, ERBB2, ERBB4,* etc.). The *ERBB2* amplification has been widely reported as a driver CNV, which is most commonly observed in breast cancer (16) and is relatively fewer in gastric cancer, colorectal cancer, etc. (17–19). In breast cancer, the *ERBB2* amplification has been reported as a strong predictor of efficacy for trastuzumab that interferes with the *HER2*. Patients whose tumors harbor *ERBB2* amplification display a remarkable response rate in prospective trials of an anti-*HER2* monoclonal antibody, including randomized phase III trials (20). Accordingly, the patient received 8 cycles of TCbHP chemotherapy with subsequent *HER2*-targeted maintenance therapy. Despite the poor characteristics at baseline, the patient has survived to date with a high quality of life after the precise diagnosis and targeted therapy directed by comprehensive genomic profiling. Although advanced technology including the 90-gene expression assay and ctDNA profiling could help guide personalized therapy, it also has some limitations. Firstly, suboptimal specimens, including limited tissues, samples with excess necrosis, or little tumor contents, are not suitable for the 90-gene expression assay. Second, both methods, the 90-gene expression assay and the NGS tests, are not yet available in most resource-limited centers.

In summary, we present a patient who was first diagnosed as CUP was later successfully clarified as breast carcinomas and treated with targeted therapy based on transcriptomic classification and actionable mutation analysis. Our results suggest that molecular identification of the primary site and genetic mutations in tumors as a guide to selecting appropriate targeted therapy can result in a durable response for CUP patients. Further randomized studies comparing this systematic approach with empirical chemotherapy for CUP patients are warranted.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Fudan University Shanghai Cancer Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BY, QFW, and XWZ analyzed and interpreted the patient data and were major contributors in writing the article. XL, SLH, LPZ, XCH, QHX, and YFS contributed to the acquisition and analysis of data. XWZ and ZGL contributed to the concept and revised the article. All authors read and approved the final article.

FUNDING

This work was financially supported by the Natural Science Foundation of China (Grant No. 81871948 and No. 81401986).

REFERENCES

1. Pavlidis N, Pentheroudakis G. Cancer of Unknown Primary Site. *Lancet (Lond Engl)* (2012) 379:1428–35. doi: 10.1016/S0140-6736(11)61178-1
2. Massard C, Loriot Y, Fizazi K. Carcinomas of an Unknown Primary Origin–Diagnosis and Treatment. *Nat Rev Clin Oncol* (2011) 8:701–10. doi: 10.1038/nrclinonc.2011.158
3. Hainsworth J, Greco F. Treatment of Patients With Cancer of an Unknown Primary Site. *N Engl J Med* (1993) 329:257–63. doi: 10.1056/NEJM199307223290407
4. Meric-Bernstam F, Brusco L, Shaw K, Horombe C, Kopetz S, Davies M, et al. Feasibility of Large-Scale Genomic Testing to Facilitate Enrollment Onto Genomically Matched Clinical Trials. *J Clin Oncol: Off J Am Soc Clin Oncol* (2015) 33:2753–62. doi: 10.1200/JCO.2014.60.4165

5. Lang G, Jiang Y, Shi J, Yang F, Li X, Pei Y, et al. Characterization of the Genomic Landscape and Actionable Mutations in Chinese Breast Cancers by Clinical Sequencing. *Nat Commun* (2020) 11:5679. doi: 10.1038/s41467-020-19342-3
6. MacConaill L, Van Hummelen P, Meyerson M, Hahn W. Clinical Implementation of Comprehensive Strategies to Characterize Cancer Genomes: Opportunities and Challenges. *Cancer Discov* (2011) 1:297–311. doi: 10.1158/2159-8290.CD-11-0110
7. Kato S, Kim K, Lim H, Boichard A, Nikanjam M, Weihe E, et al. Real-World Data From a Molecular Tumor Board Demonstrates Improved Outcomes With a Precision N-Of-One Strategy. *Nat Commun* (2020) 11:4965. doi: 10.1038/s41467-020-18613-3
8. Hyman D, Puzanov I, Subbiah V, Faris J, Chau I, Blay J, et al. Vemurafenib in Multiple Nonmelanoma Cancers With BRAF V600 Mutations. *N Engl J Med* (2015) 373:726–36. doi: 10.1056/NEJMoa1502309
9. Varadhachary G. New Strategies for Carcinoma of Unknown Primary: The Role of Tissue-of-Origin Molecular Profiling. *Clin Cancer Res: Off J Am Assoc Cancer Res* (2013) 19:4027–33. doi: 10.1158/1078-0432.CCR-12-3030
10. Greco F, Lennington W, Spigel D, Hainsworth J. Molecular Profiling Diagnosis in Unknown Primary Cancer: Accuracy and Ability to Complement Standard Pathology. *J Natl Cancer Inst* (2013) 105:782–90. doi: 10.1093/jnci/djt099
11. Ye Q, Wang Q, Qi P, Chen J, Sun Y, Jin S, et al. Development and Clinical Validation of a 90-Gene Expression Assay for Identifying Tumor Tissue Origin. *J Mol Diagn: JMD* (2020) 22:1139–50. doi: 10.1016/j.jmoldx.2020.06.005
12. Moran S, Martinez-Cardús A, Boussios S, Esteller M. Precision Medicine Based on Epigenomics: The Paradigm of Carcinoma of Unknown Primary. *Nat Rev Clin Oncol* (2017) 14:682–94. doi: 10.1038/nrclinonc.2017.97
13. Pavlidis N, Khaled H, Gaafar R. A Mini Review on Cancer of Unknown Primary Site: A Clinical Puzzle for the Oncologists. *J Adv Res* (2015) 6:375–82. doi: 10.1016/j.jare.2014.11.007
14. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, et al. Site-Specific and Targeted Therapy Based on Molecular Profiling by Next-Generation Sequencing for Cancer of Unknown Primary Site: A Nonrandomized Phase 2 Clinical Trial. *JAMA Oncol* (2020) 6(12):1931–38. doi: 10.1001/jamaoncol.2020.4643
15. Hemminki K, Bevier M, Hemminki A, Sundquist J. Survival in Cancer of Unknown Primary Site: Population-Based Analysis by Site and Histology. *Ann Oncol: Off J Eur Soc Med Oncol* (2012) 23:1854–63. doi: 10.1093/annonc/mdr536
16. Slamon D, Clark G, Wong S, Levin W, Ullrich A, McGuire W. Human Breast Cancer: Correlation of Relapse and Survival With Amplification of the HER-2/Neu Oncogene. *Sci (New York NY)* (1987) 235:177–82. doi: 10.1126/science.3798106
17. Kapitanović S, Radosević S, Kapitanović M, Andelinović S, Ferencić Z, Tavassoli Z, et al. The Expression of P185(HER-2/Neu) Correlates With the Stage of Disease and Survival in Colorectal Cancer. *Gastroenterology* (1997) 112:1103–13. doi: 10.1016/S0016-5085(97)70120-3
18. Tanner M, Hollmén M, Junttila T, Kapanen A, Tommola S, Soini Y, et al. Amplification of HER-2 in Gastric Carcinoma: Association With Topoisomerase IIalpha Gene Amplification, Intestinal Type, Poor Prognosis and Sensitivity to Trastuzumab. *Ann Oncol: Off J Eur Soc Med Oncol* (2005) 16:273–8. doi: 10.1093/annonc/mdi064
19. Langer C, Stephenson P, Thor A, Vangel M, Johnson D. Trastuzumab in the Treatment of Advanced Non-Small-Cell Lung Cancer: Is There a Role? Focus on Eastern Cooperative Oncology Group Study 2598. *J Clin Oncol: Off J Am Soc Clin Oncol* (2004) 22:1180–7. doi: 10.1200/JCO.2004.04.105
20. Behr T, Béhé M, Wörmann B. Trastuzumab and Breast Cancer. *New Engl J Med* (2001) 345:995–6. doi: 10.1056/NEJM200109273451312

Conflict of Interest: QHX and YFS were employed by the company Canhelp 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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership