frontiers | Research Topics

Insights in Precision Medicine 2021

Edited by
Alice Chen

Published in
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-8325-2471-8
DOI 10.3389/978-2-8325-2471-8

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Insights in precision medicine: 2021

Topic editor
Alice Chen — Consultant, MD, United States

Citation
Chen, A., ed. (2023). *Insights in precision medicine: 2021*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2471-8

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 03 March 2023
DOI 10.3389/fmed.2023.1145876

OPEN ACCESS

EDITED AND REVIEWED BY
Michal Rosen-Zvi,
IBM Research, United States

*CORRESPONDENCE
Alice P. Chen
✉ chen.precisionmedicine@gmail.com

SPECIALTY SECTION
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

RECEIVED 16 January 2023
ACCEPTED 31 January 2023
PUBLISHED 03 March 2023

CITATION
Chen AP (2023) Editorial: Insights in precision medicine: 2021. *Front. Med.* 10:1145876. doi: 10.3389/fmed.2023.1145876

Editorial: Insights in precision medicine: 2021

Alice P. Chen*

AP Chen Consultant, Potomac, MD, United States

KEYWORDS

precision medicine, genomics, education, hyperthyroidism, technology

Editorial on the Research Topic
Insights in precision medicine: 2021

Precision medicine has been a hot topic in recent years as I have the privilege of seeing these changes in all fields especially in oncology. It offers the potential to revolutionize healthcare by tailoring treatment to a specific target. This approach, which considers genetic, environmental, and lifestyle factors, as well as more recently metabolomic and proteomic approaches has the potential to improve patient outcomes and reduce the risk of negative side effects. The recent Research Topic, "*Insights in precision medicine: 2021*," shed light into the diverse research even on the eve of COVID-19 and provide future challenges in the field of precision medicine. The nine articles published showcase the growth of the area and the increasing unanswered questions that each discovery brings. These articles are wide range from using AI to implement treatment to education and covers cancer to chronic conditions such as epilepsy and diabetes.

One key area of focus in precision medicine is the use of genomic information to guide treatment decisions. This is particularly relevant in the context of diseases such as renal cell carcinoma, hepatocellular carcinoma, and liver diseases, where specific genomic factors can be used to predict patient outcomes and guide treatment choices.

Four articles highlight the potential use of precision medicine in improving patient outcomes for chronic disorder outside of oncology. "*Long non-coding RNA MALAT1: A key player in liver diseases*" by Lu et al. discuss the multiple roles of MALAT1 which as often in nature can be beneficial or harmful. It is beneficial when regeneration of liver is needed but overexpression could lead to proliferation and metastasis. "*Low-dose everolimus maintenance therapy for renal angiomyolipoma associated with tuberous sclerosis complex*" by Luo et al. albeit a small study showed the long-term treatment of everolimus in renal angiomyolipoma. "*EEG-driven prediction model of oxcarbazepine treatment outcomes in patients with newly-diagnosed focal epilepsy*" by Wang et al. presents a prediction model based on electroencephalography data that is able to accurately predict treatment outcomes in patients with focal epilepsy being treated with oxcarbazepine. "*Genome-wide association study of hyperthyroidism based on electronic medical record from Taiwan*" by Liu et al. identifies genetic risk factors for hyperthyroidism, which can be used to guide treatment.

Schaibley et al.'s and Field's articles highlight the challenges and opportunities involved in implementing genomics-based precision medicine in clinical practice. "*Limited genomics training among physicians remains a barrier to genomics-based implementation of precision medicine*" raise the awareness that genomics training for healthcare professionals needs to start at the medical school level to teach physicians how to use genomic data in patient care. "*Bioinformatic challenges detecting genetic variation in precision medicine programs*" discusses the challenges involved in using bioinformatics techniques to analyze and interpret

genomic data in precision medicine programs, including the need to develop robust data management systems and to address issues of data privacy and security. The use of artificial intelligence (AI) in training and use for precision medicine can help to address these challenges by providing tools and algorithms that can automate the analysis of genomic data and support the development of personalized treatment plans. However, it is important to ensure that these AI-based approaches are validated and that they are used in a responsible manner that considers the unique needs and circumstances of individual patients.

The rest of the articles emphasize precision medicine in oncology; all highlight the potential of precision medicine in improving patient outcomes and tailoring treatment to the specific needs of individual patients in oncology. "*A somatic mutation signature predicts the best overall response to anti-programmed cell death protein-1 treatment in epidermal growth factor receptor/anaplastic lymphoma kinase-negative non-squamous non-small cell lung cancer*" by Peng et al. demonstrates the potential of using a somatic mutation signature, a profile of genetic changes present in a tumor, to predict the best overall response to anti-PD-1 treatment in non-small cell lung cancer. "*Precision medicine: An optimal approach to patient care in renal cell carcinoma*" by Sharma et al. discusses the genomic changes that are associated with renal cell carcinoma and how these changes can be used to guide treatment decisions. "*Development and validation of a prognostic signature associated with tumor microenvironment based on autophagy-related lncRNA analysis in hepatocellular carcinoma*" by Deng et al. presents a prognostic signature based on the analysis of long non-coding RNAs associated with the tumor microenvironment that is able to accurately predict patient outcomes in hepatocellular carcinoma. These studies show that precision medicine has the potential to revolutionize the management of cancer by providing personalized treatment strategies based on the unique characteristics of each patient's disease.

In conclusion, precision medicine is changing the way medicine is implemented and continues to improve patient outcomes. As significant as the advances are in this field, there are still challenges to be addressed including education of physicians in this area and new technologies to drive the science toward precision, the benefits of precision medicine are clear, and it is essential that we continue to invest in and support research in this field.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

AC is the sole proprietor of AP Chen Consultant.

Publisher's note

frontiers in Medicine

ORIGINAL RESEARCH
published: 24 November 2021
doi: 10.3389/fmed.2021.744050

Low-Dose Everolimus Maintenance Therapy for Renal Angiomyolipoma Associated With Tuberous Sclerosis Complex

Cong Luo[1†], Wen-Rui Ye[2†], Xiong-Bin Zu[1], Min-Feng Chen[1], Lin Qi[1], Yang-Le Li[1] and Yi Cai[1*]*

[1] Department of Urology, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha City, China, [2] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha City, China

OPEN ACCESS

Edited by:
Liangyou Gu,
People's Liberation Army General Hospital, China

Reviewed by:
Elzbieta Radzikowska,
National Institute of Tuberculosis and Lung Diseases, Poland
Alice Chen,
National Cancer Institute (NCI), United States

***Correspondence:**
Yang-Le Li
lyle2003@csu.edu.cn
Yi Cai
cai-yi@csu.edu.cn

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

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *19 July 2021*
Accepted: *05 November 2021*
Published: *24 November 2021*

Citation:
Luo C, Ye W-R, Zu X-B, Chen M-F, Qi L, Li Y-L and Cai Y (2021) Low-Dose Everolimus Maintenance Therapy for Renal Angiomyolipoma Associated With Tuberous Sclerosis Complex. Front. Med. 8:744050. doi: 10.3389/fmed.2021.744050

Objective: To assess the safety and efficacy of low-dose everolimus maintenance therapy for tuberous sclerosis complex-related renal angiomyolipoma (TSC-RAML) patients that had previously undergone standard-dose treatment for a minimum of 6 months.

Materials and Methods: In total, 24 patients with a definitive TSC diagnosis were enrolled from April 2018 – April 2019 at Xiangya Hospital, Central South University. All patients underwent low-dose everolimus maintenance therapy following standard-dose everolimus induction therapy for a minimum of 6 months. Patients additionally underwent TSC1/TSC2 genetic testing, And they were followed-up at 3, 6, 12, 18, and 24 months. The Response Evaluation Criteria in Solid Tumors (RECIST, version 1.1) criteria were used to monitor patient RAML responses, while adverse events (AEs) were assessed as per the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE, version 4.0). $P < 0.05$ was the significance level for all analyses, which were performed using SPSS 19.0.

Results: TSC1/TSC2 gene mutations were present in all 24 patients, all of whom achieved a significant reduction in TSC-RAML volume within the initial 6-month induction therapy period, and exhibited volume stabilization during the low-dose maintenance therapy treatment period without any instances of TSC-RAML regrowth. Adverse events (AEs) were significantly less severe and less frequent over the course of maintenance therapy relative to standard therapy.

Conclusions: Low-dose everolimus maintenance therapy represents an effective approach to achieving TSC-RAML control following a minimum of 6 months of full-dose induction therapy, and may be associated with decreases in everolimus-related AE frequency and severity.

Keywords: tuberous sclerosis complex, renal angiomyolipoma, everolimus, low-dose maintenance therapy, safety

INTRODUCTION

Tuberous sclerosis complex (TSC) is an autosomaldominant syndrome that impacts between 1 in 6,000 and 1 in 10,000 individuals, resulting in characteristic neurodevelopmental features and the development of multiple tumors in organs including the skin, heart, lungs, brain, and kidneys (1). Upwards of 80% of TSC patients are affected by renal angiomyolipoma (RAML) (2), which is characterized by multiple bilateral lesions in the smooth muscles, adipose tissue, and vasculature (3). As these tumors typically grow over time, TSC-RAML can result in arterial hypertension and imposes a risk of life-threatening hemorrhage, which is the leading cause of TSC-associated mortality among adults with this condition (4).

Most TSC patients present with mutations in the *TSC1* or *TSC2* genes, which encode proteins that form the TSC1-TSC2 complex that serves to antagonize the signaling pathway downstream of mammalian target of rapamycin (mTOR) by promoting the activation of the small GTPase Rheb and thereby inhibiting cellular growth and proliferation. Pathogenic TSC1/TSC2 variants result in constitutive mTOR pathway hyperactivation, thereby contributing to the growth of benign tumors or hamartomas in multiple systems (5).

Everolimus is an mTOR inhibitor that has shown promise for the treatment of complications associated with TSC including RAML, seizures, facial angiofibromas, and subependymal giant cell astrocytomas (SEGAs) (6–9). Indeed, everolimus treatment can result in an initial rapid decrease in TSC-RAML volume, followed by a secondary phase during which these tumors slowly shrink or stabilize (7). The International Tuberous Sclerosis Complex Consensus Conference held in 2012 recommended the first-line use of mTOR inhibitors for the treatment of RAML $\geq$ 3 cm in diameter, even when not associated with any clinical symptoms (10).

Prior work suggests that TSC-RAML regrowth may occur following the cessation of mTOR inhibitor therapy, and the

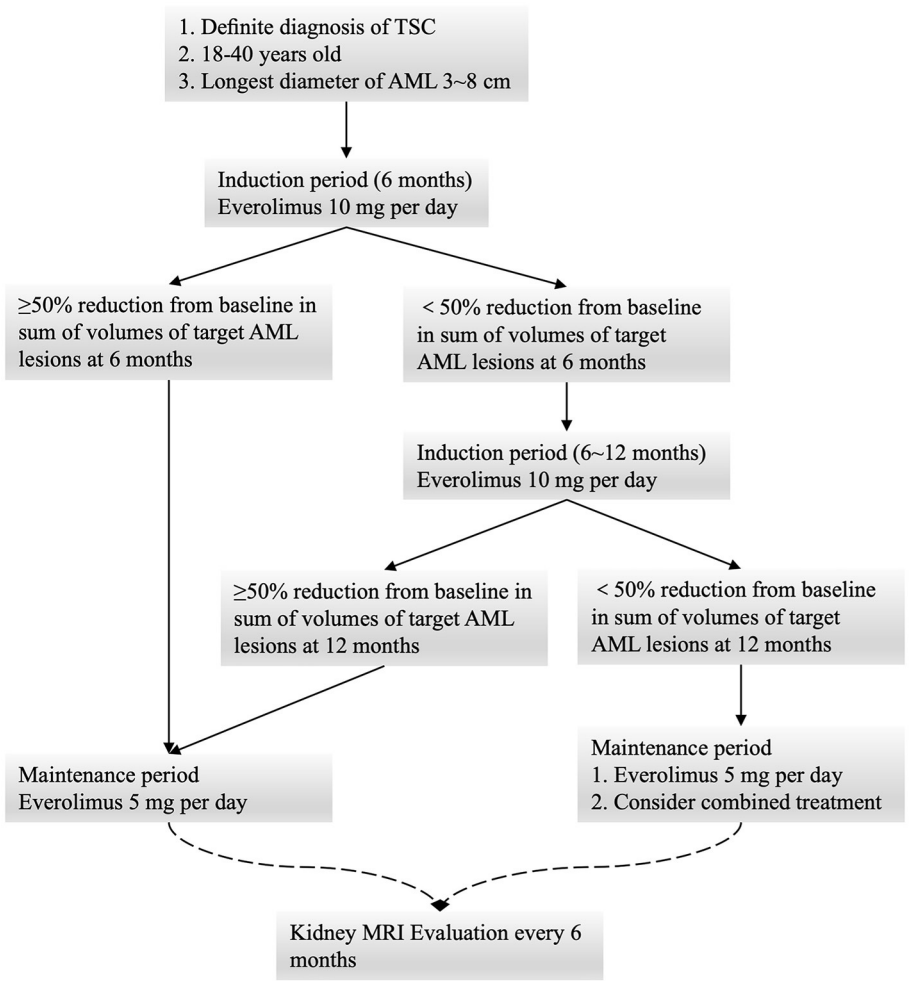

FIGURE 1 | Low-dose maintenance treatment schematic diagram.

ideal duration for this therapeutic strategy remains to be defined optimal duration of mTOR inhibitor treatment has yet to be determined (11, 12). With respect to safety, the short-term adverse effects associated with everolimus are typically acceptable, although in a few instances more severe events have been reported (13). Long-term treatment-related safety outcomes, however, remain to be established. Wheless and Klim (14) proposed a dose reduction algorithm designed to minimize the negative impact of mTOR inhibitor treatment for patients with SEGAs that are shrinking or stable in size. We were thus interested in whether the everolimus dose could similarly be reduced to control the frequency and severity of adverse events (AEs) in patients with controlled TSC-RAML. This study was therefore designed to examine the safety and efficacy of low-dose everolimus maintenance therapy in TSC-RAML patients that had previously undergone treatment with a standard everolimus dose for a minimum of 6 months.

METHODS

Study Group

This was a single-center, open-label, single-arm, prospective interventional study performed between April 2018 and April 2019 at Xiangya Hospital, Central South University. The Human Ethics Committee of Xiangya Hospital, Central South University approved this study prior to patient enrollment, and all protocols were performed in accordance with the Declaration of Helsinki (15). Patients provided written informed consent prior to voluntary study participation. Patients eligible for inclusion were: 1) individuals with a definitive TSC diagnosis as defined by meeting 2 major criteria or 1 major criterion and $\geq$ 2 minor criteria recommended by the 2012 International Tuberous Sclerosis Complex Consensus Conference (10); 2) individuals $\geq$ 18 years old; 3) individuals with a minimum of one RAML $\geq$ 30 mm in diameter.

All patients underwent oral everolimus induction therapy (10 mg/day) for 6 months, after which they underwent radiographic follow-up and a safety evaluation. All patients that achieved a $\geq$ 50% decrease in the total volume of the target AML (relative to baseline) were assigned to the low-dose oral everolimus maintenance group (5 mg/day), while patients not meeting these criteria underwent induction therapy for an additional 6 months. Follow-up was then repeated at 12 months, at which time patients were assigned to undergo low-dose maintenance therapy regardless of the observed reduction in AML size. Combination treatment options were considered for individuals exhibiting a poor response to everolimus. Patient follow-up was performed at 3, 6, 12, 18, and 24 months (**Figure 1**).

Patient Evaluation and Follow-Up

Abdominal magnetic resonance imaging (MRI) was used to visualize RAML tumors at baseline, with up to four RAMLs with a maximum diameter $\geq$ 3.0 cm being identified as target lesions in each patient. The sum of the diameters of these target lesions was calculated. Over the course of follow-up, patients underwent routine urine, blood, physical, and radiographic analyses. The Response Evaluation Criteria in Solid Tumors (RECIST, version 1.1) criteria were used to monitor patient RAML responses, while adverse events (AEs) were assessed as per the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE, version 4.0).

TABLE 1 | Baseline patient demographic and disease characteristics.

	Everolimus (*N* = 24)
Age in years, median (range)	27 (19–33)
<30	18 (75%)
≥30	6 (25%)
Sex	
Male	8 (33.3%)
Female	16 (66.7%)
Gene mutation	
TSC1	4 (16.7%)
TSC2	20 (83.3%)
Epilepsy	3 (12.5%)
Diagnosis of LAMs	5 (20.8%)
Skin lesions (≥1)	24 (100%)
Presence of SEGAs	1 (4.1%)
Renal impairment (GFR < 60 mL/min)	1 (4.1%)
Diameter of the largest RAML lesions	
6–8 cm	11 (45.8%)
4–6 cm	11 (45.8%)
3–4 cm	2 (8.3%)
Sum of volumes of target renal angiomyolipoma lesions, cm^3	
Mean (SD, cm^3)	155.7 (100.4)
Median (range, cm^3)	121.6 (29.6∼348)
Bilateral angiomyolipoma	20 (88.3%)
Number of target RAML lesions	
1∼2	14 (58.3%)
3∼4	10 (41.7%)
Previous angiomyolipoma therapy	
Surgery/invasive procedure	9 (37.5%)
Renal embolization	3 (12.5%)
Partial nephrectomy	2 (8.3%)
Nephrectomy	4 (16.7%)
Medication	0 (0)

LAMs, lymphangioleiomyomatosis; SEGAs, subependymal giant cell astrocytomas; RAML, renal angiomyolipoma.

Genetic Analysis

TSC1/TSC2 mutational status was assessed via a next-generation sequencing (NGS) approach at the NHC Key Laboratory of Cancer Proteomics (Hunan Province, China). Pathogenic mutations were confirmed through reference to the LOVD databases (www.lovd.nl/TSC1; www.lovd.nl/TSC2). The potential impact of newly identified mutations resulting in amino acid substitutions on protein function was assessed with the online SIFT and PolyPhen2 tools.

TABLE 2 | Mutations detected by next-generation sequencing.

No	Sex	Age	Mutant gene	Nucleotide change	Protein change	Mutation Type
1	Female	19	TSC2	c.1922G>T	Ser641Ile	Missense mutations
2	Female	22	TSC2	c.2407C>T	Gln803*	Nonsense mutations
3	Female	23	TSC2	c.2407C>T	Gln803*	Nonsense mutations
4	Female	24	TSC2	c.3838G>T	Gln1280*	Nonsense mutations
5	Male	24	TSC2	c.5161del	Met1721Trpfs*105	Frameshift mutations
6	Male	27	TSC2	c.1001T>G	Val334Gly	Missense mutations
7	Female	28	TSC1	c.1431_1434del	Glu478Lysfs*53	Frameshift mutations
8	Female	29	TSC2	c.2098-2A>G	p?	Intron mutation
9	Male	32	TSC2	c.3707T>C	Met1236Thr	Missense mutations
10	Female	24	TSC1	c.1960C>T	Gln654*	Nonsense mutations
11	Male	26	TSC2	c.2785G>T	Glu929*	Nonsense mutations
12	Female	26	TSC2	c.1348G>T	Glu450*	Nonsense mutations
13	Female	25	TSC1	c.1960C>T	Gln654*	Nonsense mutations
14	Male	26	TSC2	c.5024C>T	Pro1675Leu	Missense mutations
15	Male	26	TSC2	c.820T>A	Tyr274Asn	Missense mutations
16	Female	27	TSC1	309G>A	Trp103*	Nonsense mutations
17	Female	28	TSC2	c.2251C>T	Arg751*	Nonsense mutations
18	Male	29	TSC2	c.2988del	Ser997Valfs*19	Frameshift mutations
19	Female	29	TSC2	c.4604A>T	Asp1535Val	Missense mutations
20	Male	30	TSC2	c.3180G>A	Trp1060*	Nonsense mutations
21	Female	31	TSC2	c.3412C>T	Arg1138*	Nonsense mutations
22	Female	31	TSC2	c.4708A>T	Arg1570Trp	Missense mutations
23	Female	33	TSC2	c.1547_1559delinsGTGCTGCC	Ala516Glyfs*71	Frameshift mutations
24	Female	33	TSC2	EX25_36 DEL	–	Large rearrangements

TABLE 3 | Response of AML volume to everolimus therapy.

	3 months	6 months	12 months	18 months	24 months
Patients (*n*)	24	24	24	24	22
No. of response (*n*, %)	12 (50)	12 (50)	13 (54)	13 (54)	12 (55)
*% of baseline value (Mean ± SD, %)	48 ± 18	52 ± 19	53 ± 20	53 ± 19	52 ± 19

**The average percentage change of baseline in the total volume of all target AML lesions.*

Statistical Analysis

Continuous variables are given as mean ± standard deviation (M ± SD), while categorical variables are given in the form of frequencies (n) and percentages (%). Categorical variables were compared using the chi-square test. SPSS 19.0 (SPSS, IL, USA) was used for all statistical testing, with $P < 0.05$ as the significance threshold.

RESULTS

Patient Characteristics

In total, 24 patients (8 male, 16 female) were enrolled in this study, with a median age of 27 years. Patient demographics are compiled in **Table 1**. Of these patients, 20 and 4 were found to harbor *TSC2* and *TSC1* mutations, respectively, via NGS (**Table 2**). Three of these patients had a history of epilepsy, two were treated with antiepileptic monotherapy (oxcarbazepine, lamotrigine), while the remaining one with antiepileptic combination therapy (oxcarbazepine + topiramate). Five of these patients were diagnosed with lymphangioleiomyomatosis (LAM), all suffered from skin lesions, and one presented with SEGAs. One patient had a renal impairment of GFR < 60 mL/min. With respect to the baseline RAML characteristics in these patients, 11 exhibited RAMLs with a maximum diameter ≥ 6 cm, 20 exhibited bilateral RAMLs, and 10 presented with 3-4 target RAML lesions.

Before enrollment in this study, 3, 2, and 4 of these patients had respectively undergone renal embolization, partial nephrectomy, and nephrectomy. No patients had undergone prior medication therapy. Over the follow-up period, two patients withdrew from the study at 24 months, leaving 22 patients for the assessment of RAML status. One gave up medication for economic reason, and the other withdrew from the study due to the unavailability of everolimus resulted from COVID-19 pandemic.

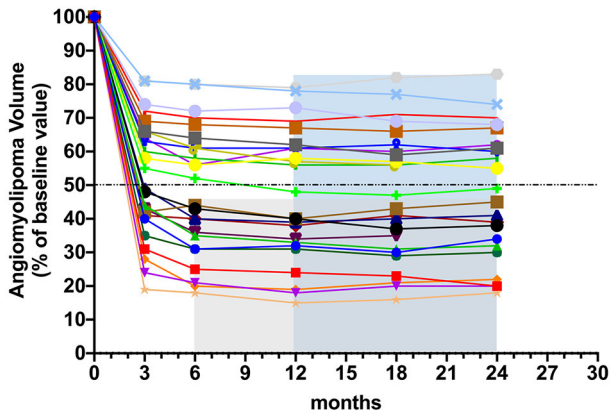

FIGURE 2 | Changes in TSC-RAML volume from baseline during the induction therapy and the maintenance therapy. Each line represents TSC-RAML volume change in one patient.

Treatment Efficacy

The patients' response of AML volume to everolimus during treatment is detailed in **Table 3**. The number of patients who achieved ≥ 50% reduction in RAML volume was 12 (50%) at 6 months and 13 (54%) at 12 months, respectively. The change in RAML volume for each patient over the study period is displayed in **Figure 2**, with the most significant decrease in tumor volume having been observed within the initial 6 months of standard-dose everolimus therapy. In total, 12 patients achieved ≥ 50% reduction in total target AML volume at 6 months, whereupon they initiated low-dose everolimus maintenance therapy. Just one of the remaining 12 patients achieved ≥ 50% reduction in target AML volume after an additional round of full-dose everolimus treatment. Target RAML volumes were well-controlled in all patients during the maintenance therapy period.

The pulmonary function in 5 female patients with LAM during everolimus therapy are detailed in **Table 4**. At baseline, four patients showed moderate airflow obstruction (forced expiratory volume in 1 second (FEV1): 50–70% of the predicted value), while one patient showed severe airflow obstruction (FEV1 < 50% of the predicted value). During the medication period, an increase was observed in FEV1, forced vital capacity (FVC), total lung capacity, and diffusion capacity for carbon monoxide (DLCO) among all the five patients, while the residual volume decreased. The changes in FEV1, FVC, and residual volume for each patient are shown in **Figure 3**. It further displayed the improved pulmonary function in patients with LAM. These results indicated that both full-dose and low-dose everolimus treatment have a protective effect on pulmonary function in patients with TSC.

Adverse Events

All AEs recorded during the induction and maintenance phases of everolimus treatment are compiled in **Table 5**. While six grade 3–4 AEs occurred during full-dose induction therapy, none occurred during low-dose maintenance therapy. The most common AEs during full-dose induction therapy included oral mucositis (22/24), abdominal pain (10/24), hypertriglyceridemia (10/24), and headache (9/24), while the most common AEs during low-dose maintenance treatment were oral mucositis (10/24) and hypertriglyceridemia (9/24). AEs with significant incidence reductions during low-dose maintenance therapy included oral mucositis ($P < 0.001$), irregular menstruation ($P = 0.04$), upper respiratory infections ($P = 0.02$), and fever ($P = 0.04$). No unexpected AEs or mortality were reported, and no patients declined treatment or withdrew from the study due to AEs.

TABLE 4 | Pulmonary functional characteristics of patients with LAM.

Value	Baseline ($n = 5$)	12 months ($n = 4$)	24 months ($n = 4$)
FEV1			
Least-square mean (liters)	1.30 ± 0.15	1.63 ± 0.27	1.69 ± 0.29
Percent of predicted value	55.08 ± 9.26	62.06 ± 10.09	73.02 ± 14.93
FVC			
Least-square mean (liters)	2.52 ± 0.54	3.02 ± 0.62	3.67 ± 0.72
Percent of predicted value	73.29 ± 12.37	85.52 ± 12.13	100.39 ± 11.37
Total lung capacity			
Least-square mean (liters)	4.80 ± 0.52	5.18 ± 0.29	5.91 ± 0.53
Percent of predicted value	92.09 ± 6.64	100.23 ± 4.22	112.92 ± 9.13
Residual volume			
Least-square mean (liters)	2.28 ± 0.07	1.97 ± 0.14	1.90 ± 0.18
Percent of predicted value	120.40 ± 5.35	111.72 ± 6.69	112.30 ± 7.06
DLCO			
Least-square mean (ml/mmHg/min)	12.03 ± 2.52	12.70 ± 2.62	15.34 ± 2.10
Percent of predicted value	49.43 ± 9.53	52.13 ± 8.01	62.08 ± 9.35

FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; DLCO, diffusion capacity for carbon monoxide.

DISCUSSION

This study is the first to our knowledge to have assessed the efficacy and safety of low-dose everolimus maintenance therapy for the treatment of TSC-RAML patients after a minimum 6-month full-dose induction therapy period.

Owing to the potential for rebound after withdrawal, sustained everolimus therapy is necessary to effectively control TSC-RAML. However, continuous everolimus treatment is associated with a number of issues. First, prolonged standard everolimus treatment is expensive and can impose a major economic burden on patients that decreases their compliance. Second, lifelong mTOR inhibitor treatment is often required for TSC patients,

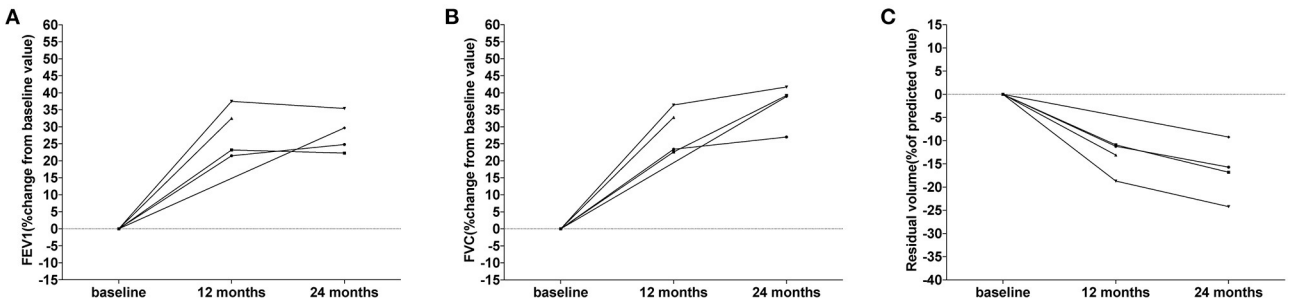

FIGURE 3 | Changes of pulmonary function in the five patients with LAM. **(A)** The changes of FEV1 from baseline during treatment. Each line represents FEV1 change in one patient. **(B)** The changes of FVC from baseline during treatment. Each line represents FVC change in one patient. **(C)** The changes of residual volume from baseline during treatment. Each line represents residual volume change in one patient.

TABLE 5 | Adverse events associated with everolimus in the study group during induction and maintenance therapy.

Adverse events	Induction therapy		Maintenance therapy		*P*-value
	All grade	Grade 3/4	All grade	Grade 3/4	
Mucositis oral	22/24	0	10/24	0	**<0.001**
Irregular menstruation	8/16	4/16	1/16	0	**0.04**
Abdominal pain	10/24	0	5/24	0	0.12
Hypertriglyceridemia	10/24	0	9/24	0	0.77
Headache	9/24	0	6/24	0	0.35
Diarrhea	8/24	0	7/24	0	0.76
Upper respiratory infection	7/24	0	1/24	0	**0.02**
Proteinuria	6/24	1/24	5/24	0	0.73
Malaise	6/24	0	5/24	0	0.73
Rash acneiform	5/24	1/24	2/24	0	0.22
Cholesterol high	5/24	0	3/24	0	0.44
Fever	4/24	0	0	0	**0.04**
Urinary tract infection	4/24	0	2/24	0	0.38
Hematuria	3/24	0	0	0	0.07
Alkaline phosphatase increased	3/24	0	4/24	0	0.68
Constipation	3/24	0	1/24	0	0.30
GGT increased	3/24	0	2/24	0	0.64
Hypophosphatemia	3/24	0	1/24	0	0.30
Seizures	2/24	0	0	0	0.15
Pneumonitis	2/24	0	0	0	0.15
Vomiting	2/24	0	0	0	0.15
Lymphocyte count decreased	2/24	0	1/24	0	0.55
Anemia	2/24	0	2/24	0	1
Neutrophil count decreased	1/24	0	0	0	0.31
Hyperuricemia	1/24	0	0	0	0.31
Creatinine increased	1/24	0	1/24	0	1

The bold values refers to $p < 0.05$ with statistic significance.

particularly for individuals < 40 years of age, emphasizing the need to explore more feasible or cost-effective solutions. Third, sustained standard everolimus treatment can result in potentially severe AEs. Lastly, the most prominent tumor growth reduction generally occurs within the initial 3–6 months of treatment in patients, whereafter tumor volumes tend to stabilize or decrease gradually (16). We therefore designed the present study to assess the ability of low-dose everolimus maintenance to control RAML volumes and to reduce AE incidence, given that such an approach has previously been reported to be successful in the treatment of TSC-related SEGAs (17).

Herein, the everolimus dose utilized for low-dose maintenance therapy was reduced to 5 mg/day from 10 mg/day. All patients achieved a significant reduction in TSC-RAML volume over the first 6 months of induction treatment, and maintained a stable TSC-RAML volume during the low-dose maintenance period without any evidence of target AML lesion growth or progression. This suggests that low-dose everolimus maintenance therapy is an effective therapeutic option for TSC-AML patients. Moreover, pulmonary functions, including FEV1, FVC, total lung capacity, DLCO, and residual volume, were improved in five patients with LAM during both full- and low-dose everlimus therapy, which further confirmed the efficacy of low-dose everlimus therapy.

Everolimus therapy is commonly associated with a range of AEs that can affect treatment efficacy and compliance in some patients. We found that these everolimus therapy-related AEs were significantly less frequent and less severe during the low-dose maintenance therapy compared with the standard treatment period, consistent with a previous study (17). Reducing the incidence of oral mucositis is critical to improving patient compliance. Most importantly, no grade 3-4 AEs were observed in the context of low-dose maintenance therapy, in contrast to the incidence of such complications during full-dose everolimus treatment. These results thus suggest that low-dose everolimus maintenance therapy is a feasible and well-tolerated option for patients with TSC-RAML.

While these results are promising, this study is nonetheless limited by the fact that it is a single-center analysis of a relatively small patient population. Even so, we hope that this study can provide a reference for the everolimus treatment of TSC-RAML patients, particularly those patients that exhibit poor tolerance for full-dose everolimus therapy. We plan to recruit more patients for treatment with this low-dose maintenance therapy regimen, and as such, our available efficacy and safety data will continue to expand in the future. Another noteworthy limitation of this study is that we were unable to assess drug concentrations in patient blood owing to technical limitations. However, we believe that consistent dosing will largely mitigate the potential bias associated with this limitation.

CONCLUSIONS

Low-dose everolimus maintenance therapy is an effective therapeutic approach to controlling TSC-RAML following full-dose induction therapy, and may reduce the frequency and severity of AEs associated with everolimus.

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 Xiangya Hospital, Central South University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CL, W-RY, Y-LL, and YC participated in its design and coordination and drafted the manuscript. CL, W-RY, and Y-LL participated in the design of the study and performed the statistical analysis. CL, W-RY, M-FC, LQ, X-BZ, Y-LL, and YC conceived of the study, participated in its design, and coordination and helped to draft the manuscript. All authors have read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant No. 81800590) and the Natural Science Foundation of Hunan Province (Grant No. 2020JJ5882).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.744050/full#supplementary-material

REFERENCES

1. Osborne JP, Fryer A, Webb D. Epidemiology of tuberous sclerosis. *Ann N Y Acad Sci.* (1991) 615:125–7. doi: 10.1111/j.1749-6632.1991.tb37754.x
2. Henske EP, Józwiak S, Kingswood JC, Sampson JR, Thiele EA. Tuberous sclerosis complex. *Nat Rev Dis Primers.* (2016) 2:16035. doi: 10.1038/nrdp.2016.35
3. Cai Y, Li H, Zhang Y. Assessment of tuberous sclerosis complex associated with renal lesions by targeted next-generation sequencing in Mainland China. *Urology.* (2017) 101:70.e1. doi: 10.1016/j.urology.2016.10.056
4. Kapoor A, Girard L, Lattouf JB, Pei Y, Rendon R, Card P, et al. Evolving Strategies in the Treatment of Tuberous Sclerosis Complex-associated Angiomyolipomas (TSC-AML). *Urology.* (2016) 89:19–26. doi: 10.1016/j.urology.2015.12.009
5. Zhang Y, Gao X, Saucedo LJ, Ru B, Edgar BA, Pan D. Rheb is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nat Cell Biol.* (2003) 5:578–81. doi: 10.1038/ncb999
6. Mizuguchi M, Ikeda H, Kagitani-Shimono K, Yoshinaga H, Suzuki Y, Aoki M, et al. Everolimus for epilepsy and autism spectrum disorder in tuberous sclerosis complex: EXIST-3 substudy in Japan. *Brain Dev.* (2019) 41:1-10. doi: 10.1016/j.braindev.2018.07.003
7. Bissler JJ, Kingswood JC, Radzikowska E, Zonnenberg BA, Frost M, Belousova E, et al. Everolimus for angiomyolipoma associated with tuberous sclerosis complex or sporadic lymphangioleiomyomatosis (EXIST-2): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet.* (2013) 381:817–24. doi: 10.1016/S0140-6736(12)61767-X
8. Wataya-Kaneda M, Nakamura A, Tanaka M, Hayashi M, Matsumoto S, Yamamoto K, et al. Efficacy and safety of topical sirolimus therapy for facial

angiofibromas in the tuberous sclerosis complex : a randomized clinical trial. *JAMA Dermatol.* (2017) 153:39–48. doi: 10.1001/jamadermatol.2016.3545
9. Franz DN, Belousova E, Sparagana S, Bebin EM, Frost M, Kuperman R, et al. Efficacy and safety of everolimus for subependymal giant cell astrocytomas associated with tuberous sclerosis complex (EXIST-1): a multicentre, randomised, placebo-controlled phase 3 trial. *Lancet.* (2013) 381:125–32. doi: 10.1016/S0140-6736(12)61134-9
10. Northrup H, Krueger DA. Tuberous sclerosis complex diagnostic criteria update: recommendations of the 2012 Iinternational Tuberous Sclerosis Complex Consensus Conference. *Pediatr Neurol.* (2013) 49:243–54. doi: 10.1016/j.pediatrneurol.2013.08.001
11. Bissler JJ, McCormack FX, Young LR, Elwing JM, Chuck G, Leonard JM, et al. Sirolimus for angiomyolipoma in tuberous sclerosis complex or lymphangioleiomyomatosis. *N Engl J Med.* (2008) 358:140–51. doi: 10.1056/NEJMoa063564
12. Cai Y, Guo H, Wang W, Li H, Sun H, Shi B, et al. Assessing the outcomes of everolimus on renal angiomyolipoma associated with tuberous sclerosis complex in China: a two years trial. *Orphanet J Rare Dis.* (2018) 13:43. doi: 10.1186/s13023-018-0781-y
13. Trelinska J, Dachowska I, Kotulska K, Fendler W, Jozwiak S, Mlynarski W. Complications of mammalian target of rapamycin inhibitor anticancer treatment among patients with tuberous sclerosis complex are common and occasionally life-threatening. *Anticancer Drugs.* (2015) 26:437–42. doi: 10.1097/CAD.0000000000000207
14. Wheless JW, Klimo P Jr. Subependymal giant cell astrocytomas in patients with tuberous sclerosis complex: considerations for surgical or pharmacotherapeutic intervention. *J Child Neurol.* (2014) 29:1562–71. doi: 10.1177/0883073813501870
15. Gandevia B, Tovell A. Declaration of Helsinki. *Med J Aust.* (1964) 2:320–1. doi: 10.5694/j.1326-5377.1964.tb115781.x
16. Bissler JJ, Budde K, Sauter M, Franz DN, Zonnenberg BA, Frost MD, et al. Effect of everolimus on renal function in patients with tuberous sclerosis complex: evidence from EXIST-1 and EXIST-2. *Nephrol Dial Transplant.* (2019) 34:1000–8. doi: 10.1093/ndt/gfy132
17. Trelinska J, Dachowska I, Baranska D, Stawiski K, Kotulska K, Fendler W, et al. Maintenance therapy with everolimus for subependymal giant cell astrocytoma in patients with tuberous sclerosis (the EMINENTS study). *Pediatr Blood Cancer.* (2017) 64:347. doi: 10.1002/pbc.26347

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 Medicine

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fmed.2021.762570

Development and Validation of a Prognostic Signature Associated With Tumor Microenvironment Based on Autophagy-Related lncRNA Analysis in Hepatocellular Carcinoma

Yan Deng[1†], *Feng Zhang*[2†], *Zhen-Gang Sun*[1] *and Shuai Wang*[1*]

[1] *Department of Hepatobiliary Surgery, Jing Zhou Central Hospital, The Second Clinical Medical College, Yangtze University, Jing Zhou, China,* [2] *Department of Ophthalmology, Jing Zhou Central Hospital, The Second Clinical Medical College, Yangtze University, Jing Zhou, China*

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute,
United States

Reviewed by:
Peng Chen,
Guangzhou University of Chinese Medicine, China
Xianhan Jiang,
Fifth Affiliated Hospital of Guangzhou Medical University, China

***Correspondence:**
Shuai Wang
m18171998538@163.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *22 August 2021*
Accepted: *15 November 2021*
Published: *14 December 2021*

Citation:
Deng Y, Zhang F, Sun Z-G and Wang S (2021) Development and Validation of a Prognostic Signature Associated With Tumor Microenvironment Based on Autophagy-Related lncRNA Analysis in Hepatocellular Carcinoma. Front. Med. 8:762570. doi: 10.3389/fmed.2021.762570

Objective: The present study aimed to establish a prognostic signature based on the autophagy-related long non-coding RNAs (lncRNAs) analysis in patients with hepatocellular carcinoma (HCC).

Methods: Patients with HCC from The Cancer Genome Atlas (TCGA) were taken as the training cohort, and patients from the International Cancer Genome Consortium (ICGC) were treated as the validation cohort. Autophagy-related lncRNAs were obtained *via* a co-expression network analysis. According to univariate and multivariate analyses, a multigene prognostic signature was constructed in the training cohort. The predictive power of the signature was confirmed in both cohorts. The detailed functions were investigated using functional analysis. The single-sample gene set enrichment analysis (ssGSEA) score was used to evaluate the tumor microenvironment. The expression levels of immunotherapy and targeted therapy targets between the two risk groups were compared. Finally, a nomogram was constructed by integrating clinicopathological parameters with independently predictive value and the risk score.

Results: Four autophagy-related lncRNAs were identified to establish a prognostic signature, which separated patients into high- and low-risk groups. Survival analysis showed that patients in the high-risk group had a shorter survival time in both cohorts. A time-independent receiver-operating characteristic (ROC) curve and principal component analysis (PCA) confirmed that the prognostic signature had a robust predictive power and reliability in both cohorts. Functional analysis indicated that the expressed genes in the high-risk group are mainly enriched in autophagy- and cancer-related pathways. ssGSEA revealed that the different risk groups were associated with the tumor microenvironment. Moreover, the different risk groups had positive correlations with the expressions of specific mutant genes. Multivariate analysis showed that the risk score also exhibited excellent predictive power irrespective of clinicopathological characteristics in both cohorts. A nomogram was established. The nomogram showed good discrimination, with Harrell's concordance index (C-index) of 0.739 and good calibration.

Conclusion: The four autophagy-related lncRNAs could be used as biological biomarkers and therapeutic targets. The prognostic signature and nomogram might aid clinicians in individual treatment optimization and clinical decision-making for patients with HCC.

Keywords: hepatocellular carcinoma, long non-coding RNA, autophagy, prognostic signature, TCGA, ICGC

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the fatal tumors occurring worldwide due to its aggressive biological behavior, rapidly increasing frequency, and high mortality (1, 2). Undoubtedly, the most difficult challenges that most clinicians face are early diagnosis and surgical intervention (3). Despite significant improvements in diagnosis and multimodal therapies, the survival benefit remains limited, owning to high heterogeneity (4–6). Hence, reliable predictive and prognostic biomarkers should be discovered to improve risk prediction ability and guide individualized therapy.

Autophagy is a multistep lysosomal degradation system that facilitates metabolic adaptability and nutrition cycling. These are biological processes that keep cell functioning properly (7–9). Autophagy has also been implicated in a variety of diseases, including cancer (10). However, the roles of autophagy in cancer are bilateral. On one hand, autophagy could offer the essential circulating metabolic substrates and enzymes to respond to various adverse circumstances; on the other hand, inappropriate autophagy enables malignant cells to proliferate rapidly, especially in advanced cancer (11, 12). Many studies have looked into a novel possible target therapy by investigating autophagy mechanisms (13, 14).

Long non-coding RNA (lncRNA) is a class of newly found RNA transcripts that cannot code for proteins. It usually has more than 200 nucleotides (15). By controlling transcriptionally or post-transcriptionally biological processes such as autophagy, an increasing number of lncRNAs have been linked to various physiological and physiological progress, including gene expression regulation, RNA decay, microRNA regulation, and protein folding (16, 17). Accumulating evidence suggested that lncRNAs could inhibit or activate the autophagy process through altering autophagy-related genes or pathways (18, 19).

With rapid advances in the RNA-sequencing technology, the potential for utilizing a lncRNA as a biomarker to aid the cancer detection, treatment, or prognosis has been gradually revealed (20). Using a comprehensive analysis of microarray data from The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) databases, the current study aimed to establish an autophagy-related lncRNA prognostic signature and a prognostic nomogram to predict the clinical outcome of patients with HCC.

Abbreviations: lncRNAs, long non-coding RNAs; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium; HCC, hepatocellular carcinoma; HADb, Human Autophagy Database; KM, Kaplan–Meier; AIC, Akaike information criterion; ROC, receiver-operating characteristic; AUC, area under the ROC curve; C-index, Harrell's concordance index; GO, Gene Ontology; KEGG, Kyoto Gene and Genomic Encyclopedia; GSEA, Gene set enrichment analysis; OS, overall survival.

MATERIALS AND METHODS

Patient Data Acquisition

RNA-sequencing of patients with HCC and accompanying clinical data were downloaded from the TCGA (https://portal.gdc.cancer.gov/) and ICGC (https://icgc.org/). Patients with a follow-up duration of <1 month were excluded for survival analysis. The training group consisted of 343 patients with HCC from the TCGA database, and the clinical data were shown in **Supplementary Excel S1**. At the same time, the validation group consisted of 230 patients with HCC from the ICGC database. The clinical information is shown in **Supplementary Excel S2**.

Due to the collection of all the data directly from public databases, no protocol was required from the ethical committee.

Autophagy-Related lncRNAs Screening

A total of 232 autophagy-related genes were obtained from the Human Autophagy Database (HADb, http://autophagy.lu/clustering/index.html). Then, the expression levels of these autophagy-related genes were retrieved from the TCGA and ICGC data sets.

The co-expression network between the expression of lncRNAs and autophagy-associated genes was investigated. LncRNAs with a correlation coefficient $|R| > 0.5$ and $p < 0.050$ were considered to be autophagy-related lncRNAs.

The lncRNA–mRNA co-expression network was constructed to explore the relationships between the autophagy-related lncRNAs and their mRNA counterparts. Cytoscape software (version 3.7.2) was used to visualize the co-expression network. Sankey plot was utilized to reveal the detailed relationships by the R studio software using the "ggalluvial" R package.

Construction of an Autophagy-Related lncRNA Signature

The "survival" R package performed the Kaplan–Meier (KM) method and univariate Cox regression analysis to screen out prognostic autophagy-related lncRNAs with both significant values of $p < 0.050$ in the training cohort. Then, among these nominated autophagy-related lncRNAs, the multivariate Cox regression analysis was employed by the "survival" R package to assess their contributions as prognostic factors. The lowest Akaike information criterion (AIC) value was used to find the best autophagy-related lncRNAs. Subsequently, the risk score was established by the multiplication of the sum of the coefficients using autophagy-related lncRNAs expressions.

Evaluation and Validation of the Prognostic Signature

A risk score was assigned to each patient with HCC. Based on the median value of their risk scores, all patients were classified into high- (high-risk score) and low-risk (low-risk score) groups. The prognosis of the two groups was compared using the KM survival curve, and the difference was assessed using a two-sided log-rank test. Time-dependent receiver-operating characteristic (ROC) curve analysis was performed using the "survival," "survminer," and "timeROC" R packages to evaluate the specificity and sensitivity of the prognostic signature. The prognosis accuracy was measured by the area under the ROC curve (AUC), a measure of discrimination. AUC ranges from 0.5 (little predictive power) to 1 (perfect prediction). Principal component analysis (PCA) was performed using the "ggplot2" R package to explore distinguishability. Following that, the distribution of patient's risk scores and scatter dots plot were depicted to visualize the detailed correlations of dead states with risk scores.

The subgroup survival analysis stratified by clinicopathological variables was conducted to evaluate the prognostic signature's accuracy across multiple cohorts.

Functional Analysis

The Gene Ontology (GO) and the Kyoto Gene and Genomic Encyclopedia (KEGG) were used to enhance the potential functional pathways and categories based on co-expressed genes of autophagy-related lncRNAs. Significant values of p and q were defined as <0.050. GO and KEGG analyses were conducted by applying the "org.Hs.eg.db," "colorspace," "stringi," "ggplot2," "dose," "clusterProfiler," and "enrichplot" R packages.

The gene set enrichment analysis (GSEA) was utilized to interpret the functional enrichment of gene expression data. This method derives its function by analyzing gene sets to determine whether the gene set shows a statistically significant difference between the two biological states. Within the "Molecular Signatures Database" of c2.cp.kegg. v6.2. Symbols by GSEA with a Java software, underlying mechanisms were studied. The random sample permutation number was set as 1,000, and the significance threshold $p < 0.050$.

Evaluation of Immune Cell Infiltration Level, Tumor Purity, and Stromal Content

ESTIMATE was performed to investigate the immune cell infiltration level (immune score), tumor purity, and stromal content for each sample (21). The single-sample GSEA (ssGSEA) score was used to quantify the activity and enrichment level of immune cell types, functions, and pathways applying the "limma," "GSVA," and "GSEABase" R packages to all samples. The "pheatmap" R package exhibited heatmap results. The Spearman correlation was utilized to identify the correlations between risk score and tumor purity as well as stromal score. The Wilcoxon rank-sum test was performed to assess the difference between high- and low-risk groups, and the result was exhibited by the "ggpubr" R package.

Correlation of the Prognostic Signature With Targets of Targeted Therapy and Immunotherapy

For the treatment of malignant tumors, targeted therapy and immunotherapy have become practical approaches.

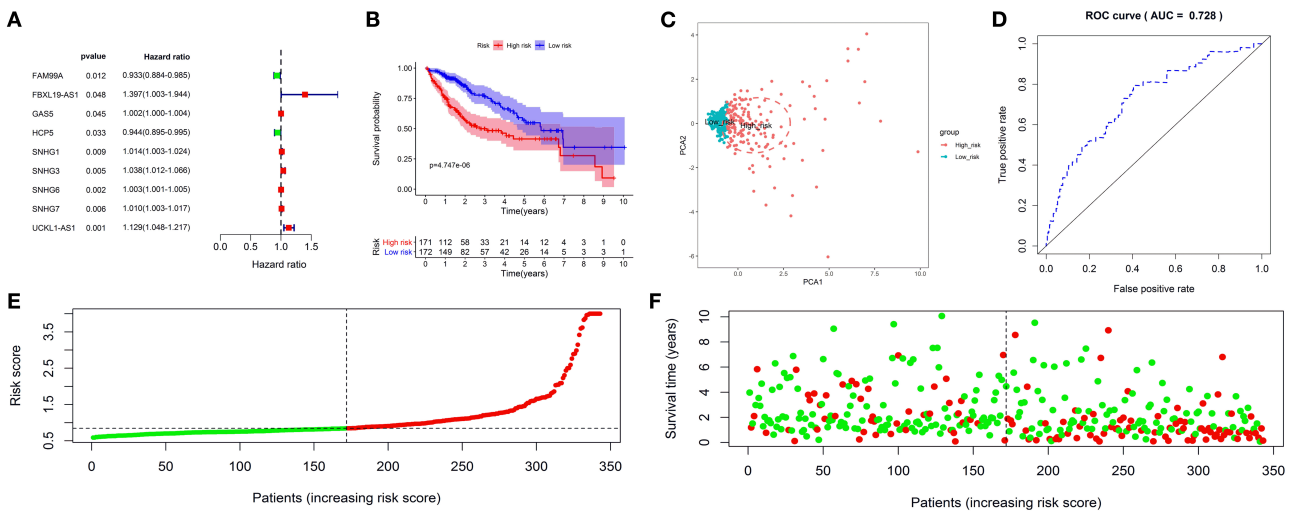

FIGURE 1 | Construction and validation of an autophagy-related long non-coding RNA (lncRNA) prognostic signature in the training cohort. The forest map showed that 9 autophagy-related lncRNAs might be correlated with overall survival based on the Kaplan–Meier (KM) method and univariate Cox regression analysis **(A)**. The KM survival analysis showed that patients in the high-risk group had a shorter overall survival time **(B)**. Principal component analysis (PCA) showed that the high- and low-risk patients were located in two distinct distribution clusters; the red dots represented high-risk patients, whereas the blue dots represented low-risk patients **(C)**. The time-dependent receiver-operating characteristic (ROC) curve showed that the area under the ROC curve (AUC) value for the prognostic signature was 0.728 **(D)**. The distribution of risk scores between low- and high-risk groups; The red dots represented high-risk patients, whereas the green dots represented low-risk patients **(E)**. The scatter plot showed the relationship between the risk score and survival time; the red dots represented high-risk patients, whereas the green dots represented low-risk patients **(F)**.

Now, the expression levels of immunotherapy and targeted therapy target genes between high- and low-risk groups were compared. We sought to predict therapeutic effectiveness using our risk score. The therapy targets were given as follows: programmed cell death 1 (PD-1, also known as PDCD1), vascular Endothelial Growth Factor Receptor (VEGFR1, also known as FLT1), Fms-like tyrosine kinase 3 (FLT3), VEGFR 3 (VEGFR3, also known as FLT4), platelet-derived growth factor receptor alpha (PDGFRA), platelet-derived growth factor receptor beta (PDGFRB), KIT proto-oncogene (KIT), ret proto-oncogene (RET), and MET proto-oncogene (MET), programmed cell death ligand 1 (PD-L1, also known as CD274), and mammalian target of rapamycin (mTOR). These correlations were drawn using the "ggpubr" R package, and the difference was evaluated by a Wilcoxon rank-sum test.

LncRNA Expression Analysis

First, the raw data of the GSE101728 and GSE62232D data sets were freely downloaded from the Gene Expression Omnibus (GEO) database. GSE101728 data set contained seven pairs of tumor and normal tissues. There were 81 tumor and 10 normal tissues in the GSE62232D data set. Differential analysis between the targeted lncRNAs expressions was further investigated in the tumor and normal tissues. The different expressions of these lncRNAs were further explored in the TCGA and ICGC databases. Differential analysis was visualized using the "ggpubr" R package. Finally, these targeted lncRNA expression levels were

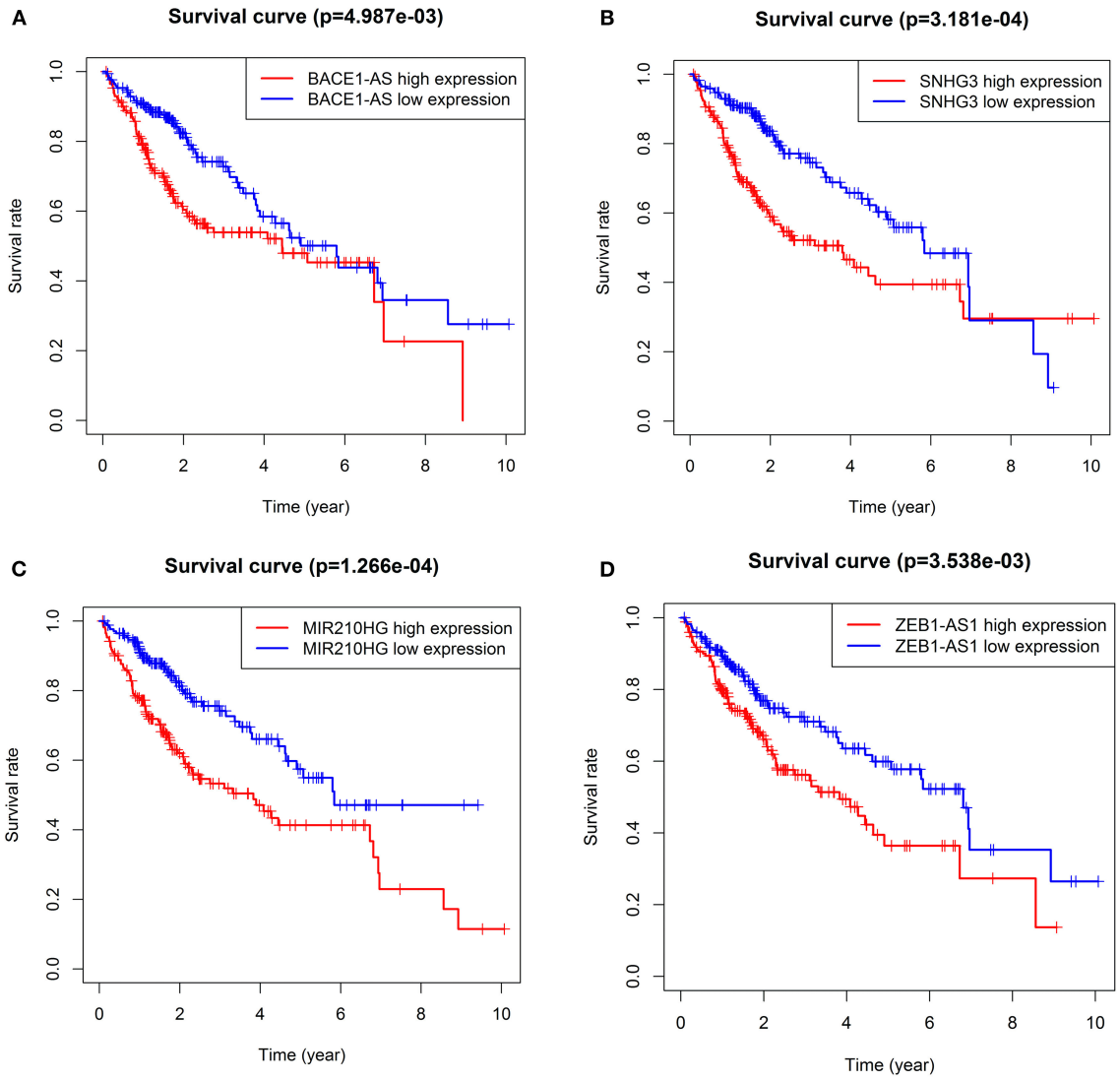

FIGURE 2 | KM survival curves showed the four autophagy-related lncRNA risk factors for hepatocellular carcinoma (HCC). BACE1-AS **(A)**, SNHG3 **(B)**, MIR210HG **(C)**, and ZEB1-AS1 **(D)**.

compared based on previous original studies' quantitative real-time PCR results (22–31).

Independence of the Prognostic Signature From Clinicopathological Parameters

Univariate and multivariate Cox proportional hazard regression analyses were performed by the "survival" R package to see if the predictive power of the prognostic signature was independent of clinicopathological parameters in both cohorts.

Establishment and Evaluation of a Nomogram for Survival Prediction

To accurately predict the 1-, 3-, and 5-year overall survival (OS) probability, a prognostic nomogram was constructed by integrating clinicopathological parameters with independently predictive value and the risk score. Harrell's concordance index (C-index) was performed to evaluate the predictive accuracy. C-index ranges from 0.5 (no predictive power) to 1 (perfect prediction). Calibration plots were used to assess the nomogram's performance characteristics. Each patient would get the total points from the nomogram, namely Nomo-score, and patients were classified into three risk groups using the tertiles of Nomo-scores as the cut-off values. The performance of the nomogram was further investigated *via* a KM curve analysis.

Statistical Analysis

All statistical analyses and figure generations were performed by the Perl programming language (version 5.30.2, http://www.perl.org) or R software (version 4.0.2, https://www.r-project.org/). A co-expression network was constructed using Cytoscape 3.6.1. A two-sided value of $p < 0.050$ was deemed statistically significant.

RESULTS

Identification of Autophagy-Related lncRNAs in Tissue Samples of a Patient With HCC

The expression levels of 232 autophagy-related genes were extracted from the TCGA and ICGC database. Subsequently, the co-expression network analysis identified autophagy-related lncRNAs with $|R| > 0.5$ and $p < 0.050$ as the selection criteria. Finally, autophagy-related lncRNAs from the two cohorts were intersected, yielding 19 autophagy-related lncRNAs.

Construction and Validation of an Autophagy-Related lncRNAs Prognostic Signature in the Training Cohort

Survival analysis showed that nine autophagy-related lncRNAs significantly correlated with OS (**Figure 1A**). Subsequently, a multivariate analysis revealed that four of nine autophagy-related lncRNAs were excellent candidates for constructing a prognostic signature. The candidates were BACE1-AS, SNHG3, MIR210HG,

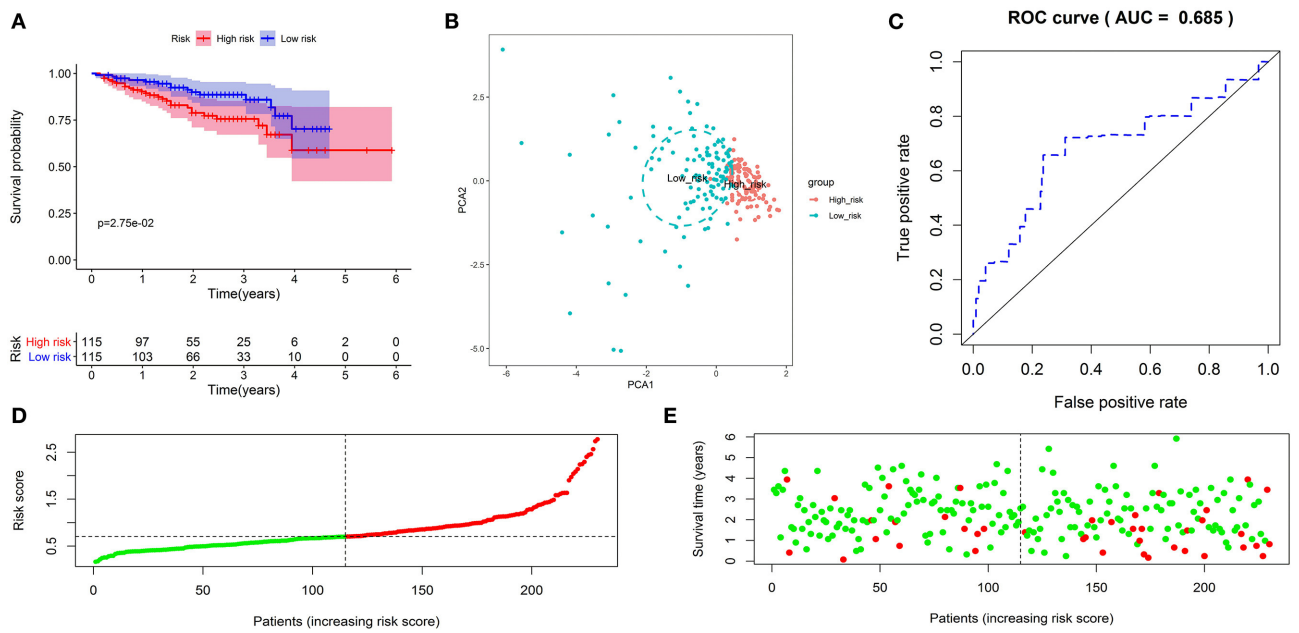

FIGURE 3 | Validation of the autophagy-related lncRNA prognostic signature in the validation cohort. KM survival curves showed that patients in the high-risk group had significantly poorer overall survival **(A)**. Principal component analysis showed that patients with different risks were significantly divided into two clusters; the red dots represented high-risk patients, whereas the blue dots represented low-risk patients **(B)**. The time-dependent ROC curve showed that the AUC value for the prognostic signature was 0.685 **(C)**. The distribution of risk scores between low- and high-risk groups; The red dots represented high-risk patients, whereas the green dots represented low-risk patients **(D)**. The scatter plot showed the relationship between the risk score and survival time; The red dots represented high-risk patients, whereas the green dots represented low-risk patients **(E)**.

and ZEB1-AS1. The four lncRNAs have been confirmed to be risk factors (**Figure 2**). Following that, these autophagy-related lncRNAs were utilized to construct the following prognostic signature: risk score = (0.142 × the expression level of BACE1-AS) + (0.032 × the expression level of SNHG3) + (0.067 × the expression level of MIR210HG) + (0.112 × the expression level of ZEB1-AS1).

A risk score was assigned to each subject. The median risk score was the cut-off point to separate patients into high- or low-risk groups. The KM survival analysis revealed that the high-risk group had a shorter OS than the low-risk group ($p < 0.050$) (**Figure 1B**). Patients in the high-risk group had 1-, 3-, and 5-year survival rates of 75.60, 49.90, and 41.50%, whereas patients in the low-risk group had 1-, 3-, and 5-year survival rates of 93.40, 76.30, and 57.00%. The PCA analysis revealed that the high- and low-risk patients were located in the two distinct distribution clusters (**Figure 1C**). Time-dependent ROC curve analysis further showed that the AUC value for the prognostic signature was 0.728 (**Figure 1D**). The distribution ranking of patients' risk scores in different groups was shown in **Figure 1E**. The correlations of dead status with the risk score was shown using the scatter dots plot (**Figure 1F**). These results

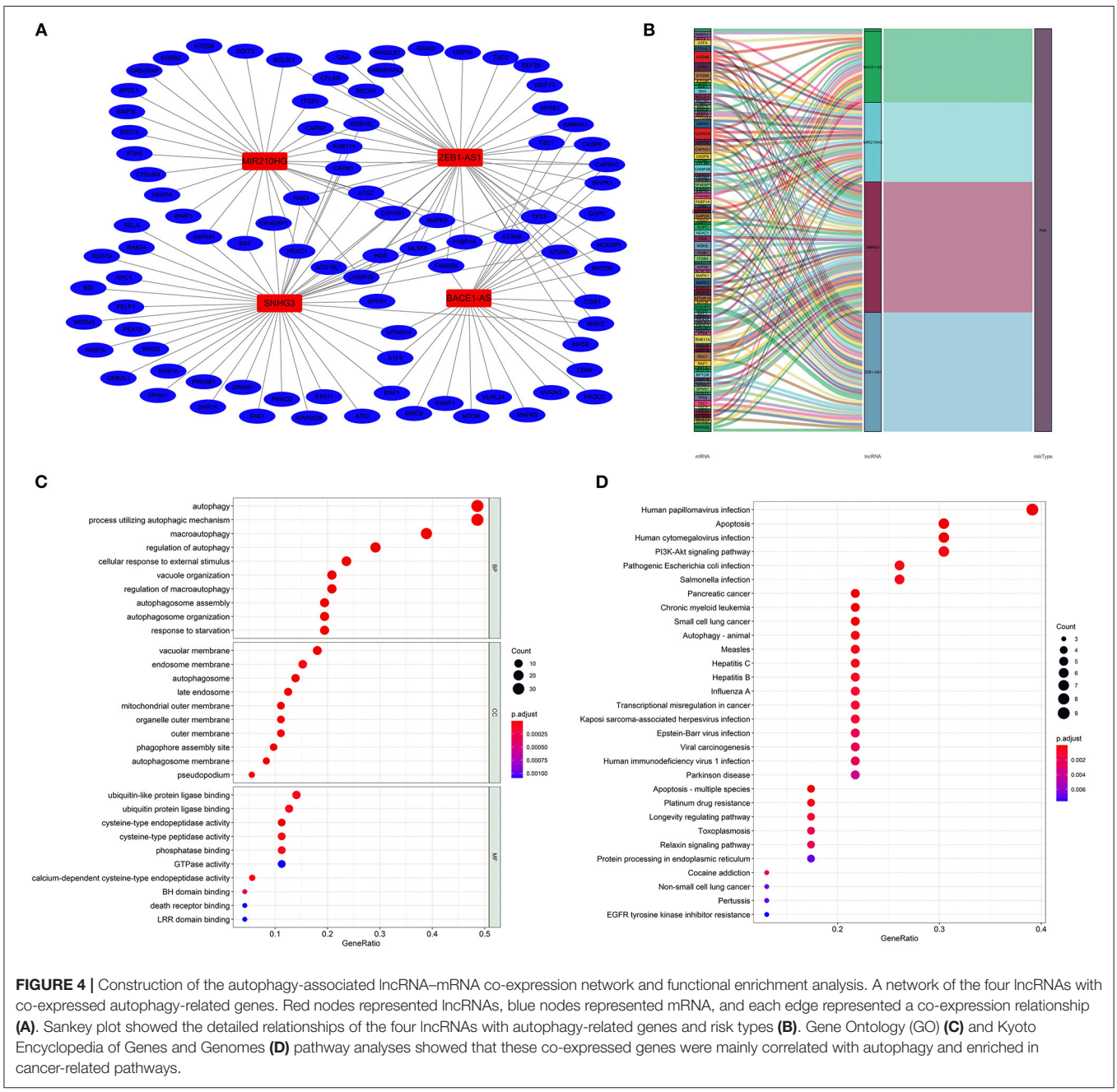

FIGURE 4 | Construction of the autophagy-associated lncRNA–mRNA co-expression network and functional enrichment analysis. A network of the four lncRNAs with co-expressed autophagy-related genes. Red nodes represented lncRNAs, blue nodes represented mRNA, and each edge represented a co-expression relationship **(A)**. Sankey plot showed the detailed relationships of the four lncRNAs with autophagy-related genes and risk types **(B)**. Gene Ontology (GO) **(C)** and Kyoto Encyclopedia of Genes and Genomes **(D)** pathway analyses showed that these co-expressed genes were mainly correlated with autophagy and enriched in cancer-related pathways.

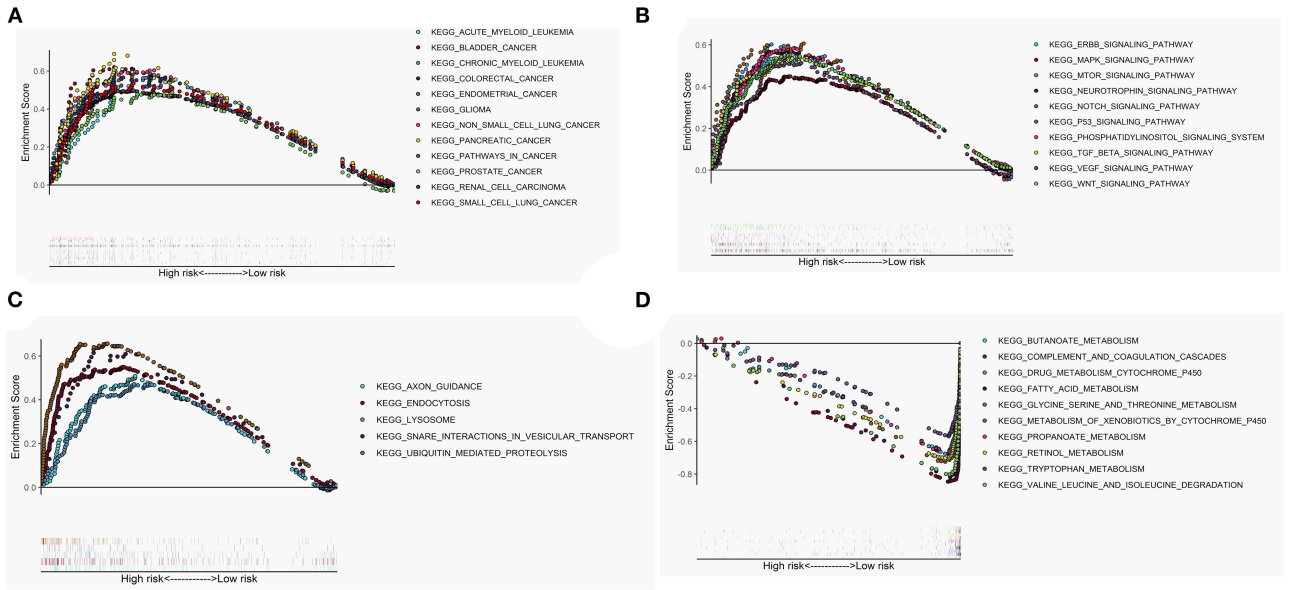

FIGURE 5 | Gene set enrichment analysis (GSEA) between high- and low-risk groups. The altered gene sets in the high-risk group were mainly enriched in tumorigenesis **(A)**, cancer- and autophagy-related signaling pathways **(B)**, and physical actions of autophagy **(C)**. The protective pathways involved in metabolism were significantly enriched in the low-risk group **(D)**.

demonstrated that patients with a higher risk score suffered from a shorter survival duration and a poorer survival rate.

Validation of the Prognostic Signature in the Validation Cohort

Subsequently, we further investigate the predictive value of the prognostic signature in the validation cohort. The same formula calculated the risk score and separated patients into low- or high-risk groups. As expected, the KM curve demonstrated that patients in the high-risk group had a shorter survival time (**Figure 3A**). PCA showed that patients of different risk groups were significantly split into two clusters (**Figure 3B**). Moreover, a time-dependent ROC curve was generated to validate the prognosis accuracy (AUC = 0.685), confirming the robust prediction of the signature (**Figure 3C**). **Figures 3D,E** depict the distribution of risk scores with survival status. These findings supported the hypothesis that the prognostic signature could reliably predict the prognosis of patients with HCC.

Construction of the Autophagy-Associated lncRNA–mRNA Co-expression Network and Functional Enrichment Analysis

The lncRNA–mRNA co-expression network was constructed to probe potential functions. As shown in **Figure 4A**, the network contains 4 lncRNAs, 91 mRNAs, and 141 lncRNA–mRNA pairs. The detailed correlations of these lncRNAs with genes and risk types are also shown on the Sankey plot (**Figure 4B**). The GO and KEGG pathway analyses demonstrated that the genes encoded by these mRNAs were mainly correlated with autophagy and enriched in pathways in cancer (**Figures 4C,D**).

Gene Set Enrichment Analysis

The GSEA performed a functional annotation. The GSEA results revealed that the altered gene sets in the high-risk group were directly involved in carcinogenesis and progression (**Figure 5A**). Besides, differentially expressed genes between the two risk groups were mainly enriched in the autophagy-associated and tumor-related pathways, including ERBB signaling pathway, MAPK signaling pathway, mTOR signaling pathway, VEGF signaling pathway, WNT signaling pathway, and P53 signaling pathway (**Figure 5B**). In addition, the altered expression genes in the high-risk group were discovered to be involved in autophagy's physical effects (**Figure 5C**). In contrast, the protective metabolic pathways were significantly enriched in the low-risk group (**Figure 5D**).

Comparison of the Immune Activity and Tumor Microenvironment Between High- and Low-Risk Groups

We looked at 29 immune-associated gene sets that represented various immune cell types, functions, and pathways. The activity and enrichment levels of immune cell types, functions, and pathways in each sample were measured using the ssGSEA score. Then, the enrichment scores were compared between low- and high-risk groups. **Figure 6** showed that the low-risk group exhibited more significant immune cell infiltration than the high-risk group. Furthermore, the 13 immunological pathways in the low-risk group were more active than those in the high-risk group. When comparing the tumor purity and stromal scores between the two risk groups, we discovered that the stromal score was significantly higher in the low-risk group. In contrast,

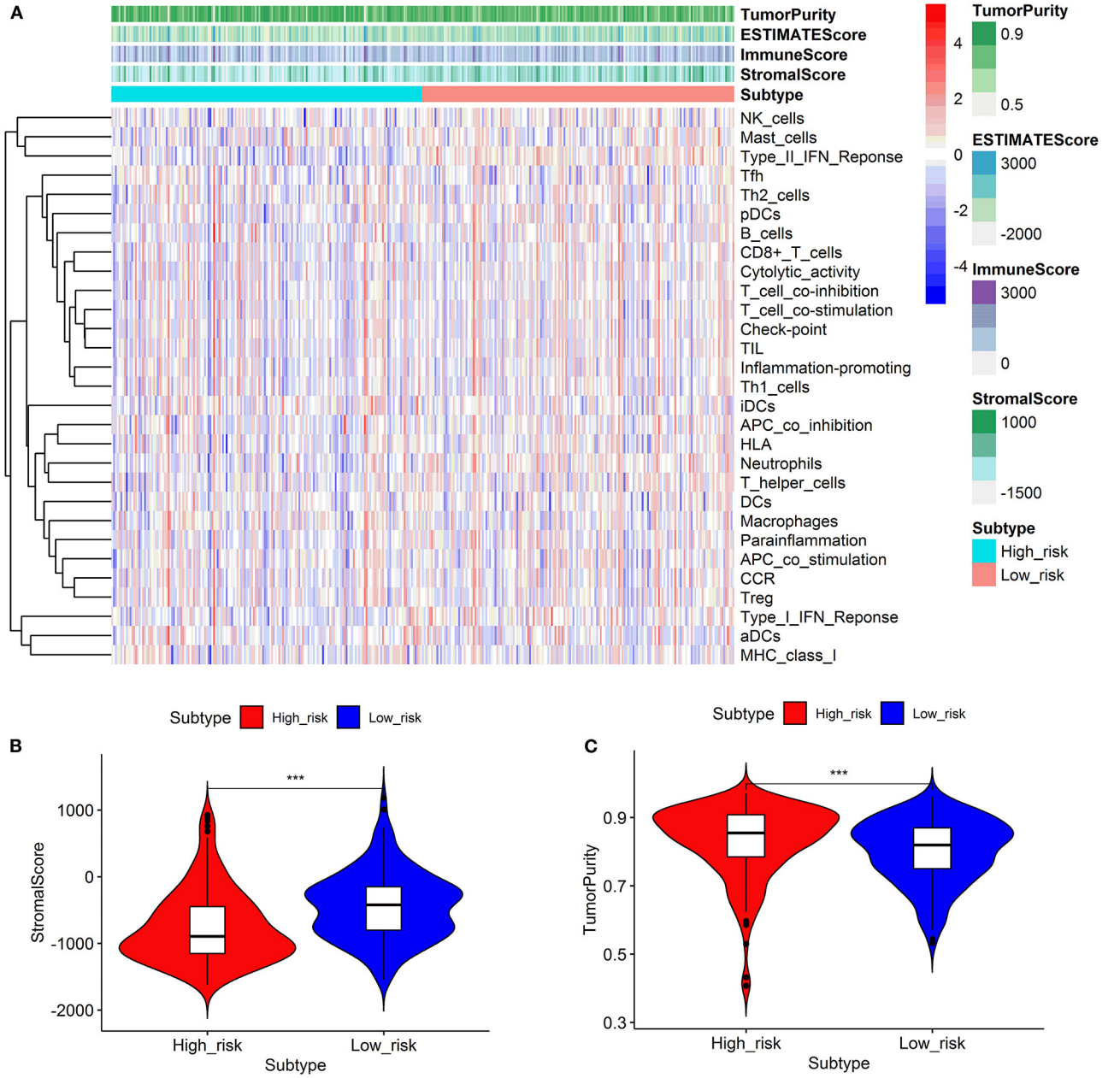

FIGURE 6 | Comparison of the immune activity and tumor microenvironment between high- and low-risk groups. Hierarchical clustering showed that the low-risk group exhibited more immune cell infiltration than the high-risk group **(A)**. Comparison of stromal scores between the high- and low-risk groups **(B)**. Comparison of the tumor purity between high- and low-risk groups **(C)**. ***$p < 0.001$.

the tumor purity trended in the opposite direction, with tumor purity increasing from low risk to high risk (Kruskal–Wallis test, $p < 0.001$).

Effectiveness Prediction of Immunotherapy and Targeted Therapy With the Prognostic Signature

As shown in **Figure 7**, the expression levels of immunotherapy and targeted therapy target genes were compared between high- and low-risk categories. The expression levels of PDGFRB, PDCD1, KIT, FLT3, and FLT4 between the two risk groups were significantly different. The PDGFRB, PDCD1, and KIT expression levels were higher in the high-risk group, while FLT3 and FLT4 expressions were higher in the low-risk group. Therefore, immunotherapy and targeted therapy medicines targeting PDGFRB, PDCD1, and KIT may be more effective in patients with HCC with higher risk scores.

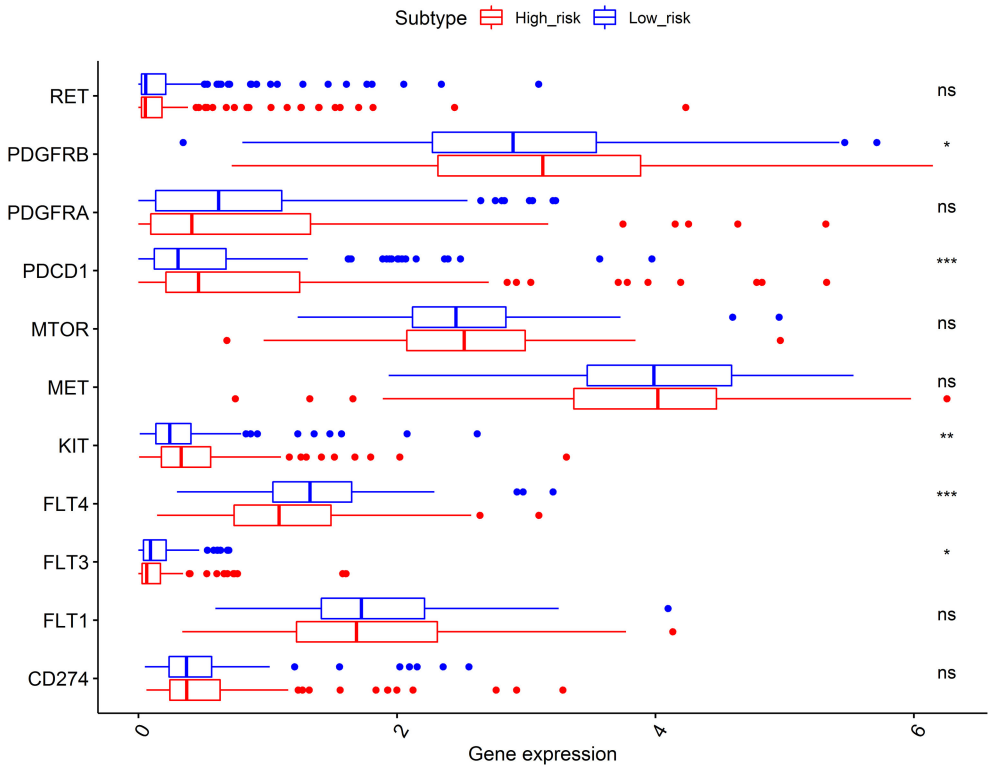

FIGURE 7 | Comparisons of the expression levels of immunotherapy and targeted therapy target genes between high- and low-risk subgroups. The platelet-derived growth factor receptor beta (PDGFRB), programmed cell death 1 (PDCD1), and KIT proto-oncogene (KIT) expression levels were higher in the high-risk group, while the FLT3 and FLT4 expressions were higher in the low-risk group. *$p < 0.050$, **$p < 0.010$, and ***$p < 0.001$; ns, no significance.

Expression Analysis of lncRNAs

Subsequently, a difference in the expression of the four targeted lncRNAs was investigated between the tumor tissues and normal tissues. In the GSE101728 data set, the SNHG3 and ZEB1-AS1expression levels were higher in the tumor tissues, while the expression levels of BACE1-AS and MIR210HG showed no difference (**Figure 8A**). In the analysis results of the GSE62232 data set, ZEB1-AS1 was highly expressed in tumor tissues (**Figure 8B**). The analysis of HCC samples from the TCGA and ICGC databases both exhibited that BACE1-AS, SNHG3, and ZEB1-AS1 were highly expressed in the tumor tissue, and there was no significant difference in MIR210HG expression (**Figures 8C,D**). What is more, we have found ten original studies involving the differential expressions of the four targeted lncRNAs between normal and tumor tissues. Interestingly, the lncRNAs expression levels in HCC tumor tissues were significantly higher than those in the normal control group in each study (**Supplementary Figure S1**).

Correlation Analysis of the Autophagy-Related lncRNA Prognostic Signature With Clinicopathological Features

Correlation analysis was done to investigate the clinical value of the prognostic signature in different subgroups stratified by the clinicopathological characteristics of the patients. As shown in **Table 1**, patients with high-risk scores were prone to be found in those with greater creatinine or Alpha-fetoprotein (AFP). The increasing risk score appeared to be highly connected with advanced T and TNM stages, suggesting that the prognostic signature may be considerably associated with HCC progression.

Prognostic Value of the Autophagy-Related lncRNAs Signature Among Different Subgroups

Subgroup analysis was conducted to investigate the prognostic value of the autophagy-related lncRNA signature among different subgroups stratified by clinicopathological variables. As indicated in **Table 2**, the prognostic signature performs better in male patients without liver cirrhosis and family history, whereas obese individuals in poor physical condition may benefit more from the prognostic signature. The prognostic signature seemed to be more applicable to patients with relatively lower serum AFP, albumin, and creatinine levels in terms of laboratory index. Besides, the prognostic signature showed excellent predictive power independent of various clinicopathological features such as gender, age, alcohol consumption history, tumor stage, and histological grade.

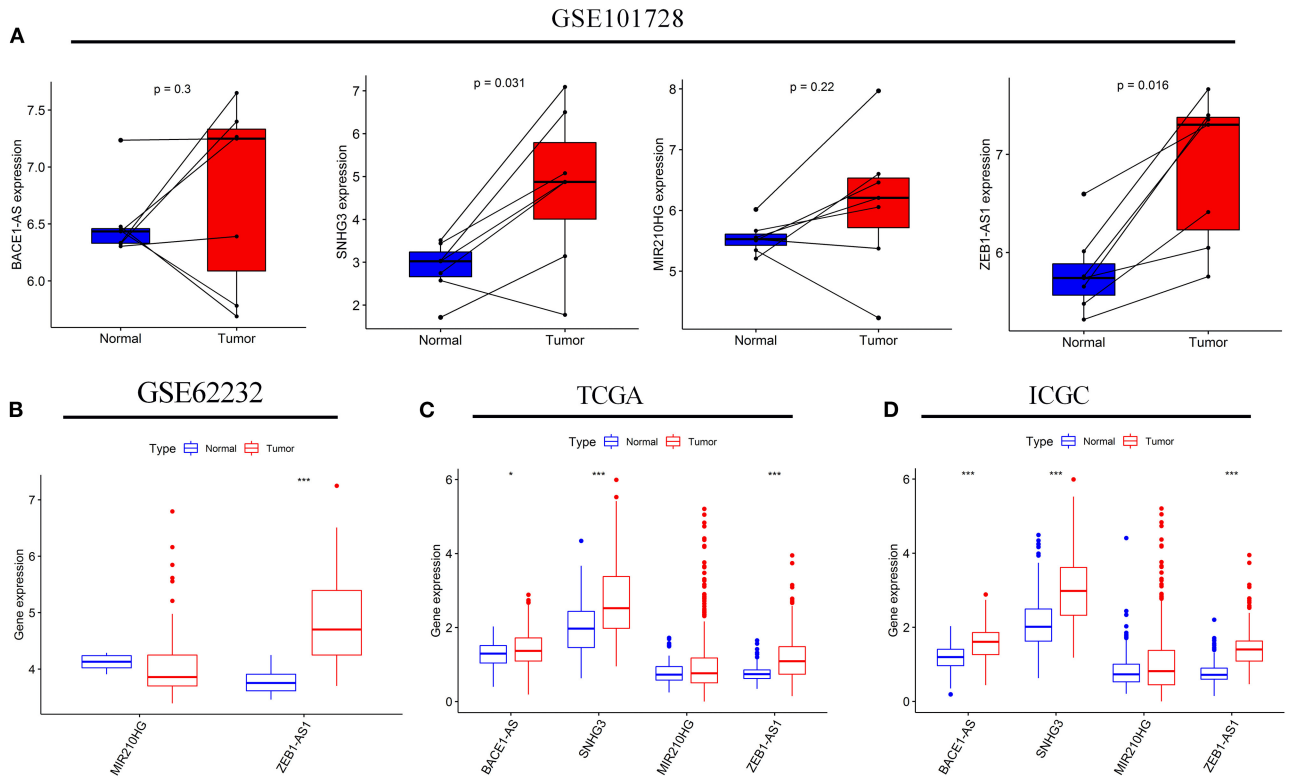

FIGURE 8 | The expression levels of the four autophagy-related lncRNAs in HCC tissues and adjacent normal tissues. The expressions of the four lncRNAs in seven paired HCC and adjacent normal tissues from the GSE10728 data set **(A)**. Comparison of the MIR210HG and ZEB1-AS1 expressions between 10 normal tissues and 81 tumor tissues in the GSE62232 data set **(B)**. Comparison of the four autophagy-related lncRNAs expressions in normal and tumor tissues from the Cancer Genome Atlas (TCGA) **(C)** and International Cancer Genome Consortium (ICGC) **(D)** cohorts. *$p < 0.050$ and ***$p < 0.001$.

Independence of the Prognostic Signature From Clinicopathological Parameters

Univariate and multivariate Cox regression analyses were performed to assess whether the prognostic signature was a prognostic factor independent of clinicopathological features in both cohorts. As shown in **Figure 9A**, univariate analysis indicated that ECOG [HR: 2.390; 95% CI: 1.894–3.016; $p < 0.001$], TNM stage [HR: 1.784; 95% CI: 1.446–2.202; $p < 0.001$], liver cirrhosis [HR: 2.426; 95% CI: 1.516–3,881.; $p < 0.001$], and the risk score [HR: 1.539; 95% CI: 1.339–1.769; $p < 0.001$] were significantly correlated with OS in the training cohort. T stage was not enrolled in multivariate Cox regression modeling because the TNM stage was derived based on the T, N, and M stages. ECOG [HR: 1.680; 95% CI: 1.168–2.417; $p = 0.005$], liver cirrhosis [HR: 1.972; 95% CI: 1.060–3.669; $p = 0.032$], and the risk score [HR: 1.385; 95% CI: 1.121–1.710; $p = 0.002$] were ruled out as independently prognostic factors in multivariate analysis (**Figure 9B**). Besides, the risk score was also proven to be an independently prognostic factor in the validation cohort (**Figures 9C,D**).

Establishment of a Nomogram for OS Prediction

An OS nomogram was formulated based on three independently prognostic factors in the training cohort. Furthermore, the 1-, 3-, and 5-year OS rate was displayed in the nomogram (**Figure 10A**). The C-index value for OS prediction was 0.739. Calibration plots further identified that the nomogram performed well in predicted 1-, 3-, and 5-year survival probabilities with an ideal model, indicating that the nomogram was perfectly calibrated to predict OS at assessing the performance characteristics (**Figures 10B–D**). Each patient with complete clinical information on the ECOG score and liver cirrhosis (or not) would get the Nomo-score, and patients were classified into three risk categories based on the tertiles of Nomo-scores. The KM curve revealed significant variations across high-, intermediate-, and low-risk groups ($p < 0.001$) (**Figure 10E**).

DISCUSSION

Hepatocellular carcinoma is one of the most lethal and prevalent primary hepatic malignant neoplasms worldwide. Despite great improvements in diagnosis and multimodal therapies, the survival benefit remains limited due to high heterogeneity (32). Clinically, histological grade, tumor stage, molecular subtype, and serum indicator prognostic effects were evaluated (33). However, such clinicopathological characteristics were unable to provide predictive value, resulting in inaccurate prognosis judgment. According to the situation, certain high-risk patients may encounter tumor cell uncontrollable growth

TABLE 1 | Clinical value of the autophagy-related lncRNA prognostic signature for HCC.

Characteristics	Group	Risk score		
		N	Mean	*P* value
Age (years)	<60	157	1.154	0.787
	≥60	186	1.183	
Gender	Male	233	1.114	0.120
	Female	110	1.289	
Alcohol consumption	Present	109	1.106	0.482
	Absent	218	1.185	
Liver cirrhosis	Present	131	1.254	0.353
	Absent	65	1.112	
Family history	Present	104	1.063	0.165
	Absent	194	1.216	
Histological grade	G1+G2	211	1.159	0.699
	G3+G4	138	1.202	
Albumin (g/dl)	<4.0	128	1.055	0.633
	≥4.0	153	1.105	
Creatinine (mg/dl)	<1.1	189	0.926	**0.005**
	≥1.1	92	1.169	
BMI (kg/cm2)	<25	143	1.163	0.776
	≥25	152	1.196	
AFP (ng/ml)	≤200	187	0.977	**0.006**
	>200	73	1.408	
ECOG	=0	156	1.094	0.196
	>0	117	1.253	
T stage	I+II	252	1.042	**0.002**
	III+IV	88	1.551	
TNM stage	I+II	238	1.025	**0.001**
	III+IV	83	1.576	

Bold font represents P < 0.05 and the relevant variables are statistically significant. HCC, hepatocellular carcinoma; HBV, hepatitis B virus; HCV, hepatitis C virus; AFP, alpha-fetoprotein; BMI, body mass index; ECOG, eastern cooperative oncology group; TNM, tumor node metastasis; SD, standard deviation.

TABLE 2 | Prognostic value of the autophagy-related lncRNA prognostic signature in different subgroups stratified by clinicopathological variables.

Characteristics	Group	Low/High	HR (95% CI)	*P* value
Overall		172/171	2.249 (1.559–3.256)	**<0.001**
Age	<60	76/81	2.832 (1.586–5.056)	**<0.001**
	≥60	95/91	1.754 (1.079–2.853)	**0.023**
Gender	Male	124/109	3.125 (1.925–5.072)	**<0.001**
	Female	47/63	1.290 (0.7285–2.285)	0.383
Alcohol consumption	Present	60/49	3.718 (1.921–7.194)	**<0.001**
	Absent	106/112	1.794 (1.136–2.833)	**0.012**
Liver cirrhosis	Present	26/39	2.108 (1.010–4.396)	**0.047**
	Absent	70/67	1.922 (0.962–3.836)	0.064
Family history	Present	63/41	1.828 (1.019–3.277)	0.043
	Absent	83/111	2.944 (1.688–5.133)	**<0.001**
Histological grade	G1+G2	102/109	2.154 (1.338–3.466)	**0.002**
	G3+G4	66/61	2.549 (1.411–4.607)	**0.002**
Albumin (g/dl)	<4.0	76/52	1.908 (1.040–3.501)	**0.037**
	≥4.0	75/78	1.644 (1.893–3.025)	0.110
Creatinine (mg/dl)	<1.1	98/91	1.721 (1.034–2.862)	**0.037**
	≥1.1	53/39	1.736 (0.825–3.656)	0.147
BMI (kg/cm2)	<25	70/73	1.670 (0.917–3.040)	0.093
	≥25	77/75	3.805 (2.105–6.787)	**<0.001**
AFP (ng/ml)	≤200	116/171	1.985 (1.156–3.408)	**0.013**
	>200	21/52	2.041 (0.757–5.506)	0.159
ECOG	=0	91/65	1.878 (0.924–3.817)	0.081
	>0	53/64	3.016 (1.679–5.416)	**<0.001**
T stage	I+II	140/112	2.064 (1.280–3.329)	**0.003**
	III+IV	29/59	1.926 (1.054–3.520)	**0.033**
TNM stage	I+II	135/103	1.915 (1.155–3.174)	**0.012**
	III+IV	26/57	2.025 (1.046–3.921)	**0.036**

Bold font represents P < 0.050 and the relevant variables are statistically significant. HBV, hepatitis B virus; HCV, hepatitis C virus; AFP, alpha-fetoprotein; BMI, body mass index; ECOG, eastern cooperative oncology group; TNM, tumor node metastasis; HR, hazard ratio.

due to insufficient treatment, while low-risk patients may receive excessive treatment, resulting in long-term toxicity and morbidity. Therefore, reliable genetic signatures or biomarkers as prognostic predictors or therapeutic targets are of significance for HCC.

The overutilization of amino acids such as tryptophan, aerobic glycolysis, tricarboxylic acid (TCA) cycle, glutamine, arginine, defective mitochondrial bioenergetics, and oxidative stress phosphorylation are all involved in malignant cell progression and extinction (34, 35). Moreover, these energy metabolisms are dramatically associated with autophagy progress. Hence, knowing the specifics and direct links between autophagy and tumor progression could provide a solid foundation for creating drugs targeting these pathways and ultimately curing malignancies. Following decades of researches on prognostic gene biomarkers of tumor-related events such as microRNAs and mRNAs, lncRNAs have recently aroused much attention. The roles of lncRNA in carcinogenesis and malignant tumor growth have been gradually revealed. The prognostic value of lncRNA also has been extensively explored. However, there has yet to be a systematic method for identifying the autophagy-related lncRNAs signature that might be used to predict the prognosis of patients with HCC. Hence, developing an autophagy-related lncRNAs signature to predict the clinical outcome is critical for patients with HCC.

In the study, we used the expression profile of HCC patients' tumor tissue from the TCGA and ICGC databases to investigate the prognostic usefulness of autophagy-related lncRNAs and develop a prognostic signature. We first identified 19 autophagy-related lncRNAs based on the lncRNAs and autophagy-related gene co-expression network. After univariate and multivariate Cox regression analyses, four autophagy-related lncRNAs, including BACE1-AS, SNHG3, MIR210HG, and ZEB1-AS1, were selected to establish a prognostic signature. Each patient obtained a risk score. All patients were divided into high or low risks based on the median value of risk scores. We also discovered that patients with varied risks were considerably split into two groups, with the high-risk group having a shorter

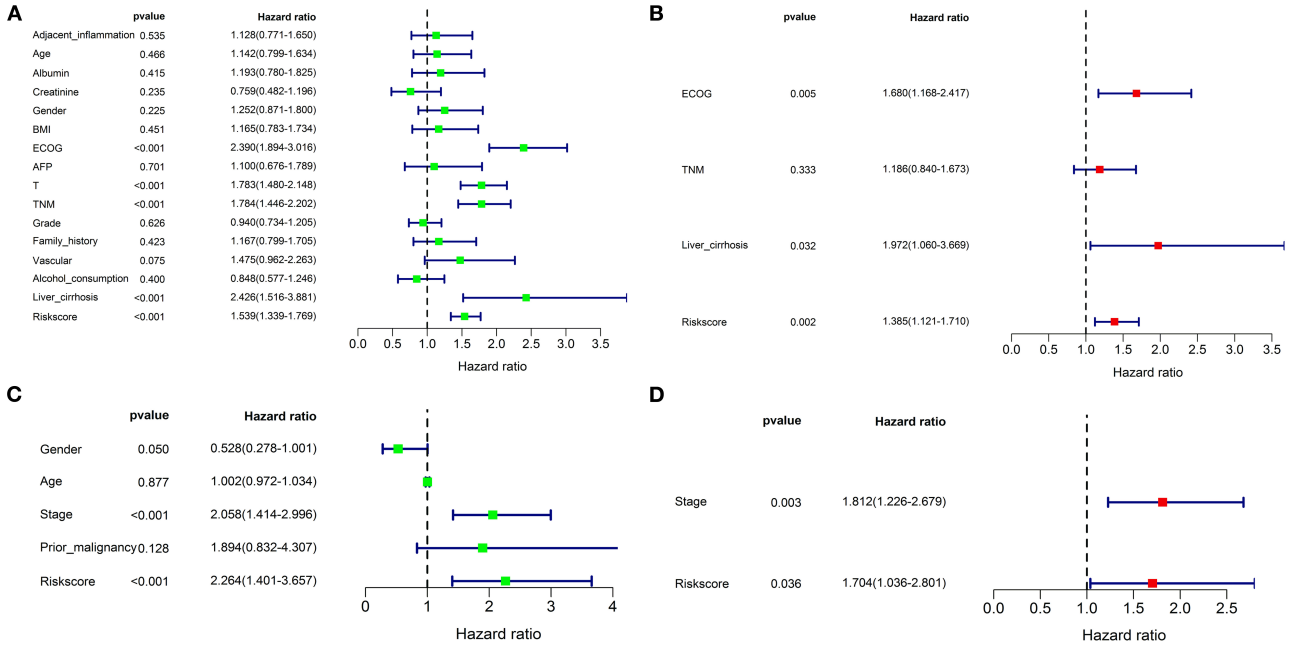

FIGURE 9 | Univariate and multivariate Cox regression analyses for the risk score combined with the clinical characteristics. Univariate analysis for the TCGA **(A)** and ICGC **(C)** cohorts. These parameters with a prognostic value were enrolled into the multivariate Cox regression analysis. The result showed that the risk score was an independently prognostic indicator in TCGA **(B)** and ICGC **(D)** cohorts.

survival time. The ROC curve analysis further confirmed the prognostic accuracy of the signature. When the predictive value of the prognostic signature was investigated in the ICGC validation cohort, similar results were achieved. Univariate and multivariate regression models showed that the risk score showed excellent predictive power independent of all clinicopathological characteristics in both cohorts. Hence, the autophagy-related lncRNA prognostic signature showed powerful potential for clinical applications.

When associations of the risk score and clinicopathological characteristics were investigated, we found that the risk score was significantly related to advanced tumor and a higher level of serum AFP. The explanation supported the findings that improper autophagy contributed to a poor tumor microenvironment, allowing the malignant cell to proliferate, invade, and migrate quickly as the tumor advanced. These alterations might lead to a poor prognosis for patients with advanced cancers (14, 36). Subgroups analyses stratified by clinicopathological variables further verified the steadied predictive value of the prognostic signature.

The role of autophagy in cancer is debatable. As the understanding of autophagy continues to deepen, the role has been increasingly revealed. On one hand, autophagy could provide the essential circulating metabolic substrates and enzymes to respond to various poor circumstances such as tumor microenvironment; on the other hand, inappropriate autophagy also promotes malignant cell rapid growth, particularly in the advanced tumor. LncRNAs' roles have recently been discovered to mediate tumorigenesis, progression, metastasis, and treatment resistance by regulating genes or microRNAs. The present study identified four autophagy-related lncRNAs to establish a prognostic signature. Previous studies confirmed that BACE1-AS was significantly associated with the prognosis of patients with cancer (22, 23). BACE1-AS could also enhance autophagy-related neuronal damage *via* the miR-214-3p/ATG5 signaling axis in Alzheimer's disease (37). Additionally, the functions of SNHG3 in cancer have been steadily revealed. A growing body of evidence showed that SNHG3 appeared to influence tumor formation and progression by modulating autophagy-related microRNAs, genes, or pathways (24–26). MIR210HG (27, 38–40) and ZEB1-AS1 (28–31, 41, 42) have also been found to alter tumorigenesis, progression, and tumor metastasis, resulting in a poor prognosis for patients with cancer. Unquestionably, the established prognostic signature based on the four robust autophagy-related lncRNAs had a more excellent predictive value. Subsequently, we also identified the genes governed by the four autophagy-related lncRNAs and established a lncRNA–mRNA co-expression network. GO and KEGG functional enrichment analyses showed that these genes were mainly enriched in autophagy- and tumor-related signaling pathways.

The GSEA functional enrichment analysis showed that autophagy- and cancer-related pathways were overrepresented in the high-risk group. The altered gene sets in the high-risk group were discovered to be engaged in the autophagy-associated and tumor-related pathways as well as the autophagy's physical effects. At the same time, the protective pathways involved in various metabolisms were significantly enriched in the low-risk group. As a result, our findings added to

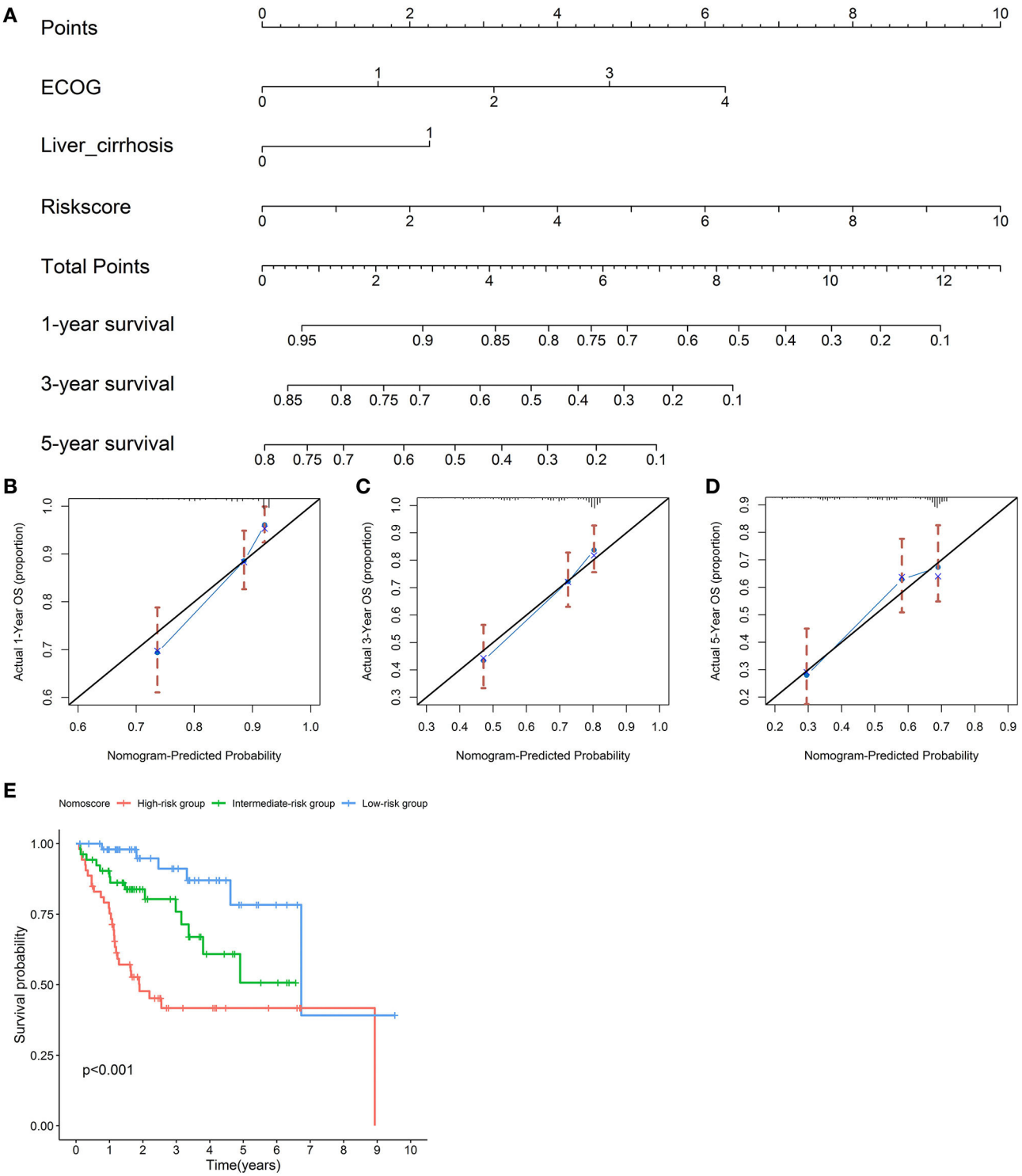

FIGURE 10 | Establishment of a nomogram for overall survival (OS) prediction. A nomogram for predicting 1-, 3-, and 5-year OS was constructed based on three independent prognostic factors: the risk score, ECOG, and liver cirrhosis. The detailed 1-, 3-, and 5-year OS rates were displayed in the nomogram **(A)**. Calibration plots showed that the nomogram performed well in the predicted 1- **(B)**, 3- **(C)**, and 5-year **(D)** survival probabilities with an ideal model. The black line represents the "ideal" line of a perfect match between the predicted and observed survival. The blue line indicates the performance of the proposed nomogram. X-axis is the nomogram predicted probability of survival, and Y-axis is actual survival. Blue dots are subcohorts of the data set; red vertical bars represent a 95% CI. KM curves of three risk subgroups stratified by the tertiles of Nomo-scores showed the healthy performance of the nomogram **(E)**.

the growing body of evidence showing that autophagy is a crucial regulator of oncogenesis and development. We also concluded that the four autophagy-related lncRNAs could be therapeutic targets. Moreover, we looked at immune cell infiltration and antitumor immunological activity between high- and low-risk groups. The results revealed that patients with low-risk scores had more immune cell infiltrations and antitumor immune activities, showing that the high-risk group's immune functions were overall impaired. The increasing antitumor immune activity could explain why patients with low risk had well clinical outcomes. Moreover, we found that the stromal score was greater in the low-risk group, whereas the tumor purity increased from the low- to the high-risk subgroup. The results further demonstrated that the poor prognosis might be due to an unbenefited tumor microenvironment. Currently, immunotherapy and targeted therapy are hot fields of investigation. The prognostic signature also revealed that the risk score was significantly associated with the effectiveness of immunotherapy and targeted therapies, thus validating the signature's prediction accuracy.

Nomogram is an effective and reliable clinical tool that can generate a probabilistic forecast for an individual patient. To improve prognosis prediction for patients with HCC, we constructed an OS nomogram based on independently prognostic factors. Calibration plots showed that the predicted 1-, 3-, and 5-year survival rates were comparable with the actual observation. A high C-index indicated robust discrimination, implying that it might function as a predive tool. However, more research is needed to confirm the prognostic signature in a larger number of patients and to reveal the potential molecular mechanisms of the four autophagy-related lncRNAs in HCC.

CONCLUSION

In conclusion, although the autophagy-related lncRNAs prognostic signature was a promising tool for predicting the prognosis of patients with HCC, there is a need for further studies to evaluate the device. The prognostic signature might aid in better understanding the role of autophagy in carcinogenesis and progression. The four autophagy-related lncRNAs might be used as potential biomarkers and therapeutic targets for patients with HCC. The prognostic signature and nomogram could be used to stratify patients at risk, aiding clinicians in treatment optimization and clinical decision-making.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YD and FZ contributed to conception and design, the acquisition of data, the analysis and interpretation of data, and drafting the article. Z-GS took part in drafting the article or revising it critically for important intellectual content and gave final approval of the version to be published. SW contributed to conception and design, revising the article critically for important intellectual content, and gave final approval of the version to be published. All authors agreed to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

We thank all the Perl and R programming package developers. All authors sincerely acknowledge the contributions from the TCGA project and the ICGC depository.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.762570/full#supplementary-material

Supplementary Table 1 | A list of 232 autophagy genes.

Supplementary Excel S1 | Clinical data of patients with hepatocellular carcinoma (HCC) from the Cancer Genome Atlas (TCGA) database.

Supplementary Excel S2 | Clinical data of patients with HCC from the International Cancer Genome Consortium (ICGC) database.

Supplementary Figure S1 | The expression levels of the four autophagy-related long non-coding RNAs (lncRNAs) in hepatocellular carcinoma (HCC) tissues and adjacent normal tissues were detected by RT-PCR in each previous study. BACE1-AS **(A–B)**, SNHG3 **(C–E)**, MIR210HG **(F)**, and ZEB1-AS1 **(G–J)**.

REFERENCES

1. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology.* (2012) 142:1264–73 e1. doi: 10.1053/j.gastro.2011.12.061
2. Shetty S, Sharma N, Ghosh K. Epidemiology of hepatocellular carcinoma (HCC) in hemophilia. *Crit Rev Oncol Hematol.* (2016) 99:129–33. doi: 10.1016/j.critrevonc.2015.12.009
3. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet.* (2012) 379:1245–55. doi: 10.1016/B978-1-4377-1454-8.00080-1
4. Lin S, Hoffmann K, Schemmer P. Treatment of hepatocellular carcinoma: a systematic review. *Liver Cancer.* (2012) 1:144–58. doi: 10.1159/000343828
5. Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives. *Gut.* (2014) 63:844–55. doi: 10.1136/gutjnl-2013-306627
6. Fong ZV, Tanabe KK. The clinical management of hepatocellular carcinoma in the United States, Europe, and Asia: a comprehensive and evidence-based comparison and review. *Cancer.* (2014) 120:2824–38. doi: 10.1002/cncr.28730
7. Amaravadi RK, Kimmelman AC, Debnath J. Targeting autophagy in cancer: recent advances and future directions. *Cancer Discov.* (2019) 9:1167–81. doi: 10.1158/2159-8290.CD-19-0292
8. Cheong H. Integrating autophagy and metabolism in cancer. *Arch Pharm Res.* (2015) 38:358–71. doi: 10.1007/s12272-015-0562-2
9. Rabinowitz JD, White E. Autophagy and metabolism. *Science.* (2010) 330:1344–8. doi: 10.1126/science.1193497
10. Dikic I, Johansen T, Kirkin V. Selective autophagy in cancer development and therapy. *Cancer Res.* (2010) 70:3431–4. doi: 10.1158/0008-5472.CAN-09-4027
11. Shintani T, Klionsky DJ. Autophagy in health and disease: a double-edged sword. *Science.* (2004) 306:990–5. doi: 10.1126/science.1099993

12. White E, DiPaola RS. The double-edged sword of autophagy modulation in cancer. *Clin Cancer Res.* (2009) 15:5308–16. doi: 10.1158/1078-0432.CCR-07-5023
13. Trejo-Solis C, Serrano-Garcia N, Escamilla-Ramirez A, Castillo-Rodriguez RA, Jimenez-Farfan D, Palencia G, et al. Autophagic and apoptotic pathways as targets for chemotherapy in glioblastoma. *Int J Mol Sci.* (2018) 19:3773. doi: 10.3390/ijms19123773
14. Janku F, McConkey DJ, Hong DS, Kurzrock R. Autophagy as a target for anticancer therapy. *Nat Rev Clin Oncol.* (2011) 8:528–39. doi: 10.1038/nrclinonc.2011.71
15. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* (2009) 10:155–9. doi: 10.1038/nrg2521
16. Zhu C, Zhang S, Fu H, Zhou C, Chen L, Li X, et al. Transcriptome and phytochemical analyses provide new insights into long non-coding RNAs modulating characteristic secondary metabolites of oolong tea (Camellia sinensis) in solar-withering. *Front Plant Sci.* (2019) 10:1638. doi: 10.3389/fpls.2019.01638
17. Barangi S, Hayes AW, Reiter R, Karimi G. The therapeutic role of long non-coding RNAs in human diseases: A focus on the recent insights into autophagy. *Pharmacol Res.* (2019) 142:22–9. doi: 10.1016/j.phrs.2019.02.010
18. Sun T. Long noncoding RNAs act as regulators of autophagy in cancer. *Pharmacol Res.* (2018) 129:151–5. doi: 10.1016/j.phrs.2017.11.009
19. Zhang J, Wang P, Wan L, Xu S, Pang D. The emergence of noncoding RNAs as Heracles in autophagy. *Autophagy.* (2017) 13:1004–24. doi: 10.1080/15548627.2017.1312041
20. Luan F, Chen W, Chen M, Yan J, Chen H, Yu H, et al. An autophagy-related long non-coding RNA signature for glioma. *FEBS Open Bio.* (2019) 9:653–67. doi: 10.1002/2211-5463.12601
21. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* (2013) 4:2612. doi: 10.1038/ncomms3612
22. Liu C, Wang H, Tang L, Huang H, Xu M, Lin Y, et al. LncRNA BACE1-AS enhances the invasive and metastatic capacity of hepatocellular carcinoma cells through mediating miR-377-3p/CELF1 axis. *Life Sci.* (2021) 275:119288. doi: 10.1016/j.lfs.2021.119288
23. Tian Q, Yan X, Yang L, Liu Z, Yuan Z, Zhang Y. Long non-coding RNA BACE1-AS plays an oncogenic role in hepatocellular carcinoma cells through miR-214-3p/APLN axis. *Acta Biochim Biophys Sin (Shanghai).* (2021) 53:1538–46. doi: 10.1093/abbs/gmab134
24. Zhang T, Cao C, Wu D, Liu L. SNHG3 correlates with malignant status and poor prognosis in hepatocellular carcinoma. *Tumour Biol.* (2016) 37:2379–85. doi: 10.1007/s13277-015-4052-4
25. Wu J, Liu L, Jin H, Li Q, Wang S, Peng B. LncSNHG3/miR-139-5p/BMI1 axis regulates proliferation, migration, and invasion in hepatocellular carcinoma. *Onco Targets Ther.* (2019) 12:6623–38. doi: 10.2147/OTT.S196630
26. Zhao Q, Wu C, Wang J, Li X, Fan Y, Gao S, et al. LncRNA SNHG3 promotes hepatocellular tumorigenesis by targeting miR-326. *Tohoku J Exp Med.* (2019) 249:43–56. doi: 10.1620/tjem.249.43
27. Wang Y, Li W, Chen X, Li Y, Wen P, Xu F. MIR210HG predicts poor prognosis and functions as an oncogenic lncRNA in hepatocellular carcinoma. *Biomed Pharmacother.* (2019) 111:1297–301. doi: 10.1016/j.biopha.2018.12.134
28. Li T, Xie J, Shen C, Cheng D, Shi Y, Wu Z, et al. Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma. *Oncogene.* (2016) 35:1575–84. doi: 10.1038/onc.2015.223
29. Ma Z-j, Wang Y, Li H-f, Liu M-H, Bi F-r, Ma L, et al. LncZEB1-AS1 regulates hepatocellular carcinoma bone metastasis via regulation of the miR-302b-EGFR-PI3K-AKT axis. *J Cancer.* (2020) 11:5118–28. doi: 10.7150/jca.45995
30. Mu B, Lv C, Liu Q, Yang H. Long non-coding RNA ZEB1-AS1 promotes proliferation and metastasis of hepatocellular carcinoma cells by targeting miR-299-3p/E2F1 axis. *J Biochem.* (2021) 170:41–50. doi: 10.1093/jb/mvab042
31. Xue S, Lu F, Sun C, Zhao J, Zhen H, Li X. LncRNA ZEB1-AS1 regulates hepatocellular carcinoma progression by targeting miR-23c. *World J Surg Oncol.* (2021) 19:121. doi: 10.1186/s12957-021-02176-8
32. Oliveri RS, Wetterslev J, Gluud C. Hepatocellular carcinoma. *Lancet.* (2012) 380:470. doi: 10.1016/S0140-6736(12)61285-9
33. Deng Y, Pang Q, Miao RC, Chen W, Zhou YY Bi JB, et al. Prognostic significance of pretreatment albumin/globulin ratio in patients with hepatocellular carcinoma. *Onco Targets Ther.* (2016) 9:5317–28. doi: 10.2147/OTT.S109736
34. Kimmelman AC, White E. Autophagy and tumor metabolism. *Cell Metab.* (2017) 25:1037–43. doi: 10.1016/j.cmet.2017.04.004
35. Lucarelli G, Loizzo D, Franzin R, Battaglia S, Ferro M, Cantiello F, et al. Metabolomic insights into pathophysiological mechanisms and biomarker discovery in clear cell renal cell carcinoma. *Expert Rev Mol Diagn.* (2019) 19:397–407. doi: 10.1080/14737159.2019.1607729
36. Yun CW, Lee SH. The roles of autophagy in cancer. *Int J Mol Sci.* (2018) 19:3466. doi: 10.3390/ijms19113466
37. Zhou Y, Ge Y, Liu Q, Li YX, Chao X, Guan JJ, et al. LncRNA BACE1-AS promotes autophagy-mediated neuronal damage through the miR-214-3p/ATG5 signalling axis in Alzheimer's disease. *Neuroscience.* (2020) 455:52–64. doi: 10.1016/j.neuroscience.2020.10.028
38. Du Y, Wei N, Ma R, Jiang SH, Song D. Long noncoding RNA MIR210HG promotes the warburg effect and tumor growth by enhancing HIF-1alpha translation in triple-negative breast cancer. *Front Oncol.* (2020) 10:580176. doi: 10.3389/fonc.2020.580176
39. Li XY, Zhou LY, Luo H, Zhu Q, Zuo L, Liu GY, et al. The long noncoding RNA MIR210HG promotes tumor metastasis by acting as a ceRNA of miR-1226-3p to regulate mucin-1c expression in invasive breast cancer. *Aging (Albany NY).* (2019) 11:5646–65. doi: 10.18632/aging.102149
40. Wang AH, Jin CH, Cui GY Li HY, Wang Y, Yu JJ, et al. MIR210HG promotes cell proliferation and invasion by regulating miR-503-5p/TRAF4 axis in cervical cancer. *Aging (Albany NY).* (2020) 12:3205–17. doi: 10.18632/aging.102799
41. Ma T, Chen H, Wang P, Yang N, Bao J. Downregulation of lncRNA ZEB1-AS1 represses cell proliferation, migration, and invasion through mediating PI3K/AKT/mTOR signaling by miR-342-3p/CUL4B axis in prostate cancer. *Cancer Biother Radiopharm.* (2020) 35:661–72. doi: 10.1089/cbr.2019.3123
42. Ni X, Ding Y, Yuan H, Shao J, Yan Y, Guo R, et al. Long non-coding RNA ZEB1-AS1 promotes colon adenocarcinoma malignant progression via miR-455-3p/PAK2 axis. *Cell Prolif.* (2020) 53:e12723. doi: 10.1111/cpr.12723

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 Medicine

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fmed.2021.781937

EEG-Driven Prediction Model of Oxcarbazepine Treatment Outcomes in Patients With Newly-Diagnosed Focal Epilepsy

Bin Wang[1,2], Xiong Han[1,2], Zongya Zhao[3], Na Wang[1], Pan Zhao[1], Mingmin Li[1], Yue Zhang[2], Ting Zhao[1], Yanan Chen[1], Zhe Ren[2] and Yang Hong[4]*

[1] Department of Neurology, Zhengzhou University People's Hospital, Zhengzhou, China, [2] Department of Neurology, Henan Provincial People's Hospital, Zhengzhou, China, [3] School of Medical Engineering, Xinxiang Medical University, Xinxiang, China, [4] Department of Neurology, Henan University People's Hospital, Zhengzhou, China

Objective: Antiseizure medicine (ASM) is the first choice for patients with epilepsy. The choice of ASM is determined by the type of epilepsy or epileptic syndrome, which may not be suitable for certain patients. This initial choice of a particular drug affects the long-term prognosis of patients, so it is critical to select the appropriate ASMs based on the individual characteristics of a patient at the early stage of the disease. The purpose of this study is to develop a personalized prediction model to predict the probability of achieving seizure control in patients with focal epilepsy, which will help in providing a more precise initial medication to patients.

Methods: Based on response to oxcarbazepine (OXC), enrolled patients were divided into two groups: seizure-free (52 patients), not seizure-free (NSF) (22 patients). We created models to predict patients' response to OXC monotherapy by combining Electroencephalogram (EEG) complexities and 15 clinical features. The prediction models were gradient boosting decision tree-Kolmogorov complexity (GBDT-KC) and gradient boosting decision tree-Lempel-Ziv complexity (GBDT-LZC). We also constructed two additional prediction models, support vector machine-Kolmogorov complexity (SVM-KC) and SVM-LZC, and these two models were compared with the GBDT models. The performance of the models was evaluated by calculating the accuracy, precision, recall, F1-score, sensitivity, specificity, and area under the curve (AUC) of these models.

Results: The mean accuracy, precision, recall, F1-score, sensitivity, specificity, AUC of GBDT-LZC model after five-fold cross-validation were 81%, 84%, 91%, 87%, 91%, 64%, 81%, respectively. The average accuracy, precision, recall, F1-score, sensitivity, specificity, AUC of GBDT-KC model with five-fold cross-validation were 82%, 84%, 92%, 88%, 83%, 92%, 83%, respectively. We used the rank of absolute weights to separately calculate the features that have the most significant impact on the classification of the two models.

Conclusion: (1) The GBDT-KC model has the potential to be used in the clinic to predict seizure-free with OXC monotherapy. (2). Electroencephalogram complexity, especially

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute (NCI),
United States

Reviewed by:
Wesley T. Kerr,
University of California, Los Angeles,
United States
James Tao,
University of Chicago, United States

***Correspondence:**
Xiong Han
hanxiong@zzu.edu.cn

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *23 September 2021*
Accepted: *06 December 2021*
Published: *03 January 2022*

Citation:
Wang B, Han X, Zhao Z, Wang N, Zhao P, Li M, Zhang Y, Zhao T, Chen Y, Ren Z and Hong Y (2022) EEG-Driven Prediction Model of Oxcarbazepine Treatment Outcomes in Patients With Newly-Diagnosed Focal Epilepsy. Front. Med. 8:781937. doi: 10.3389/fmed.2021.781937

Kolmogorov complexity (KC) may be a potential biomarker in predicting the treatment efficacy of OXC in newly diagnosed patients with focal epilepsy.

Keywords: precision medicine, machine learning, prediction model, gradient boosting decision tree (GBDT) model, EEG complexity

INTRODUCTION

Epilepsy is a chronic disease affecting more than 70 million people worldwide, it is characterized by recurrent, paroxysmal, rigid, and unpredictable alterations of sensory and motor systems, and abnormal electrical activity of neurons (1, 2). Epilepsy is classified into four types: focal, generalized, combined generalized and focal, and unknown onset (3). Focal epilepsy, accounting for 60% of all epilepsies, is the most frequent type of epilepsy and occurs in patients of all ages (4). Comparative monotherapy trials in patients with newly diagnosed focal epilepsy have shown that oxcarbazepine (OXC) is equal in efficacy to phenytoin and immediate-release carbamazepine but may have superior tolerability (5–7). Pharmacotherapy is the primary treatment modality for epilepsy, however, in some patients, seizures cannot be controlled with antiseizure medicine (ASM) and lead to significant risks of neuronal damage and cognitive decline (8). This highlights a need for the prediction of drug response at the drug initiation phase.

However, ASM response is complex and is modulated by multiple factors, including environmental, anthropometric, and genetic factors, and biological subsystems affected by the disease (9). The current standard of care relies on trial and error with sequential therapy. Although there are drug selection guidelines based on seizure types (focal or generalized onset), many drugs have similar efficacy (10). So, drug selection becomes extremely difficult as it is impossible to predict which drugs will be the most effective in a particular patient. There are also no biomarkers that can reliably predict treatment response during conventional treatment.

Since the 1980s, precision medicine has emerged as a new paradigm for improving and promoting patient-specific medicine. Its key goal is to provide personalized treatment for every patient where medical decisions are based on the individual characteristics of the patient, rather than the average characteristics of the entire patient population. Precision medicine requires the analysis of different types of multivariate data from the same individual. It has been used in the early diagnosis and prevention of diseases, reduction of the risk of side effects and adverse events of medications, and in the design of clinical trials (11, 12). The development of precision medicine is inseparable from artificial intelligence. Machine learning, as a branch of artificial intelligence, has the ability to build integrated and multi-scale models by integrating different types of features at different levels. Recently, De Jong et al. (13) integrated pharmacogenetics and clinical data to achieve an accurate prediction of brivaracetam treatment response, but this method is not cost-effective to be integrated in the clinical practice (14).

Studies have shown that the Electroencephalogram (EEG) signal is an internal "fingerprint" of individuals (15). Although with the increase of age, EEG frequency, amplitude, and other aspects will change to a certain extent, the oscillation network of brain waves in each adult brain is relatively stable, and many genetic, structural, and functional abnormalities related to diseases are more or less directly involved in the generation and/or synchronization of brain wave oscillations. Electroencephalogram can be used as a biomarker for the treatment of brain diseases such as epilepsy (16). Other studies and our previous work have shown that it is possible to predict ASM response using EEG-based artificial intelligence (17–20). So, here we test whether the integration of EEG and clinical data can be used to construct a prediction model of OXC treatment outcomes, that can facilitate the correct selection of ASM in newly-diagnosed patients with focal epilepsy patients.

Abbreviations: ASM, antiseizure medicine; CI, confidence interval; GBDT, gradient boosting decision tree; GBM, gradient boosting machine; KC, Kolmogorov complexity; LZC, Lempel-Ziv complexity; NSF, not seizure-free; OXC, oxcarbazepine; RFE, recursive feature elimination; SF, seizure-free; SVM, support vector machine; TLE, temporal lobe epilepsy; WGS, whole genome sequencing.

MATERIALS AND METHODS

Participants and Data Acquisition

Participants

The retrospective study was approved by the Henan Provincial People's Hospital ethics committees, and informed consent was obtained from all participants. Six thousand three hundred seventy patients with epilepsy were registered between January 2014 and April 2021 at the Epilepsy Center of Henan Provincial People's Hospital. Focal epilepsy is defined as seizures originating within networks limited to one hemisphere and the seizures may be discretely localized or widely distributed (21). Patients who meet the following criteria were included: newly diagnosed focal epilepsy with drug-naïve; OXC is the only ASM after diagnosis; long-term scalp EEG recordings were conducted before drug initiation; more than 1 year of follow-up. Exclusion criteria were the following: generalized epilepsy and epileptic syndrome; other ASM were taken before OXC; the combination of other ASMs; lack of EEG data; follow-up data of <1 year and poor adherence; pregnant or lactating women.

Seventy-four individuals with newly diagnosed patients with focal epilepsy patients, initially treated with OXC, were enrolled at our center. After 1 year of follow-up, according to Engel class (22), SF was defined as patients with epilepsy who met Class I while not seizure-free (NSF) was defined as patients who met Engel Class II, III, and IV. Finally, 52 patients were enrolled in the SF group and 22 in the NSF group (**Figure 1**).

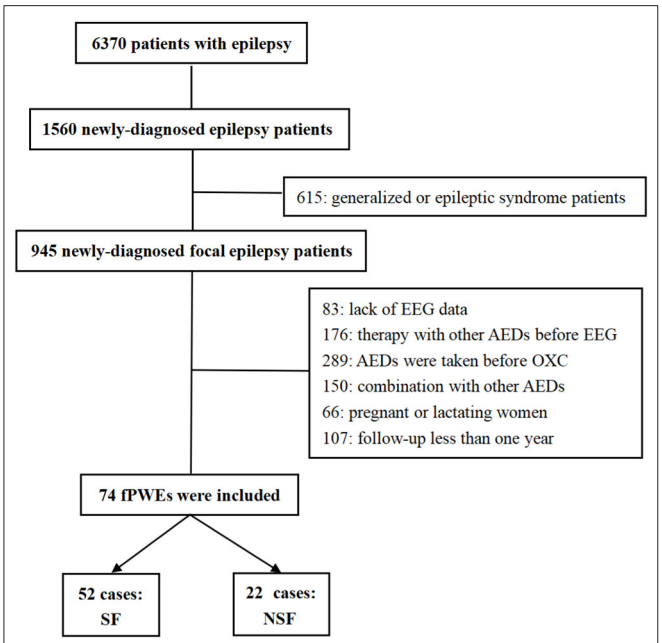

FIGURE 1 | Flow chart. EEG, electroencephalogram; AEDs, antiepileptic drugs; OXC, oxcarbazepine; fPWEs, patients with focal epilepsy; SF, seizure-free; NSF, not seizure-free.

Clinical Data

We included 15 clinical features: sex, age, age at the onset of the disease, follow-up time, seizure frequency before OXC, seizure circadian rhythm, comorbidities, inducement, history of perinatal injury, physical development, family history of epilepsy, MRI, temporal lobe epilepsy (TLE), history of central nervous system infection, and history of head injuries (23–26) (**Table 1**).

EEG Data

EEG Acquisition

Long-term scalp EEG was carried out by EEG-1200C electroencephalograph (Nihon Kohden, Tokyo, Japan), the sampling rate was 256 Hz, the amplifier was 1,000x. Electrodes were placed according to the international 10–20 system. There were 19 scalp electrodes and 2 reference electrodes. The discharges of EEG were marked independently by two experienced electroencephalographers. If there were any disputes, another clinical neurologist was consulted. We intercepted a continuous 1-h EEG including the waking period and the sleeping period. The waking period and the sleeping period accounted for 30 min each.

EEG Preprocessing

Matlab software (Mathworks Inc., USA) equipped with the EEGLAB toolbox was used for EEG preprocessing (27). Electroencephalogram preprocess was as follows: firstly, 0.5–30 Hz EEG fragments were retained using bandpass filter. Then independent component analysis was used to remove artifacts. Next, EEG data without epileptic charges or artifacts, were taken while the patient was awake, and their eyes were open. The EEG data were divided into 15 s periods and recalculated based on a reference average. Finally, 15 time periods for each subject were randomly selected for subsequent analysis.

EEG Complexity Estimators

Lempel–Ziv Complexity

Lempel–Ziv complexity (LZC) is a simple non-parametric measure to calculate the randomness of a one-dimensional finite-length sequence. It was related to the number of different substrings and their occurrence rate along the sequence. The larger the value is, the more complex the corresponding data is (28, 29). The following procedure was done with MathWorks in MATLAB. Before calculating EEG complexity, the binarization was performed according to the median value of the EEG time series. For binary sequences S (S1, S2,..., Sn), the sequence length is n, and $c(n)$ is defined as the LZC value of the EEG time series. When a new subsequence appears in the time series, $c(n)$ increases by one unit, and the pattern search continues until the last string is scanned. For a sufficiently long random 0–1 sequence, the following formula holds if 0 and 1 are equally likely to occur:

$$\lim_{n\to\infty} c(n) = b(n) = n/\log 2^n \tag{1}$$

The $b(n)$ is used to normalize $c(n)$ to obtain a value independent of the sequence length n, so LZC is:

$$LZC = c(n)/b(n) \tag{2}$$

Kolmogorov Complexity

Kolmogorov complexity (KC), known as algorithmic complexity, is defined as a new algorithmic measure of randomness for generating quantitative definitions of information (30). Kolmogorov complexity describes the randomness of an object, which is a string based on the length of a computer program; the complexity of a string, consisting of 0 and 1, is estimated by the number of bits of the shortest computer program that produces the string. The KC is described as follows:

$$k_u(x) = \min p : u(p) = x^{l(p)} \tag{3}$$

Where p is the computer program and $l(p)$ is the length of x output strings of u general Turing machine (computer). Kolmogorov complexity is the minimum length of the output of a computer program. To calculate the KC of an EEG, the data were first converted into discrete binary sequences. Then, KC estimation methods could be used to analyze the bits of the shortest computer program associated with the discrete sequence. Based on previous reports, the KC estimation was carried out by the difference method. When the difference between two sequential samples was positive, the method assigned 1, and when the difference was negative, it assigned 0.

TABLE 1 | Demographical and clinical status of the participants.

	SF ($n = 52$)	NSF ($n = 22$)	$x^2/Z/t$-value	P-value
Sex (Male/female)	33/19	11/11	1.162[a]	0.281
Age, year	14.5 ± 11.50	16.50 ± 12.00	−0.681[b]	0.496
Age at onset, year	13.5 ± 12.25	15.50 ± 13.38	−0.361[b]	0.718
Follow-up time, months	32.58 ± 9.83	36.18 ± 8.91	−1.481[c]	0.143
Seizure frequency before OXC, times/month	0.65 ± 0.70	15.50 ± 13.37	−1.983[b]	**0.045***
Seizure circadian rhythm (day/night/both)	19/16/17	8/7/7	0.009[a]	0.995
Comorbidity (Y/N)	27/25	16/6	2.749[a]	0.097
Inducement (Y/N)	29/23	17/5	3.039[a]	0.081
History of perinatal injury (Y/N)	12/40	12/10	6.986[a]	**0.008***
Physical development (N/AN)	9/43	3/19	4.469[a]	**0.035***
Family history (Y/N)	2/50	2/20	0.832[a]	0.362
MRI(P/N)	17/35	13/9	4.469[a]	**0.035***
TLE(Y/N)	16/36	8/14	0.221[a]	0.638
History of CNS infection (Y/N)	7/45	2/20	0.261[a]	0.599
History of head injury (Y/N)	5/47	2/20	0.005[a]	0.944

SF, seizure-free; NSF, not seizure-free; OXC, oxcarbazepine; Y/N, yes/no; N/AN, normal/abnormal; P/N, positive/negative; MRI, magnetic resonance imaging; TLE, temporal lobe epilepsy; CNS, central nervous system.

[a]For qualitative data, Chi-square tests were used.

[b]For quantitative data, after Shapiro-Wilk normality test, the Mann-Whitney U-test was applied for data with abnormal distributions, data that did not conform to normal distributions were presented as the median ± interquartile range.

[c]Data with a normal distribution were compared by the independent sample t-tests, mean ± standard deviation was used to describe. $p < 0.05$ is considered as statistically significant.

**Defined as features that have statistically significant between SF group and NSF group.*

Bold values are statistically significant.

Model Process

For imbalance in a sample, SMOTE was used to strike an equilibrium during the training process (31). Toolkits: Python's sklearn toolkits (32). To avoid overfitting, default parameters were used unless otherwise specified. The parameters were set as follows: The number of nearest neighbors is 5 ($K = 5$); Degree of over-sampling: making the number of positive and negative samples consistent. SMOTE in this paper was done independently within each training set, not used before cross-validation. The SMOTE consists of two functions, SMOTE (*T, N, K*) and Populate (*N, i, nnarray*). The SMOTE code idea is very simple: scan every sample point, calculate *K* nearest neighbor of every sample point, record the index of each nearest neighbor point in *nnarray*, then pass it into Populate (*N, i, nnarray*), and complete a sample point. Populate is responsible for randomly generating *N* samples similar to the observed sample *i* based on the index in the *nnarray*. The function calculates the gap dif between random neighboring point *nn* and each feature of observed sample point *i*, multiplying the gap by a [0,1] random factor gap, and then combining the value of dif * gap plus the observation point *i* (33, 34).

SVM Model

The support vector machine (SVM) is a classical classifier with good performance in dichotomies (35). It has good performance for small samples (36). Lib-SVM was used for the classification process. The core of SVM is to establish an optimized hyperplane. A linear SVM classifier was constructed based on kernel parameter and regularization C parameter. In this study, the C parameter was set to 1.

The five-fold cross-validation was used for the classification process, four-fold pats as the training sets, and one-fold part as the validation sets. This process was repeated five times until all subjects went through it once. The recursive feature elimination (RFE) was used for feature selection. In cross-validation, the RFE occurs in the training sets but not the validation sets, with the results of the RFE feature screening from the training sets to guide the feature selection in the validation sets. Absolute weight value was applied to the feature selection procedure; the greater the absolute weight of features, the greater the influence on classification. We used EEG complexity features such as KC and LZC combined with clinical features to establish two SVM-RFE models for predicting SF with OXC monotherapy in patients with newly diagnosed focal epilepsy.

GBDT Model

Based on the theory of Gradient boosting machine (GBM), Gradient Boosting Decision Tree (GBDT) is a typical representative of ensemble learning, which is a lifting algorithm (37). Gradient Boosting Decision Tree can effectively avoid overfitting by combining decision trees with gradient algorithms. It is considered that all machine learning algorithms can be used as the basic learning machine of gradient lifting by GBM. Because decision trees are easier to understand and calculate compared with other algorithms, GBDT chooses decision tree as the base learning machine. Decision tree can combine multiple features, and has good processing ability for non-parameterized features. Therefore, when there are outliers or non-linearly separable data in the data, decision tree can be used to process these data. However, the decision tree suffers from the drawback of overfitting. So, combining the decision tree (formed by the

combination of multiple gradient lifting methods) with the gradient lifting algorithm can reduce the overfitting of the decision tree (38, 39).

Two predictive models were established by GBDT using KC-clinical, LZC-clinical characteristics. Five-fold cross-validation was also used for this process. In this process, absolute weights were used to rank the features that influence classification, as well.

Descriptive Statistics

Statistical analysis was calculated with SPSS. The Shapiro-Wilk normality test was used to assess the normality distribution of data. For quantitative data, the independent sample *t*-tests were used to compare the data with a normal distribution (Mean ± standard deviation); the Mann-Whitney U-test was applied for data with abnormal distributions (median ± interquartile range). A strict false discovery rate based on the Benjamini–Hochberg correction was applied to *p*-values to correct for multiple comparisons. While, for qualitative data, Chi-square tests were used. $P < 0.05$ was considered statistically significant in this study.

RESULTS

The study included 44 males and 30 females. The age range was 5–70, the overall follow-up time ranged from 12 to 60 months. The average follow-up time was 21.97 months in 74 individuals—24.17 months in the SF group and 25.13 in the NSF group. There were no significant differences in gender, age, and follow-up time between the two groups. However, significant differences were found in seizure frequency before OXC ($p = 0.045$), history of perinatal injury ($p = 0.008$), physical development (p=0.035), and MRI ($p = 0.035$) (**Table 1**).

GBDT-LZC

The NSF group showed higher LZC than the SF group. The top 10 features that influenced classification were δ band from F8 channel, θ band from T3 channel ($p < 0.05$), α band form Cz channel ($p < 0.05$), θ band from F3 channel ($p < 0.05$), α band form Fz channel ($p < 0.05$), θ band from T6 channel, TLE, β band from T3 and Pz channel ($p < 0.05$), α band form T6 channel ($p < 0.05$) (**Table 2A**; **Figure 2**). The mean accuracy, precision, recall, F1-score, sensitivity, specificity, and AUC of the GBDT model after five-fold cross-validation were 81%, 84%, 91%, 87%, 91%, 64%, 81%, respectively (**Table 3A**; **Figure 3**). The mean accuracy, precision, recall, F1-score, sensitivity, specificity, and AUC of the SVM-RFE model after five-fold cross-validation were 62%, 77%, 91%, 87%, 91%, 64%, 81%, respectively (**Table 3B**; **Figure 3**).

GBDT-KC

Like LZC, it was apparent that the NSF group has higher KC than the SF group. The top ten features that have the highest

TABLE 2 | The top 10 features that impacting the GBDT classifier mostly.

	SF	NSF	*Z*/*t* value	*P*-value	*P*′-value
(A) GBDT-LZC					
δ-F8	0.0240 ± 0.0214	0.0298 ± 0.0132	−1.656	0.098	0.098
θ-T3	0.0518 ± 0.0356	0.0566 ± 0.0558	−2.010	0.044	0.060
α-Cz	0.0631 ± 0.0383	0.0745 ± 0.0500	−2.472	0.013	**0.032***
θ-F3	0.0510 ± 0.0352	0.0561 ± 0.0526	−2.032	0.042	0.057
α-Fz	0.0649 ± 0.0384	0.0757 ± 0.0582	−2.424	0.015	**0.036***
θ-T6	0.0531 ± 0.0394	0.0587 ± 0.0528	−1.880	0.060	0.068
TLE					
β-T3	0.1119 ± 0.0742	0.1261 ± 0.1133	−2.081	0.037	0.050
β-Pz	0.1100 ± 0.7230	0.1227 ± 0.1024	−2.081	0.037	0.050
α-T6	0.0624 ± 0.0334	0.0742 ± 0.0569	−2.389	0.017	**0.038***
(B) GBDT-KC					
δ-T3	0.2245 ± 0.0068	0.2531 ± 0.2333	−1.809	0.070	0.070
θ-F7	0.0503 ± 0.0204	0.0575 ± 0.0555	−1.904	0.057	0.065
Seizure frequency before OXC					
θ-FP1	0.0521 ± 0.038	0.0573 ± 0.0566	−1.928	0.054	0.061
θ-T6	0.0537 ± 0.0398	0.0594 ± 0.0535	−1.904	0.057	0.065
θ-Fz	0.0505 ± 0.0371	0.0564 ± 0.0540	−2.105	0.035	**0.044***
β-T6	0.1116 ± 0.0728	0.1250 ± 0.1084	−2.105	0.035	**0.044***
β-O2	0.1109 ± 0.0704	0.1244 ± 0.1148	−2.200	0.028	**0.031***
Seizure circadian rhythm					
α-Pz	0.0645 ± 0.037	0.0753 ± 0.0542	−2.306	0.021	**0.028***

GBDT, gradient boosting decision tree; LZC, Lempel-Ziv complexity; KC, Kolmogorov complexity; SF, seizure-free; NSF, not seizure-free; TLE, temporal lobe epilepsy; δ-F8, δ band from F8 channel; OXC, oxcarbazepine; P′-value refers to P-value that is corrected by false discovery rate correction.

**The features that have statistically significance. Although the selected features may not be statistically significant, they did have a classification value in the model.*

Bold values are statistically significant.

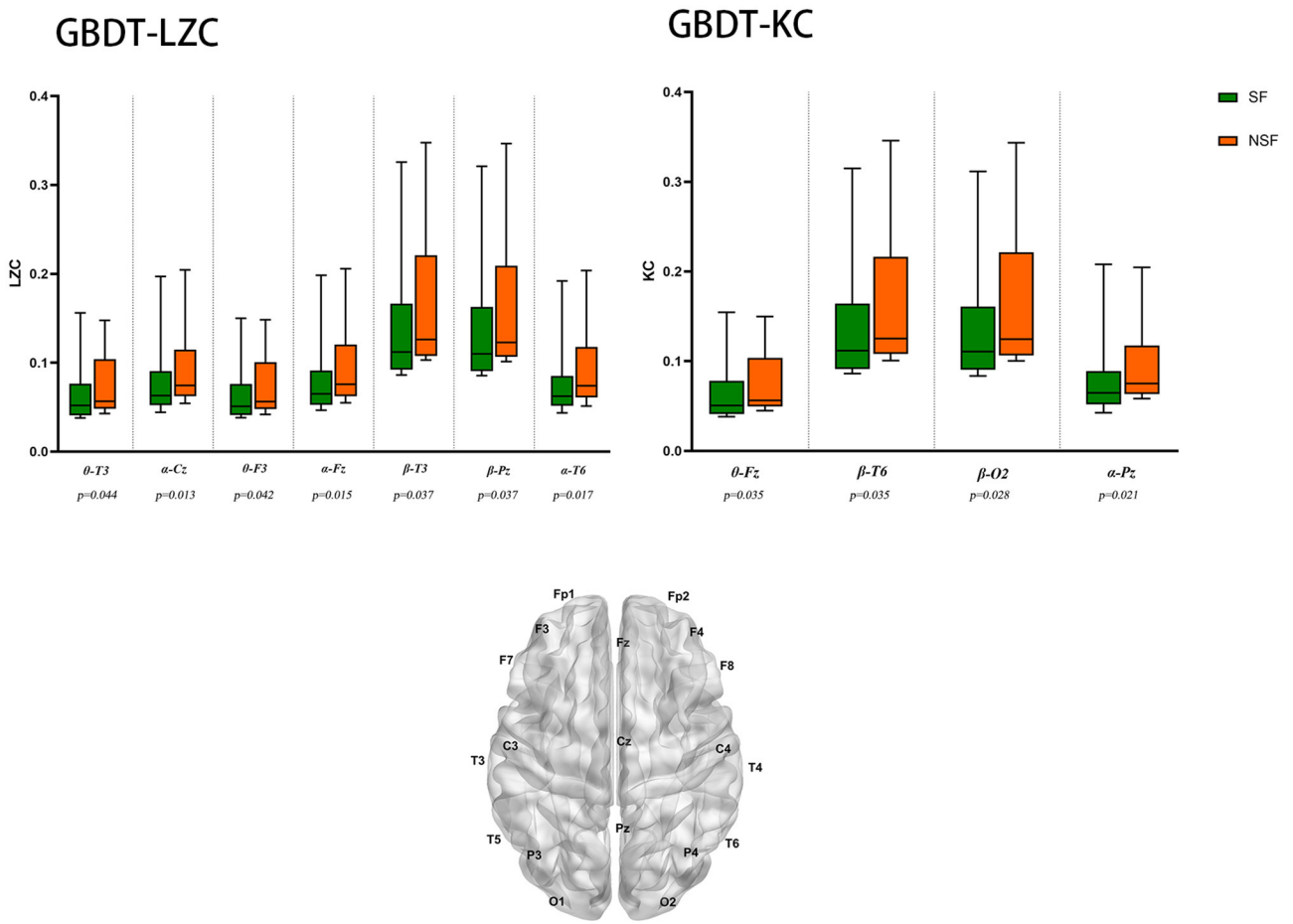

FIGURE 2 | EEG features that have the most significant impact on classification. θ-T3: θ band from T3 channel, $p < 0.05$ is considered statistically significant.

absolute weights were δ band from T3 channel, θ band from F7 channel, seizure frequency before OXC, θ band from FP1 channel, T6 and Fz ($p < 0.05$) channel, β band from T6 and O2 channel ($p < 0.05$), seizure circadian rhythm, and α band form Pz channel ($p < 0.05$) (**Table 2B**; **Figure 2**). Although the selected features may not be statistically significant, they did have a classification value in the model. The model yielded average accuracy of 82%, precision of 84%, recall of 92%, F1-score of 88%, sensitivity of 83%, specificity of 92%, and AUC of 83% after five-fold cross-validation, respectively (**Table 3C**; **Figure 3**). Compared with the GBDT model, SVM-RFE model yielded mean accuracy, precision, recall, F1-score, sensitivity, specificity, AUC of five-fold cross-validation were 62%, 77%, 67%, 71%, 67%, 55%, 63%, respectively (**Table 3D**; **Figure 3**). The results of each fold were presented in **Figure 4**.

DISCUSSION

We constructed a model for predicting OXC treatment outcomes. Our GBDT-KC model (EEG complexity and clinical data) performed better in terms of the performance merits compared with De Jong's study (pharmacogenetics and clinical data) (13). Our research has a more clinical application because it is cost-efficient. To our knowledge, this is the first study that applied EEG complexity to predict OXC response in patients with focal epilepsy, and achieved good performance.

EEG Complexity as a Biomarker for Epilepsy

Electroencephalogram plays an important role in the diagnosis, treatment, and prognosis of epilepsy (40–42). Electroencephalogram signals have non-linear structures in the time dimension. Recently, new methods for studying EEG signals have been developed from non-linear systems theory since non-linear measurements are more suitable to reflect the complex, irregular, and non-stationary behavior of neural processes. The non-linear analysis quantifies the complexity of EEG and reflects the state of brain neural networks. Electroencephalogram complexity correlates with synchronization (43); highly synchronized signals (e.g., epileptic seizures) give rise to low complexity values (44). Complexity is related to the degree of entropy, so some of these estimators

TABLE 3 | The performance of the four classifier models.

	Fold1	Fold2	Fold3	Fold4	Fold5	Mean-Value
(A) GBDT-LZC						
Accuracy (%)	67	87	80	93	79	81
Precision (%)	69	80	81	100	90	84
Recall (%)	90	100	90	92	82	91
F1-score (%)	78	89	86	96	86	87
AUC (%)	64	89	82	100	70	81
Sensitivity (%)	90	100	90	92	82	91
Specificity (%)	20	71	60	100	67	64
(B) SVM-LZC						
Accuracy (%)	60	53	67	67	64	62
Precision (%)	70	56	78	100	80	77
Recall (%)	70	63	70	62	73	67
F1-score (%)	70	59	74	76	76	71
AUC (%)	54	64	64	100	33	63
Sensitivity (%)	70	63	70	62	73	67
Specificity (%)	40	43	60	100	33	55
(C) GBDT-KC						
Accuracy (%)	67	87	80	100	79	82
Precision (%)	69	80	82	100	90	84
Recall (%)	90	100	90	100	82	92
F1-score (%)	78	89	86	100	86	88
AUC (%)	66	89	88	100	73	83
Sensitivity (%)	66	89	88	100	73	83
Specificity (%)	90	100	90	100	82	92
(D) SVM-KC						
Accuracy (%)	60	53	67	67	64	62
Precision (%)	70	56	78	100	80	77
Recall (%)	70	63	70	62	73	67
F1-score (%)	70	59	74	76	76	71
AUC (%)	54	63	64	100	33	63
Sensitivity (%)	70	63	70	62	73	67
Specificity (%)	40	43	60	100	33	55

LZC, Lempel-Ziv complexity; GBDT, gradient boosting decision tree; AUC, the area under the curve; SVM, support vector machine; RFE, recursive feature elimination; KC, Kolmogorov complexity.

are called entropy estimators. Based on the complexity of the algorithm, LZC does not rely on large amounts of EEG data and is suitable for short and non-stationary time series (45). Kolmogorov complexity is defined as the complexity of a sequence and is based on the length of the shortest program that could generate the sequence (46). Kolmogorov complexity was found to be more sensitive to the detection in patients with schizophrenia compared with other measures (47). However, the application of EEG complexity in epilepsy remains limited.

In our study, the LZC and KC showed a complexity decrease in the SF group compared with the NSF group. Though the relationship between EEG complexity and epilepsy is not clear, EEG complexity is related to the severity and prognosis of the disease. Cerquera et al. (48) analyzed the difference between cognitive deficit schizophrenia (DS) and non-cognitive deficit schizophrenia (NDS), and found that the DS group showed less LZC in the frontal lobe than the NSD group. Another study found a significant reduction in EEG complexity, 2 min before the seizure, compared with the inter-seizure period (about 6–8 min before the seizure) (49). Valproic acid treatment also decreased the overall complexity of 19 EEG channels in patients with idiopathic epilepsy (50).

The Prediction Models of Drug Response

Antiseizure medicines are still the mainstream treatment for patients with epilepsy. Non-standard treatment in the early stage has been shown to contribute to poor prognosis (51). Therefore, it is necessary to choose appropriate ASMs for epileptic patients. Although there are many ASMs for focal epilepsy, the differences between these ASMs are unclear and many ASMs are cross-referenced for both focal and generalized epilepsy. Further, there are different adverse effects and different treatment responses of these ASMs. Therefore, ASMs suitable for patient A, may not be the right choice for other patients (B, C,…). Previous studies of drug response were based mostly on clinical characteristics, without an in-depth analysis of individual patients (52, 53).

Recently, precision medicine has made significant development where the goal is to make personalized medical decisions based on the individual characteristics of the patient. Precision medicine is closely related to pharmacogenetics, but mostly in oncology, and has considerable impact on drug prescription (54). Outside oncology, genetic information has not yet played a major role in drug selection. However, the field is an active area of research. Precision medicine is usually associated with gene manipulation or gene targeting. De Jong et al. (13) designed a phase III clinical trial in which 235 participants were randomly assigned to brivaracetam or placebo groups. Only the genomes of brivaracetam treated patients were sequenced. Clinical characteristics and whole genome sequencing (WGS) pharmacogenetics data were combined to predict brivaracetam drug response in patients with focal epilepsy. The GBDT classifier was confirmed as the best performing model with an AUC of 0.76 in the discovery datasets and 0.75 in the validation datasets; the asymptotic 95% confidence interval (CI) was wide (0.6–0.9). However, the study had several limitations which are as follows: the dimensions of WGS data are huge, concerns over overfitting, WGS data were generated only for patients with brivaracetam therapy, the 95% CI in validation datasets as wide and it was too expensive to be clinically applicable. Our GBDT model, based on clinical and EEG complexity features, achieved better prediction performance than De Jong's study, with an average AUC of 0.832. At present, using EEG to predict drug response has high clinical value, the price is more affordable, and the prediction performance is not bad. Although, pharmacogenetics is not cost-effective currently, if the cost of genetic sequencing decreases and/or the demonstrated benefit of genetically-guided ASM selection is increased, then it may be cost-effective. It cannot be denied that EEG combined with pharmacogenetics would be more clinically beneficial.

Lin et al. (18) used 24 univariate EEG features extracted from EEG fragments from 11 drug-refractory epilepsy patients

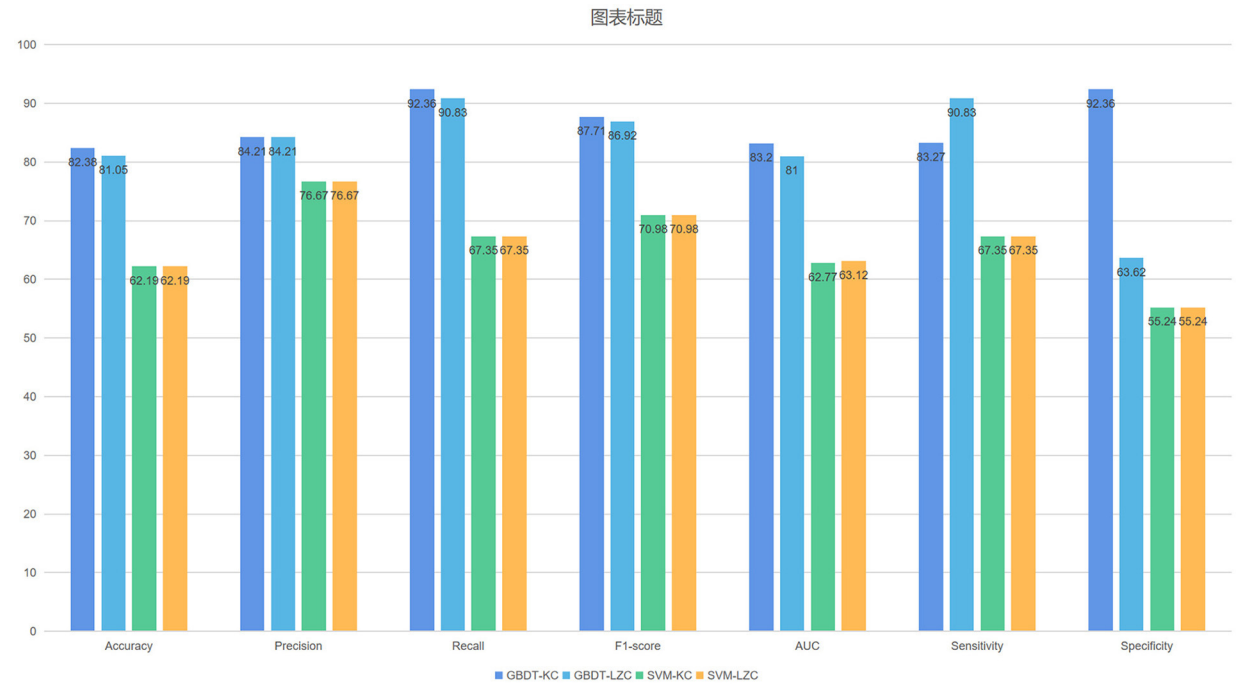

FIGURE 3 | The mean evaluation indexes after five-fold cross-validation.

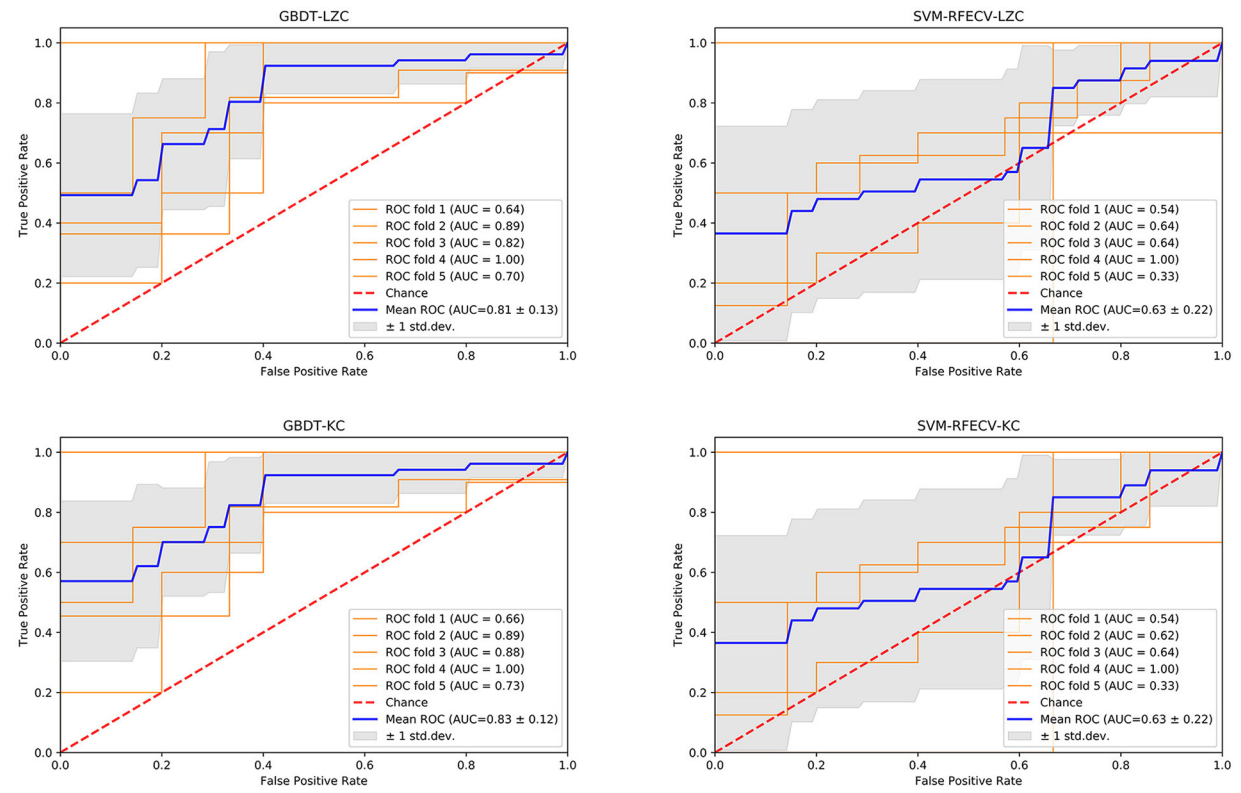

FIGURE 4 | The performance of four models. GBDT, gradient boosting decision tree; LZC, Lempel-Ziv complexity; AUC, area under the curve; ROC, receiver operating-characteristic curve; std. dev, standard deviation; SVM, support vector machine, RFE, recursive feature elimination, CV, cross-validation; KC, Kolmogorov complexity.

and 16 control epilepsy patients to predict drug-resistant epilepsy; the study yielded a precision rate of 0.942, ROC area 0.938. We have devised an integrated model combining clinical features and EEG functional connectivity. Our study used the phase lag index functional connectivity to predict drug-refractory epilepsy for newly diagnosed epilepsy patients and achieved good performance with AUC of 0.98, an accuracy of 0.94, sensitivity of 0.95, and specificity of 0.93(17). We show that EEG has good prediction performance in drug response.

Zhang et al. (20) used clinical and EEG sample entropy features to predict drug response with levetiracetam therapy via SVM achieved good performance. Our SVM-RFE model was inferior to Zhang's, while, the GBDT model achieved better performance than theirs'. In this study, we established a GBDT-KC model to predict SF for patients with focal epilepsy with OXC monotherapy. Our study yielded an average accuracy of 82%, a precision of 84%, recall of 92%, F1-score of 88%, sensitivity of 83%, specificity of 92%, and AUC of 83% after five-fold cross-validation, respectively. Focal-onset epilepsy accounts for the majority of all epilepsy cases. The selection of ASMs for focal epilepsy is of great clinical significance. However, there is still no referenced study for personalized drug selection for focal epilepsy. Our study may facilitate future studies in this field.

Limitations and Prospects

There are limitations to our study. The study was retrospective, and selection bias is inevitable. Prospective studies need to be conducted in the future. The sample size in our study was small, and the model was only suitable for Asians. Multi-center studies with large sample sizes and diverse populations are required. Currently, the clinical problem is the choice between multiple potential ASMs. For example, there are many ASMs for focal epilepsy, such as carbamazepine, OXC, lamotrigine, valproate, clobazam, topiramate, phenytoin, phenobarbital, and zonisamide. Although this study was focused on the drug response of OXC, our final purpose was to make personalized and optimal treatment with less adverse effects for newly diagnosed epilepsy patients. We demonstrated that it is feasible to predict ASMs' response in combination with clinical and EEG complexity features, EEG complexity could be used as a biomarker to predict drug response, the concept is still in its theoretical stage. Our study proposed the possibility of this research in this area, there is still a long way to go in the future.

CONCLUSION

We established a GBDT-KC prediction model for seizure outcome of patients with focal epilepsy with OXC monotherapy. EEG complexity, especially KC can be used as a biomarker for predicting outcomes of ASMs treatment.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

XH and NW obtained funding. BW designed the study, acquired the data, analyzed EEG recordings, worked on EEG preprocessing and machine learning process, drafted, and revised the manuscript. XH designed the study and revised the manuscript. ZZ aided assistant in machine learning process. PZ and ML analyzed EEG recordings. NW, TZ, and YC analyzed and interpreted the data. YZ, ZR, and YH conducted the statistical analysis. All authors revised this draft, read, and approved the final manuscript.

FUNDING

This study was sponsored by Henan Province's Gong Jian Program (Authorization number: SB201901074), 23456 Talent Engineering (Authorization number: ZC20200371), and the National Natural Youth Fund (Authorization number: 81801291).

ACKNOWLEDGMENTS

We would like to thank the staff members at the Department of Neurology at the Henan Provincial People's Hospital for their assistance with this research study and we would like to express our most profound gratitude to our patients and their families.

REFERENCES

1. Jacobs MP, Fischbach GD, Davis MR, Dichter MA, Dingledine R, Lowenstein DH, et al. Future directions for epilepsy research. *Neurology.* (2001) 57:1536–42. doi: 10.1212/wnl.57.9.1536
2. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. *Lancet (London, England).* (2019) 393:689–701. doi: 10.1016/s0140-6736(18)32596-0
3. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia.* (2017) 58:512–21. doi: 10.1111/epi.13709
4. Wiebe S. Epilepsy: a comprehensive textbook on CD-ROM. *BMJ (Clinical research ed).* (2000) 320:810.
5. Abou-Khalil BW. Update on antiepileptic drugs 2019. *Continuum (Minneapolis, Minn).* (2019) 25:508–36. doi: 10.1212/con.0000000000000715
6. Koch MW, Polman SK. Oxcarbazepine versus carbamazepine monotherapy for partial onset seizures. *Cochrane Database Syst Rev.* (2009) 4:Cd006453. doi: 10.1002/14651858.CD006453.pub2
7. Nolan SJ, Muller M, Tudur Smith C, Marson AG. Oxcarbazepine versus phenytoin monotherapy for epilepsy. *Cochrane Database Syst Rev.* (2013) 5:Cd003615. doi: 10.1002/14651858.CD003615.pub3
8. Oostrom KJ, van Teeseling H, Smeets-Schouten A, Peters AC, Jennekens-Schinkel A. Three to four years after diagnosis: cognition and behaviour in children with 'epilepsy only'. A prospective, controlled study. *Brain.* (2005) 128:1546–55. doi: 10.1093/brain/awh494

9. Armstrong M. The genetics of adverse drug reactions: promises and problems. *Methods Pharmacol. Toxicol.* (2008) 2008:121–47. doi: 10.1007/978-1-59745-439-1-7
10. Perucca E, Brodie MJ, Kwan P, Tomson T. 30 years of second-generation antiseizure medications: impact and future perspectives. *Lancet Neurol.* (2020) 19:544–56. doi: 10.1016/s1474-4422(20)30035-1
11. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. *BMC Med.* (2018) 16:150. doi: 10.1186/s12916-018-1122-7
12. Vogenberg FR, Isaacson Barash C, Pursel M. Personalized medicine: part 1: evolution and development into theranostics. *P T.* (2010) 35:560–76.
13. de Jong J, Cutcutache I, Page M, Elmoufti S, Dilley C, Fröhlich H, et al. Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. *Brain.* (2021) 144:1738–50. doi: 10.1093/brain/awab108
14. Chen Z, Anderson A, Ge Z, Kwan P. One step closer towards personalized epilepsy management. *Brain.* (2021) 144:1624–6. doi: 10.1093/brain/awab199
15. Buzsáki G, Logothetis N, Singer W. Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron.* (2013) 80:751–64. doi: 10.1016/j.neuron.2013.10.002
16. Finelli LA, Achermann P, Borbély AA. Individual 'fingerprints' in human sleep EEG topography. *Neuropsychopharmacology.* (2001) 25(5 Suppl):S57–62. doi: 10.1016/s0893-133x(01)00320-7
17. Colic S, Wither RG, Lang M, Zhang L, Eubanks JH, Bardakjian BL. Prediction of antiepileptic drug treatment outcomes using machine learning. *J Neural Eng.* (2017) 14:016002. doi: 10.1088/1741-2560/14/1/016002
18. Lin LC, Ouyang CS, Chiang CT, Yang RC, Wu RC, Wu HC. Early prediction of medication refractoriness in children with idiopathic epilepsy based on scalp EEG analysis. *Int J Neural Syst.* (2014) 24:1450023. doi: 10.1142/s0129065714500233
19. Wang B, Han X, Yang S, Zhao P, Li M, Zhao Z, et al. An integrative prediction algorithm of drug-refractory epilepsy based on combined clinical-EEG functional connectivity features. *J Neurol.* (2021). doi: 10.1007/s00415-021-10718-z. [Epub ahead of print].
20. Zhang JH, Han X, Zhao HW, Zhao D, Wang N, Zhao T, et al. Personalized prediction model for seizure-free epilepsy with levetiracetam therapy: a retrospective data analysis using support vector machine. *Br J Clin Pharmacol.* (2018) 84:2615–24. doi: 10.1111/bcp.13720
21. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia.* (2010) 51:676–85. doi: 10.1111/j.1528-1167.2010.02522.x
22. Engel J Jr, Wiebe S, French J, Sperling M, Williamson P, Spencer D, et al. Practice parameter: temporal lobe and localized neocortical resections for epilepsy: report of the Quality Standards Subcommittee of the American Academy of Neurology, in association with the American Epilepsy Society and the American Association of Neurological Surgeons. *Neurology.* (2003) 60:538–47. doi: 10.1212/01.wnl.0000055086.35806.2d
23. Kalilani L, Sun X, Pelgrims B, Noack-Rink M, Villanueva V. The epidemiology of drug-resistant epilepsy: a systematic review and meta-analysis. *Epilepsia.* (2018) 59:2179–93. doi: 10.1111/epi.14596
24. Sander JW. Comorbidity and premature mortality in epilepsy. *Lancet (London, England).* (2013) 382:1618–9. doi: 10.1016/s0140-6736(13)61136-8
25. van Campen JS, Valentijn FA, Jansen FE, Joëls M, Braun KP. Seizure occurrence and the circadian rhythm of cortisol: a systematic review. *Epilepsy Behav.* (2015) 47:132–7. doi: 10.1016/j.yebeh.2015.04.071
26. Wu T, Chen CC, Chen TC, Tseng YF, Chiang CB, Hung CC, et al. Clinical efficacy and cognitive and neuropsychological effects of levetiracetam in epilepsy: an open-label multicenter study. *Epilepsy Behav.* (2009) 16:468–74. doi: 10.1016/j.yebeh.2009.08.026
27. Delorme A, Makeig S, EEGLAB. an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods.* (2004) 134:9–21. doi: 10.1016/j.jneumeth.2003.10.009
28. Estevez-Rams E, Lora Serrano R, Aragón Fernández B, Brito Reyes I. On the non-randomness of maximum Lempel Ziv complexity sequences of finite size. *Chaos (Woodbury, NY).* (2013) 23:023118. doi: 10.1063/1.4808251
29. Jiménez-Montaño MA, Ebeling W, Pohl T, Rapp PE. Entropy and complexity of finite sequences as fluctuating quantities. *Biosystems.* (2002) 64:23–32. doi: 10.1016/s0303-2647(01)00171-x
30. Liau BY, Yeh SJ, Chiu CC, Tsai YC. Dynamic cerebral autoregulation assessment using chaotic analysis in diabetic autonomic neuropathy. *Med Biol Eng Comput.* (2008) 46:1–9. doi: 10.1007/s11517-007-0243-5
31. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *ArXiv.* (2012) arXiv:1201.0490.
33. Nakamura M, Kajiwara Y, Otsuka A, Kimura H, LVQ-SMOTE. - learning vector quantization based synthetic minority over-sampling technique for biomedical data. *BioData Min.* (2013) 6:16. doi: 10.1186/1756-0381-6-16
34. Japkowicz N. Stephen SJ. The class imbalance problem: a systematic study. *Intell Data Anal.* (2002) 6:429–49. doi: 10.3233/ida-2002-6504
35. Luts J, Ojeda F, Van de Plas R, De Moor B, Van Huffel S, Suykens JA, et al. tutorial on support vector machine-based methods for classification problems in chemometrics. *Anal Chim Acta.* (2010) 665:129–45. doi: 10.1016/j.aca.2010.03.030
36. Way TW, Sahiner B, Hadjiiski LM, Chan HP. Effect of finite sample size on feature selection and classification: a simulation study. *Med Phys.* (2010) 37:907–20. doi: 10.1118/1.3284974
37. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* (2001) 29:1198–232. doi: 10.1214/aos/1013203451
38. Lv ZB. *Escherichia coli* DNA N-4-methycytosine site prediction accuracy improved by light gradient boosting machine feature selection technology. *IEEE Access.* (2020) 8:14851–9. doi: 10.1109/ACCESS.2020.2966576
39. Sahin EK. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Appl Sci.* (2020) 2:1308. doi: 10.1007/s42452-020-3060-1
40. Altenmüller DM, Hebel JM, Deniz C, Volz S, Zentner J, Feuerstein TJ, et al. Electrocorticographic and neurochemical findings after local cortical valproate application in patients with pharmacoresistant focal epilepsy. *Epilepsia.* (2020) 61:e60–5. doi: 10.1111/epi.16523
41. Betting LE, Mory SB, Lopes-Cendes I, Li LM, Guerreiro MM, Guerreiro CA, et al. EEG features in idiopathic generalized epilepsy: clues to diagnosis. *Epilepsia.* (2006) 47:523–8. doi: 10.1111/j.1528-1167.2006.00462.x
42. Dlugos D, Shinnar S, Cnaan A, Hu F, Moshé S, Mizrahi E, et al. Pretreatment EEG in childhood absence epilepsy: associations with attention and treatment outcome. *Neurology.* (2013) 81:150–6. doi: 10.1212/WNL.0b013e31829a3373
43. Escudero J, Ibanez-Molina A, Iglesias-Parro S. Effect of the average delay and mean connectivity of the Kuramoto model on the complexity of the output electroencephalograms. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (Milan) (2015) 7873–6. doi: 10.1109/embc.2015.7320217
44. Radhakrishnan N, Gangadhar BN. Estimating regularity in epileptic seizure time-series data. A complexity-measure approach. *IEEE Eng Med Biol Magaz.* (1998) 17:89–94. doi: 10.1109/51.677174
45. Abásolo D, James CJ, Hornero R. Non-linear analysis of intracranial electroencephalogram recordings with approximate entropy and Lempel-Ziv complexity for epileptic seizure detection. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Lyon (2007) 1953–6. doi: 10.1109/iembs.2007.4352700
46. Fernández A, López-Ibor MI, Turrero A, Santos JM, Morón MD, Hornero R, et al. Lempel-Ziv complexity in schizophrenia: a MEG study. *Clin Neurophysiol.* (2011) 122:2227–35. doi: 10.1016/j.clinph.2011.04.011
47. Akar SA, Kara S, Latifoglu F, Bilgiç V. Analysis of the complexity measures in the EEG of schizophrenia patients. *Int J Neural Syst.* (2016) 26:1650008. doi: 10.1142/s0129065716500088
48. Cerquera A, Gjini K, Bowyer SM, Boutros N. Comparing EEG nonlinearity in deficit and nondeficit schizophrenia patients: preliminary data. *Clin EEG Neurosci.* (2017) 48:376–82. doi: 10.1177/1550059417715388
49. Bob P, Roman R, Svetlak M, Kukleta M, Chladek J, Brazdil M. Preictal dynamics of EEG complexity in intracranially recorded epileptic seizure: a case report. *Medicine.* (2014) 93:e151. doi: 10.1097/md.0000000000000151

50. Kondakor I, Toth M, Wackermann J, Gyimesi C, Czopf J, Clemens B. Distribution of spatial complexity of EEG in idiopathic generalized epilepsy and its change after chronic valproate therapy. *Brain Topogr.* (2005) 18:115–23. doi: 10.1007/s10548-005-0280-z
51. Kwan P, Brodie MJ. Early identification of refractory epilepsy. *N Engl J Med.* (2000) 342:314–9. doi: 10.1056/nejm200002033420503
52. Arya R, Glauser TA. Pharmacotherapy of focal epilepsy in children: a systematic review of approved agents. *CNS Drugs.* (2013) 27:273–86. doi: 10.1007/s40263-013-0048-z
53. Leppik IE. Three new drugs for epilepsy: levetiracetam, oxcarbazepine, and zonisamide. *J Child Neurol.* (2002) 17(Suppl 1):S53–7. doi: 10.1177/088307380201700107010
54. FDA. *Table of Pharmacogenomic Biomarkers in Drug Labeling*. (2020). Available online at: https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling

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 Medicine

REVIEW
published: 25 January 2022
doi: 10.3389/fmed.2021.734643

Long Non-coding RNA MALAT1: A Key Player in Liver Diseases

Juan Lu[1†], *Jing Guo*[1†], *Jun Liu*[1], *Xiaomin Mao*[2] *and Kaijin Xu*[1*]

[1] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Medicine, The First Affiliated Hospital, Zhejiang University, Hangzhou, China, [2] Haining People' Hospital, Haining Branch, The First Affiliated Hospital, College of Medicine, Zhejiang University, Haining, China

Long non-coding RNAs (lncRNAs) exceed 200 nucleotides in length are considered to be involved in both developmental processes and various diseases. Here, we focus on lncRNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), which was one of the most important lncRNAs in proliferation, apoptosis, and migration. MALAT1 plays a regulatory role in liver diseases, including hepatic fibrosis, liver regeneration, liver cancer, and fatty liver diseases. In the current review, we summarize the latest literature about the function roles of MALAT1 in liver disorders. Probing the regulatory mechanism and cross talk of MALAT1 with other signaling pathways of pathological processes would improve the prognosis, diagnosis of liver diseases, and offer a promising candidate target for therapeutic interventions.

Keywords: lncRNAs, MALAT1, mechanism, signaling pathways, liver diseases

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute (NCI),
United States

Reviewed by:
En-Qiang Chen,
Sichuan University, China
Prasanth Puthanveetil,
Midwestern University, United States

***Correspondence:**
Kaijin Xu
zdyxyxkj@zju.edu.cn

[†] *These authors share first authorship*

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *01 July 2021*
Accepted: *13 December 2021*
Published: *25 January 2022*

Citation:
Lu J, Guo J, Liu J, Mao X and Xu K (2022) Long Non-coding RNA MALAT1: A Key Player in Liver Diseases. Front. Med. 8:734643. doi: 10.3389/fmed.2021.734643

Seventy-five percent of the human genome generates transcripts, yet only about two percent of the genome encodes proteins. The majority of transcripts, including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), do not encode proteins (1). Small non-coding RNAs of ∼22 nucleotides regulate the translation and stability of mRNA at the post-transcriptional level. Several biological processes are regulated by lncRNAs (more than 200 nucleotides), including carcinogenesis, development, and differentiation (2, 3). lncRNAs modulate various cellular processes, including nuclear organization and transcriptional and post-transcriptional modulation of gene expression (4, 5). lncRNAs serve as competitive endogenous RNAs (ceRNAs) in a regulatory network by "sponging" target miRNAs to regulate mRNA expression. This regulatory network is implicated in cancer development, apoptosis, and drug resistance (6–8).

To date, over 50,000 human lncRNAs have been identified (9). One of the most extensively studied of lncRNAs is Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1, ENSG00000251562), located in human chromosome 11q13.1 (mouse chromosome 19qA); MALAT1 is also known as nuclear-enriched abundant transcript first described to be associated with non-small cell lung cancer (10–12). Elevated MALAT1 expression is implicated in hyperproliferation, cellular and molecular functions, metastasis, and poor prognosis (13, 14). MALAT1 is overexpressed in a variety of human diseases in the form of a ceRNA network, which is important for gene expression, proliferation, and metastasis (15). The pathological processes would be activated once the level of MALAT1 expression changes to abnormal condition due to various endogenous or exogenous inducers. Here, we summarize the roles of MALAT1 in liver diseases including hepatic fibrosis, hepatic carcinoma, liver regeneration, and fatty liver diseases.

MOLECULAR FUNCTION OF MALAT1

MALAT1 is preferentially associated with transcriptionally active genes in the nucleus (16). MALAT1 plays a very critical cellular function for normal physiology. Given the nuclear localization of nuclear speckles, MALAT1 could be implicated in the regulation of alternative splicing in terms of binding splicing factors (17). The interaction between MALAT1 and the splicing factors of serine/arginine-rich (SR) proteins regulates selective splicing by modulating phosphorylation and distribution in the nuclear macular region. MALAT1 also regulates SR proteins by affecting the localization and activities of shear factor kinases such as serine/arginine protein kinases 1 (SRPK1) (18, 19). As a result of MALAT1 transcripts located in the nucleus, MALAT1 combined with release of pre-mRNA splicing factors via antisense oligonucleotides triggers could induce overexpression of the related factors. Therefore, MALAT1 might serve as an "anchor point" for localization of certain genes near nuclear speckles and induce transcriptional activation to affect the processing of their RNAs, implicating MALAT1 in transcriptional regulation (20). The major mechanisms of post-transcriptional regulation of MALAT1 include alternative splicing, protein activities, and competitive ceRNAs (15). MALAT1 is also expressed in vascular endothelial cells and plays important roles in regulating vascular growth. The cell cycle inhibitory proteins were also significantly increased accompanied by the depletion of MALAT1 (21).

MicroRNAs bind to the 3'UTR of their target genes and negatively regulate gene expression through depressing translation or promoting mRNA attenuation. This strategy was proposed initially by Poliseno to explain how mRNAs communicate with lncRNAs through miRNA response elements (MREs) as "language" (22). In the process of RNA-miRNA regulation, MREs act as binding sites for ceRNAs, which positively adjust miRNAs availability to bind their target mRNAs (23). MALAT1 also followed the ceRNA regulatory system through the negatively regulation between MALAT1 and micRNAs (**Figure 1**).

MALAT1 binds to chromatin of activated transcribing genes, transcription factors and transcriptional co-activators and regulates their expression at the transcriptional level to promote cell proliferation and inhibit cell apoptosis. MALAT1 was also triggered to upregulate transcriptional activators of proteosome genes (24).

MALAT1 –AN INDUCER IN HEPATIC FIBROSIS

Hepatic fibrosis is characterized by chronic abnormal hyperplasia and accumulation of large amounts of extracellular matrix (ECM) [including smooth muscle actin (SMA) and type I collagen], and the release of proinflammatory and profibrotic factors. Fibrosis can result from a variety of chronic liver diseases such as viral hepatitis, alcoholism, drug abuse, metabolic syndrome, genetic metabolic diseases, and autoimmune hepatitis (25). The regulatory mechanism of liver fibrosis is not entirely clear. The activation of hepatic stellate cells (HSCs), the resident perisinusoidal cell type, is critical in the development of liver fibrosis (26). Activated HSCs are considered proliferative cells that secrete profibrogenic mediators and express ECM; they may also be involved in fibrosis progression (27). Yu et al. (28) found that MALAT1 was significantly upregulated in activated HSCs and negatively correlated with the expression of miR-101b in in the mouse liver fibrosis model induced by carbon tetrachloride (CCl4). MALAT1 and ras-related C3 botulinum toxin substrate 1 (Rac1) are targets of miR-101b, and the former acts as a ceRNA to enhance the expression of Rac1, thus promoting HSC proliferation and activation. Dai et al. (29) found that MALAT1 extracted from arsenite-treated human hepatocytes promoted the activation of LX-2 HSCs (an immortalized HSC cell lines) by binding miRNA-26b. Wu et al. (30) reported that MALAT1 influenced the progression of hepatic fibrosis by repressing the expression and function of silent information regulator 1 (SIRT1), a nicotinamide adenine nicotinamide (NAD+)-dependent III class of histone deacetylases in the Sirtuin family (16). SIRT1 suppressed the ability of the promoters of fibrogenic genes, such as collagen type I, to bind Smad3, a downstream mediator of the TGF-β signaling pathway (31, 32). In short, MALAT1 promotes HSC activation by blocking SIRT1-mediated inhibition of the TGF-β signaling pathway in hepatic fibrosis.

MALAT1—A REGULATING ROLE IN LIVER REGENERATION

The normal liver has a strong regenerative ability and can maintain the original liver volume by inducing mitosis of its own cells, or tissue repair through rapid division and proliferation from endogenous liver stem cells (such as liver oval cells) to mature liver cells (33). Because of the limited regenerative ability of endogenous liver stem cells, exogenous measures are needed to promote liver regeneration and maintain the liver function of patients with severe liver injury or liver failure (34). Liver cells begin proliferate and divide in the initial period. The proliferation of liver is not only completed by parenchyma cells, but by a variety of cells which also participate in the coordination and completion at the stage of proliferation (35). The timely and automatic termination of hepatocyte proliferation is distinct from the unlimited proliferation of liver tumors in the termination stage, which are are important for the liver regulation of growth, development, and differentiation (36). The interaction and regulatory factors were activated rested during the biological processes of liver development, liver regeneration and hepatic carcinoma. Tripathi et al. (37) showed that MALAT1 promotes the proliferation of normal cells. MALAT1 also promotes the expression of cell cycle genes (such as the transcription factor B-myb) and progression from G1 to S phase, thus inducing mitosis. Knockout of MALAT1 activated p53 and its reporter gene, a downstream gene of MALAT1. Maxy et al. (38, 39) studied the expression of MALAT1 in various tissues and organs of mice and found that MALAT1 expression in mice was highest in the liver followed by the kidney, lowest in skeletal muscle. MALAT1 expression was elevated, and its inhibition or knockout may promote the proliferation, migration,

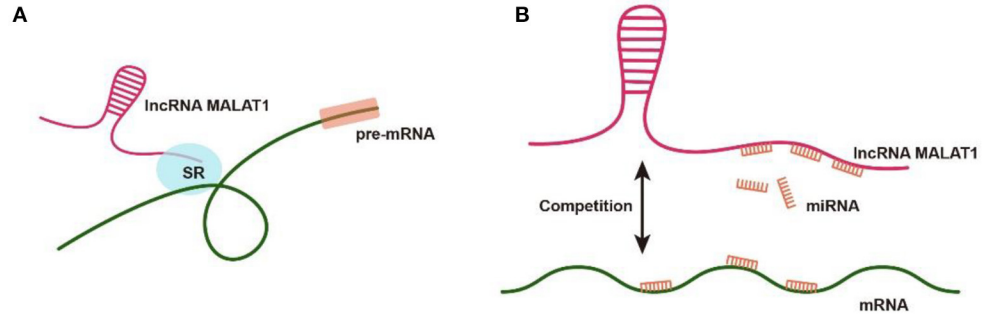

FIGURE 1 | Molecular Function of MALAT1. **(A)** The transcriptional regulation of MALAT1 by interaction between MALAT1 and SR splicing factors. **(B)** The post-transcriptional regulation of MALAT1 via competitive ceRNAs.

and tubulation, and suppress apoptosis, of endothelial cells by activating the PI3K/Akt signaling pathway (40).

Liver regeneration is initiated immediately after liver resection, and upregulation of hepatocyte growth factor (HGF) induces the expression of MALAT1, damaging the stability of the p-catenin degradation complex (41). The resulting increased total p-catenin activates the Wnt/β-catenin signaling pathway and upregulates cyclinD1, promoting progression from G1 to S phase and shortening the cell cycle, thereby accelerating liver cell proliferation and liver regeneration (42). Deficiency of MALAT1 inhibits VEGFR2 expression, reduces angiogenesis, perfusion, and functional recovery of ischemic hind limbs in mice, and inhibits blood flow recovery and capillary density in gastrocnemeal muscle tissue after ischemia, suggesting that MALAT1 affects angiogenesis via multiple mechanisms (21). MALAT1 regulates angiogenesis and immune responses (43). MiR-3064-5p inhibits the FOXA1/CD24/Src pathway to exert an antiangiogenic effect, whereas MALAT1 adsorbs miR-3064-5p, thereby alleviating the inhibition of FOXA1 and promoting hepatic hemangiogenesis (44). Hou et al. (45, 46) confirmed that MALAT1 is an important upstream target of miR-140, which inhibits VEGF-A expression and M2 macrophage polarization, thereby inhibiting angiogenesis and immunosuppression.

MALAT1—A NEW THERAPEUTIC TARGET IN HEPATIC CARCINOMA

The expression of MALAT1 is significantly increased in HCC tissues and cell lines, which promotes HCC proliferation and metastasis and inhibits HCC cell apoptosis by acting as an oncogene (47–49). Silencing of MALAT1 reduces the proliferation, invasion, and migration of cancer cells, and induces their apoptosis (50). The upstream of MALAT1 contains 5 specific protein 1/3(SP1/3) binding sites, and the combined regulation of SP1 and SP3 in cancer cells promote the expression of MALAT1. Yes-related protein (YAP) upregulates the expression of MALAT1 at the transcriptional and post-transcriptional levels, whereas serine/arginine splicing factor 1 (SRSF1) rich in serine/arginine had the opposite effect. Overexpression of YAP reduces SRSF1 nuclear retention (51).

MALAT1 plays a role in HCC cell proliferation and apoptosis by multiple pathways. Upregulation of MALAT1 promotes cancer cell proliferation, and its downregulation promotes cancer cell death and autophagy. Malakar et al. (52) found that high expression of MALAT1 in HCC cells upregulates oncogenic shear factor SRSF1 and activates the mTOR pathway, thereby promoting the proliferation and survival of HCC cells. Peng et al. (50) found that MALAT1 expression regulates the proliferation, apoptosis, and autophagy of HCC cells by adsorption of miR-146a, whereas downregulation of miR-146a upregulates PI3K, modulating the phosphorylation of downstream Akt and mTOR. Therefore, apoptosis and autophagy of HCC cells can be inhibited by targeting the PI3K/Akt/mTOR signaling axis. Liu et al. (53) showed that MALAT1 acted as a molecular sponge to absorb miR-195, which inhibited its downstream target EGFR, inducing the activation of PI3K/Akt and JAK/STAT pathways by overexpression of EGFR, thus promote the growth activity of HCC cells. Chen et al. (54) reported that MALAT1 regulates the expression of zinc finger E-box binding homeobox (ZEB1) by sponging miR-143-3p. In conclusion, the up- or downregulation of MALAT1 is related to the proliferation and apoptosis of HCC cells. MALAT1 acts as a molecular sponge to regulate miRNA signaling pathways affecting downstream factors and promotes HCC cell proliferation and inhibits their apoptosis.

HCC invasion and metastasis are aggravated via signaling pathways and MALAT1. YAP1 in vascular endothelial cells reduces vascular proliferation (55). Exosomes containing MALAT1 are released into the tumor microenvironment, inhibiting and depleting YAP1, activating ERK1/2 signal transduction, and enhancing the expression of MMP2 and MMP9, thereby promoting tumor invasion and metastasis. The tumor transcription factor FOXM1 is the target of miR-125a-3p, targeting of which by MALAT1 upregulates FOXM1 expression and promotes HCC invasion and migration (51, 56). SNAI1, a key transcription factor in the epithelial–mesenchymal transition, is also a direct target of miR-22, which could be absorbed by MALAT1, promoting the enrichment of enhancer of zeste homolog 2 (EZH2) to inhibit miR-22 transcription, thereby upregulating SNAI1 expression and facilitating HCC invasion and distant metastasis. Li et al. (57) reported that MALAT1 acted as a molecular sponge for miR-146b-5p, inhibiting HCC growth

and metastasis by targeting Akt phosphorylation mediated by TNF receptor-related factor 6. Hou et al. (58) found that MALAT1 reduced the inhibition of SIRT1 by miR-204 by competitively binding miR-204. Chen et al. (54) showed that MALAT1 regulates ZEB1 expression by sponging miR-143-3p, promoting HCC invasion and metastasis.

MALAT1 also promotes glycolysis in HCC. Aerobic glycolysis (Warburg effect) is a marker of tumor cells' ability to escape apoptosis to promote proliferation and migration. It could provide material basis for tumor cells and create an acidic microenvironment (59, 60). Malakar et al. (61) found that MALAT1 enhances the translation of metabolic transcription factor TCF7L2 by upregulating shear factor SRSF1 and activating the mTORC1-4EBP1 axis to upregulate glycolysis genes and inhibit gluconeogenesis, promoting the development of HCC (62). Additionally, it has been shown that MALAT1 enhances glycolysis in liver, and inhibits gluconeogenesis, via elevated translation of the transcription factor TCF7L2 and as such also plays a crucial role in metabolic stress (63). MALAT1 promotes the progression of inflammation-associated HCC. Huang et al. (64) reported that MALAT1 induces the secretion of inflammatory cytokines, promoting the progression of inflammation-related HCC by mobilizing chromatin and remodeling subunit brahma-related gene 1 (BRG1) to the promoter region of the inflammatory cytokines IL-6 and C-X-C motif chemokine ligand 8 (CXCL8).

Cancer stem cells (CSCs) are characterized by self-renewal and differentiation and are considered the seeds of tumor genesis, development, and metastasis. MALAT1 promotes HCC stem cell properties. The higher the proportion of CSCs, the more aggressive the tumor (65). He et al. (4) reported that HBx protein induced CSC production in HCC via the PI3K/Akt signaling pathway (66). MALAT1 emerged as the function of competing endogenous RNA, preventing miR-124-mediated inhibition of PI3K/Akt signaling. This induced CSC characteristics and ultimately promoted HBV-associated HCC. Chen et al. (67) reported that circ-MALAT1 generated by reverse shear of lncRNA MALAT1 acts as a "brake" in the ribosome, and forms a complex with the ribosome and mRNA, preventing the transcription factor PAX5 mRNA translation and promoting CSC self-renewal. However, circ-MALAT1 also acts as a sponge for miR-6887-3p to enhance the phosphorylation of JAK2, activating the JAK2/STAT3 signaling pathway and promoting CSC self-renewal (68). Thus, MALAT1 represents both a promising cancer bio-marker as well as a potential therapeutic target for limiting metastatic growth.

MALAT1 IN OTHER LIVER DISEASES

MALAT1 significantly inhibits palmitic acid-induced lipid accumulation and increases expression of SREBP-1c, an important regulator of cholesterol and fatty acid synthesis in abnormal lipid metabolism and fatty liver disease (69, 70).

TABLE 1 | Different roles of MALAT1 under different liver diseases.

Hepatic fibrosis	Promoting proliferation by ceRNA of miR-101b
Liver regeneration	Promoting proliferation by ceRNA of miR-3064-5p140
Hepatic carcinoma	Promoting proliferation by ceRNA of miR-146a/195/143-3p/ 22/204/146b/125a-3p
	Promoting proliferation by upregulating SP1/3/YAP/SRSF1
	Promoting proliferation by Warburg effect
	Promoting proliferation by ceRNA of miR-124/6887-3p
Fatty liver disease	Promoting proliferation by upregulating SREBP-1c/CXCL5/ transaminase

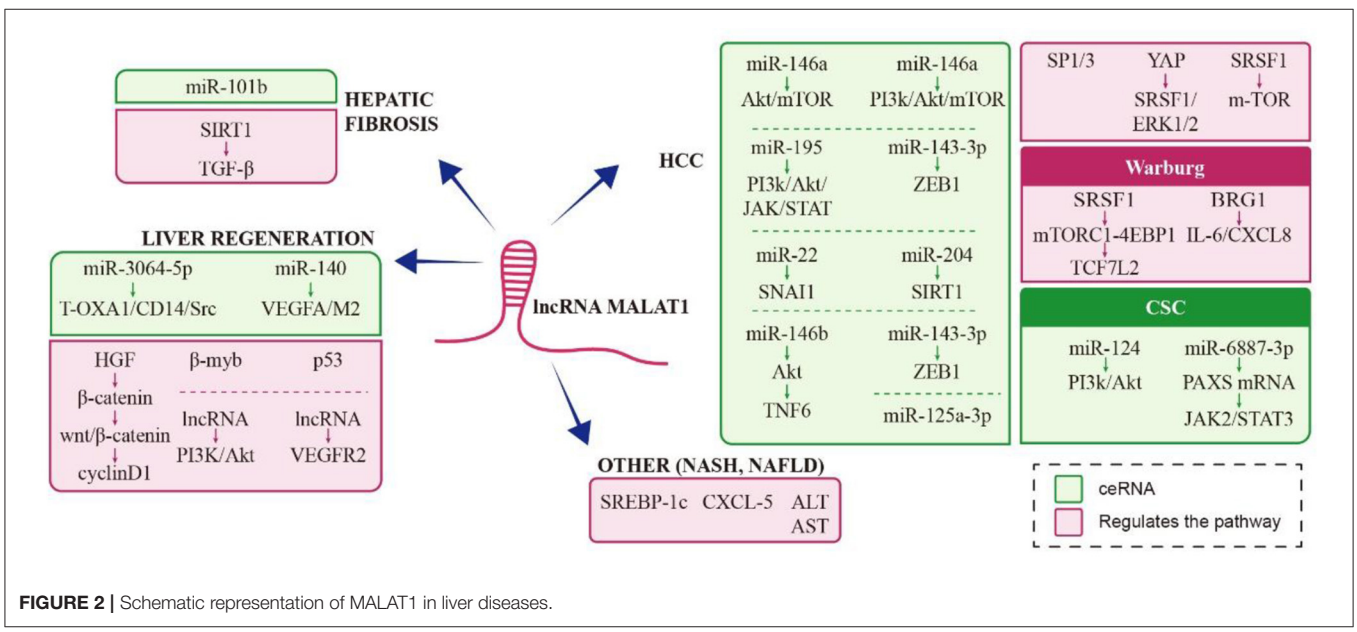

FIGURE 2 | Schematic representation of MALAT1 in liver diseases.

MALAT1 abundance in liver tissue is closely related to the pathological changes in non-alcoholic fatty liver disease (NAFLD), a clinicopathologic syndrome characterized by diffuse bullae steatosis (abnormal accumulation of lipids in liver tissues) not caused by alcohol or other hepatotoxic factors. MALAT1 may mediate chemokines by regulating C-X-C motif chemokine ligand 5 (CXCL5) in hepatic stellate cells in the occurrence of non-alcoholic steatohepatitis (NASH) and fibrosis in patients with NAFLD (71, 72). MALAT1 abundance increases significantly in NASH with hepatocyte ballooning degeneration and lobular inflammation, and in hepatocyte dysfunction with elevated alanine aminotransferase, aspartate aminotransferase, and alkaline phosphatase. The schematic representation of MALAT1 in liver diseases was **Figure 2**. A form was also presented to explain different roles of MALAT1 under different liver diseases and how MALAT1 influences pathophysiology in **Table 1**.

CONCLUSION

LncRNA is a hot research topic in the field of liver disease in recent years. As a member of the lncRNA family, MALAT1 is a multi-functional lncRNA and an important regulator in hepatic fibrosis, liver regeneration, cancer, and fatty liver diseases. The molecular mechanisms mediated by one lncRNA would be complicated. Compared to most lncRNAs, MALAT1 is expressed at relatively high level in almost all human tissues in a variety of regulating pathways, thus makes it intricate to be targeted by simply silencing or overexpressing in pathological conditions. Further challenge to therapeutically measures would include small molecules specifically designed to intervene the gene-protein interaction for MALAT1 functions (73). A deeper understanding of the functions of MALAT1 and its interaction network will lay the foundation for the development of lncRNAs as therapeutic targets and as diagnostic or prognostic biomarkers for liver diseases.

AUTHOR CONTRIBUTIONS

JLu and KX structured the text and content. JLu and JG wrote and edited the manuscript. JLi generated the figures. XM reviewed the literature. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Independent Project Fund of the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, the National Key Research and Development Program of China (No. 2016YFC1101404/3), and Zhejiang Basic Public Welfare Research Program of China (No. LQ20H030012).

REFERENCES

1. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, et al. Distinguishing protein-coding, and noncoding genes in the human genome. *Proc Natl Acad Sci USA.* (2007) 104:19428–33. doi: 10.1073/pnas.0709013104
2. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet.* (2011) 12:861–74. doi: 10.1038/nrg3074
3. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* (2016) 17:47–62. doi: 10.1038/nrg.2015.10
4. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* (2009) 10:155–9. doi: 10.1038/nrg2521
5. Malissovas N, Ninou E, Michail A, Politis PK. Targeting long non-coding RNAs in nervous system cancers: new insights in prognosis, diagnosis and therapy. *Curr Med Chem.* (2019) 26:5649–63. doi: 10.2174/0929867325666180831170227
6. Lin Z, Li X, Zhan X, Sun L, Gao J, Cao Y, et al. Construction of competitive endogenous RNA network reveals regulatory role of long non-coding RNAs in type 2 diabetes mellitus. *J Cell Mol Med.* (2017) 21:3204–13. doi: 10.1111/jcmm.13224
7. Shi X, Sun M, Liu H, Yao Y, Song Y. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett.* (2013) 339:159–66. doi: 10.1016/j.canlet.2013.06.013
8. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature.* (2010) 465:182–7. doi: 10.1038/nature09033
9. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet.* (2015) 47:199–208. doi: 10.1038/ng.3192
10. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene.* (2003) 22:8031–41. doi: 10.1038/sj.onc.1206928
11. Tani H, Nakamura Y, Ijiri K, Akimitsu N. Stability of MALAT-1, a nuclear long non-coding RNA in mammalian cells, varies in various cancer cells. *Drug Discov Ther.* (2010) 4:235–9.
12. Ren D, Li H, Li R, Sun J, Guo P, Han H, et al. Novel insight into MALAT-1 in cancer: therapeutic targets and clinical applications. *Oncol Lett.* (2016) 11:1621–30. doi: 10.3892/ol.2016.4138
13. Liang T, Xu F, Wan P, Zhang L, Huang S, Yang N, et al. Malat-1 expression in bladder carcinoma tissues and its clinical significance. *Am J Transl Res.* (2021) 13:3555–60.
14. Fu S, Wang Y, Li H, Chen L, Liu Q. Regulatory networks of LncRNA MALAT-1 in cancer. *Cancer Manag Res.* (2020) 12:10181–98. doi: 10.2147/CMAR.S276022
15. Kong X, Wang J, Cao Y, Zhang H, Lu X, Wang Y, et al. The long noncoding RNA MALAT-1 functions as a competing endogenous RNA to regulate MSL2 expression by sponging miR-338-3p in myasthenia gravis. *J Cell Biochem.* (2019) 120:5542–50. doi: 10.1002/jcb.27838
16. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell.* (2014) 55:791–802. doi: 10.1016/j.molcel.2014.07.012
17. Sridhar B, Rivas-Astroza M, Nguyen TC, Chen W, Yan Z, Cao X, et al. Systematic mapping of RNA-chromatin interactions *In Vivo*. *Curr Biol.* (2017) 27:610–2. doi: 10.1016/j.cub.2017.01.068
18. Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* (2018) 46:2169–84. doi: 10.1093/nar/gky095
19. Luco RF. Retrotransposons jump into alternative-splicing regulation via a long noncoding RNA. *Nat Struct Mol Biol.* (2016) 23:952–4. doi: 10.1038/nsmb.3318
20. Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, et al. The noncoding RNA MALAT1 is a critical regulator of the

metastasis phenotype of lung cancer cells. *Cancer Res.* (2013) 73:1180–9. doi: 10.1158/0008-5472.CAN-12-2850
21. Michalik KM, You X, Manavski Y, Doddaballapur A, Zörnig M, Braun T, et al. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ Res.* (2014) 114:1389–97. doi: 10.1161/CIRCRESAHA.114.303265
22. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell.* (2011) 146:353–8. doi: 10.1016/j.cell.2011.07.014
23. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* (2010) 465:1033–8. doi: 10.1038/nature09144
24. Amodio N, Stamato MA, Juli G, Morelli E, Fulciniti M, Manzoni M, et al. Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia.* (2018) 32:1948–57. doi: 10.1038/s41375-018-0067-3
25. Teng KY, Ghoshal K. Role of noncoding RNAs as biomarker and therapeutic targets for liver fibrosis. *Gene Expr.* (2015) 16:155–62. doi: 10.3727/105221615X14399878166078
26. Zhang H, Sun D, Wang G, Cui S, Field RA, Li J, et al. Alogliptin alleviates liver fibrosis via suppression of activated hepatic stellate cell. *Biochem Biophys Res Commun.* (2019) 511:387–93. doi: 10.1016/j.bbrc.2019.02.065
27. Schon HT, Bartneck M, Borkham-Kamphorst E, Nattermann J, Lammers T, Tacke F, et al. Pharmacological intervention in hepatic stellate cell activation and hepatic fibrosis. *Front Pharmacol.* (2016) 7:33. doi: 10.3389/fphar.2016.00033
28. Yu F, Lu Z, Cai J, Huang K, Chen B, Li G, et al. MALAT1 functions as a competing endogenous RNA to mediate Rac1 expression by sequestering miR-101b in liver fibrosis. *Cell Cycle.* (2015) 14:3885–96. doi: 10.1080/15384101.2015.1120917
29. Dai X, Chen C, Xue J, Xiao T, Mostofa G, Wang D, et al. Exosomal MALAT1 derived from hepatic cells is involved in the activation of hepatic stellate cells via miRNA-26b in fibrosis induced by arsenite. *Toxicol Lett.* (2019) 316:73–84. doi: 10.1016/j.toxlet.2019.09.008
30. Wu Y, Liu X, Zhou Q, Huang C, Meng X, Xu F, et al. Silent information regulator 1 (SIRT1) ameliorates liver fibrosis via promoting activated stellate cell apoptosis and reversion. *Toxicol Appl Pharmacol.* (2015) 289:163–76. doi: 10.1016/j.taap.2015.09.028
31. Sun L, Fan Z, Chen J, Tian W, Li M, Xu H, et al. Corrigendum: transcriptional repression of SIRT1 by protein inhibitor of activated STAT 4 (PIAS4) in hepatic stellate cells contributes to liver fibrosis. *Sci Rep.* (2016) 6:30513. doi: 10.1038/srep30513
32. Wei J, Ghosh AK, Chu H, Fang F, Hinchcliff ME, Wang J, et al. The histone deacetylase sirtuin 1 is reduced in systemic sclerosis and abrogates fibrotic responses by targeting transforming growth factor beta signaling. *Arthritis Rheumatol.* (2015) 67:1323–34. doi: 10.1002/art.39061
33. Bangru S, Kalsotra A. Cellular and molecular basis of liver regeneration. *Semin Cell Dev Biol.* (2020) 100:74–87. doi: 10.1016/j.semcdb.2019.12.004
34. Ray K. Therapy: targeting liver tissue repair and regeneration. *Nat Rev Gastroenterol Hepatol.* (2016) 13:559. doi: 10.1038/nrgastro.2016.146
35. Michalopoulos GK, Bhushan B. Liver regeneration: biological and pathological mechanisms and implications. *Nat Rev Gastroenterol Hepatol.* (2021) 18:40–55. doi: 10.1038/s41575-020-0342-4
36. Xu D, Yang F, Yuan JH, Zhang L, Bi HS, Zhou CC, et al. Long noncoding RNAs associated with liver regeneration 1 accelerates hepatocyte proliferation during liver regeneration by activating Wnt/beta-catenin signaling. *Hepatology.* (2013) 58:739–51. doi: 10.1002/hep.26361
37. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell.* (2010) 3:925–38. doi: 10.1016/j.molcel.2010.08.011
38. Jeffers LK, Duan K, Ellies LG, Seaman WT, Burger-Calderon RA, Diatchenko LB, et al. Correlation of transcription of MALAT-1, a novel noncoding RNA, with deregulated expression of tumor suppressor p53 in small DNA tumor virus models. *J Cancer Ther.* (2013) 4:3. doi: 10.4236/jct.2013.43094
39. Pruszko M, Milano E, Forcato M, Donzelli S, Ganci F, Di Agostino S, et al. The mutant p53-ID4 complex controls VEGFA isoforms by recruiting lncRNA MALAT1. *EMBO Rep.* (2017) 18:1331–51. doi: 10.15252/embr.201643370
40. Zhang SH, Zhang SG, Zhou P, Wei X, Mao XD, Lin SG, et al. LncRNA MALAT1 affects high glucose-induced endothelial cell proliferation, apoptosis, migration and angiogenesis by regulating the PI3K/Akt signaling pathway. *Eur Rev Med Pharmacol Sci.* (2019) 23:8551–9. doi: 10.26355/eurrev_201910_19170
41. Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* (2013) 9:e1003368. doi: 10.1371/journal.pgen.1003368
42. Yang F, Yi F, Han X, Du Q, Liang Z. MALAT-1 interacts with hnRNP C in cell cycle regulation. *FEBS Lett.* (2013) 587:3175–81. doi: 10.1016/j.febslet.2013.07.048
43. Ma Z, Zhang J, Xu X, Qu Y, Dong H, Dang J, et al. LncRNA expression profile during autophagy and Malat1 function in macrophages. *PLoS One.* (2019) 14:e0221104. doi: 10.1371/journal.pone.0221104
44. Zhang P, Ha M, Li L, Huang X, Liu C. MicroRNA-3064-5p sponged by MALAT1 suppresses angiogenesis in human hepatocellular carcinoma by targeting the FOXA1/CD24/Src pathway. *FASEB J.* (2020) 34:66–81. doi: 10.1096/fj.201901834R
45. Bao M, Liu G, Song J, Gao Y. Long non-coding RNA MALAT1 promotes odontogenic differentiation of human dental pulp stem cells by impairing microRNA-140-5p-dependent downregulation of GIT2. *Cell Tissue Res.* (2020) 38:487–98. doi: 10.1007/s00441-020-03246-1
46. Hou ZH, Xu XW, Fu XY, Zhou LD, Liu SP, Tan DM. Long non-coding RNA MALAT1 promotes angiogenesis and immunosuppressive properties of HCC cells by sponging miR-140. *Am J Physiol Cell Physiol.* (2020) 318:C649–63. doi: 10.1152/ajpcell.00510.2018
47. Toraih EA, Ellawindy A, Fala SY, Al Ageeli E, Gouda NS, Fawzy MS, et al. Oncogenic long noncoding RNA MALAT1 and HCV-related hepatocellular carcinoma. *Biomed Pharmacother.* (2018) 102:653–69. doi: 10.1016/j.biopha.2018.03.105
48. Goyal B, Yadav SRM, Awasthee N, Gupta S, Kunnumakkara AB, Gupta SC. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochim Biophys Acta Rev Cancer.* (2021) 1875:188502. doi: 10.1016/j.bbcan.2021.188502
49. Wang C, Zhang Q, Hu Y, Zhu J, Yang J. Emerging role of long non-coding RNA MALAT1 in predicting clinical outcomes of patients with digestive system malignancies: a meta-analysis. *Oncol Lett.* (2019) 17:2159–70. doi: 10.3892/ol.2018.9875
50. Peng N, He J, Li J, Huang H, Huang W, Liao Y, et al. Long noncoding RNA MALAT1 inhibits the apoptosis and autophagy of hepatocellular carcinoma cell by targeting the microRNA-146a/PI3K/Akt/mTOR axis. *Cancer Cell Int.* (2020) 20:165. doi: 10.1186/s12935-020-01231-w
51. Wang J, Wang H, Zhang Y, Zhen N, Zhang L, Qiao Y, et al. Mutual inhibition between YAP and SRSF1 maintains long non-coding RNA, Malat1-induced tumourigenesis in liver cancer. *Cell Signal.* (2014) 26:1048–59. doi: 10.1016/j.cellsig.2014.01.022
52. Malakar P, Shilo A, Mogilevsky A, Stein I, Pikarsky E, Nevo Y, et al. Long noncoding RNA MALAT1 promotes hepatocellular carcinoma development by SRSF1 upregulation and mTOR activation. *Cancer Res.* (2017) 77:1155–67. doi: 10.1158/0008-5472.CAN-16-1508
53. Liu D, Zhu Y, Pang J, Weng X, Feng X, Guo Y. Knockdown of long non-coding RNA MALAT1 inhibits growth and motility of human hepatoma cells via modulation of miR-195. *J Cell Biochem.* (2018) 119:1368–80. doi: 10.1002/jcb.26297
54. Chen L, Yao H, Wang K, Liu X. Long non-coding RNA MALAT1 regulates ZEB1 expression by sponging miR-143-3p and promotes hepatocellular carcinoma progression. *J Cell Biochem.* (2017) 118:4836–43. doi: 10.1002/jcb.26158
55. Li Y, Zhang X, Zheng Q, Zhang Y, Ma Y, Zhu C, et al. YAP1 inhibition in HUVECs Is associated with released exosomes and increased hepatocarcinoma invasion and metastasis. *Mol Ther Nucleic Acids.* (2020) 21:86–97. doi: 10.1016/j.omtn.2020.05.021
56. Zhao L, Lou G, Li A, Liu Y. lncRNA MALAT1 modulates cancer stem cell properties of liver cancer cells by regulating YAP1 expression via miR375 sponging. *Mol Med Rep.* (2020) 22:1449–57. doi: 10.3892/mmr.2020.11196
57. Li C, Miao R, Liu S, Wan Y, Zhang S, Deng Y, et al. Down-regulation of miR-146b-5p by long noncoding RNA MALAT1 in hepatocellular

carcinoma promotes cancer growth and metastasis. *Oncotarget.* (2017) 8:28683–95. doi: 10.18632/oncotarget.15640
58. Hou Z, Xu X, Zhou L, Fu X, Tao S, Zhou J, et al. The long non-coding RNA MALAT1 promotes the migration and invasion of hepatocellular carcinoma by sponging miR-204 and releasing SIRT1. *Tumour Biol.* (2017) 39:1010428317718135. doi: 10.1177/1010428317718135
59. Das L, Vinayak M. Long term effect of curcumin in regulation of glycolytic pathway and angiogenesis via modulation of stress activated genes in prevention of cancer. *PLoS ONE.* (2014) 9:e99583. doi: 10.1371/journal.pone.0099583
60. Ma ZJ, Yan H, Wang YJ, Yang Y, Li XB, Shi AC, et al. Proteomics analysis demonstrating rosmarinic acid suppresses cell growth by blocking the glycolytic pathway in human HepG2 cells. *Biomed Pharmacother.* (2018) 105:334–49. doi: 10.1016/j.biopha.2018.05.129
61. Malakar P, Stein I, Saragovi A, Winkler R, Stern-Ginossar N, Berger M, et al. Long noncoding RNA MALAT1 regulates cancer glucose metabolism by enhancing mTOR-mediated translation of TCF7L2. *Cancer Res.* (2019) 79:2480–93. doi: 10.1158/0008-5472.CAN-18-1432
62. Chen W. Cancer statistics: updated cancer burden in China. *Chin J Cancer Res.* (2015) 27:1. doi: 10.3978/j.issn.1000-9604.2015.02.07
63. Sridhar B, Rivas-Astroza M, Nguyen TC, Chen W, Yan Z, Cao X, et al. Systematic mapping of RNA-chromatin interactions *In Vivo*. *Curr Biol.* (2017) 27-602–609. doi: 10.1016/j.cub.2017.01.011
64. Huang M, Wang H, Hu X, Cao X. lncRNA MALAT1 binds chromatin remodeling subunit BRG1 to epigenetically promote inflammation-related hepatocellular carcinoma progression. *Oncoimmunology.* (2019) 8:e1518628. doi: 10.1080/2162402X.2018.1518628
65. Han Y, Zhou L, Wu T, Huang Y, Cheng Z, Li X, et al. Downregulation of lncRNA-MALAT1 affects proliferation and the expression of stemness markers in glioma stem cell line SHG139S. *Cell Mol Neurobiol.* (2016) 36:1097–107. doi: 10.1007/s10571-015-0303-6
66. He B, Peng F, Li W, Jiang Y. Interaction of lncRNA-MALAT1 and miR-124 regulates HBx-induced cancer stem cell properties in HepG2 through PI3K/Akt signaling. *J Cell Biochem.* (2019) 120:2908–18. doi: 10.1002/jcb.26823
67. Chen L, Kong R, Wu C, Wang S, Liu Z, Liu S, et al. Circ-MALAT1 functions as both an mrna translation brake and a microRNA sponge to promote self-renewal of hepatocellular cancer stem cells. *Adv Sci.* (2020) 7:1900949. doi: 10.1002/advs.201900949
68. Chang HL, Bamodu OA, Ong JR, Lee WH, Yeh CT, Tsai JT. Targeting the epigenetic non-coding RNA MALAT1/Wnt signaling axis as a therapeutic approach to suppress stemness and metastasis in hepatocellular carcinoma. *Cells.* (2020) 9:1020. doi: 10.3390/cells9041020
69. Oliva J, Bardag-Gorce F, French BA, Li J, French SW. The regulation of non-coding RNA expression in the liver of mice fed DDC. *Exp Mol Pathol.* (2009) 87:12–9. doi: 10.1016/j.yexmp.2009.03.006
70. Yan C, Chen J, Chen N. Long noncoding RNA MALAT1 promotes hepatic steatosis and insulin resistance by increasing nuclear SREBP-1c protein stability. *Sci Rep.* (2016) 6:22640. doi: 10.1038/srep22640
71. Leti F, Legendre C, Still CD, Chu X, Petrick A, Gerhard GS, et al. Altered expression of MALAT1 lncRNA in nonalcoholic steatohepatitis fibrosis regulates CXCL5 in hepatic stellate cells. *Transl Res.* (2017) 190:25–39.e21. doi: 10.1016/j.trsl.2017.09.001
72. Li JZ, Ye LH, Wang DH, Zhang HC, Li TY, Liu ZQ, et al. The identify role and molecular mechanism of the MALAT1/hsa-mir-20b-5p/TXNIP axis in liver inflammation caused by CHB in patients with chronic HBV infection complicated with NAFLD. *Virus Res.* (2021) 298:198405. doi: 10.1016/j.virusres.2021.198405
73. Zhang X, Tang X, Liu K, Hamblin MH, Yin KJ. Long non-coding RNA malat1 regulates cerebrovascular pathologies in ischemic stroke. *J Neurosci.* (2017) 37:1797–806. doi: 10.1523/JNEUROSCI.3389-16.2017

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

Publisher's Note: 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.

GLOSSARY

Abbreviations

The following abbreviations are used in this manuscript:

General	Abbreviation full name
4EBP1	phosphorylated 4E-binding protein 1
AKP	alkaline phosphorus
ALT	Aminotransferase
AST	aspartate aminotransferase
BRG1	brahma-related gene 1
CCl4	carbon tetrachloride
ceRNAs	competitive endogenous RNAs
CSC	cancer stem cells
CXCL5	C-X-C motif chemokine ligand 5
CXCL8	C-X-C motif chemokine ligand 8
ECM	extracellular matrix
EZH2	enhancer of zeste homolog 2
FOXA1	forkhead box A1
HCC	hepatic carcinoma
HGF	hepatocyte growth factor
HSCs	hepatic stellate cells
lncRNA	long noncoding RNA
MALAT1	metastasis-associated lung adenocarcinoma transcript 1
miRNA	microRNA
MREs	microRNA response elements
mTORC1	mechanistic target of rapamycin complex 1
NAD+	nicotinamide adenine dinucleotide
NAFLD	non-alcoholic fatty liver disease
NASH	nonalcoholic steatohepatitis
NEAT2	uuclear-enriched abundant transcript 2
Rac1	ras-related C3 botulinum toxin substrate 1
SIRT1	silent information regulator 1
SMA	smooth muscle actin
SMA	smooth muscle actin
SNAI1	snail family transcriptional repressor 1
SP1/3	specific protein 1/3
SR proteins	serine/arginine-rich proteins
SRPK1	serine/arginine protein kinases 1
SRSF1	serine/arginine splicing factor 1
STAT	signal transducer and activator of transcription
TCF7L2	transcription factor 7-like 2
TNF	tumor necrosis factor
YAP	Yes-related protein
ZEB1	zinc finger E-box binding homeobox 1

Signaling Pathways

Signaling pathway	KEGG ID
AKT/mTOR signaling pathway	hsa04150
CXCL5 signaling pathway	hsa04060
CXCL8 signaling pathway	hsa04061
cyclinD1 signaling pathway	hsa05200
EFGR signaling pathway	hsa01521
ERK1/2 signaling pathway	hsa04933
FOXM1 signaling pathway	hsa04218
IL-6 signaling pathway	hsa05200
JAK/STAT signaling pathway	hsa04630
JAK/STAT signaling pathway	hsa04630
MAPK1 signaling pathway	hsa05166
MMP2/9 signaling pathway	hsa05219
mTOR1-4EBP1signaling pathway	hsa04218
P53 signaling pathway	hsa04115
PAX5 signaling pathway	hsa05202
PI3K/AKT signaling pathway	hsa04151
SIRT1 signaling pathway	hsa04211
SNAI1 signaling pathway	hsa04520
Src signaling pathway	hsa04510
TCF7L2 signaling pathway	hsa05200
TGF-β signaling pathway	hsa04068
VEGF signaling pathway	hsa04210
Wnt/b-catenin signaling pathway	hsa04310
YAP/SRSF signaling pathway	hsa04390
ZEB1 signaling pathway	hsa05206

MicroRNAs

Mirbase ID	HGNC symbol	Mirbase accession
has-miR-124	MIR124	MI0000443
has-miR-143-3p	MIR143	MI0000459
has-miR-146a	MIR146A	MI0000477
has-miR-146b	MIR146B	MI0003129
has-miR-22	MIR22	MI0000078
has-miR-3064-5p	MIR3064	MI0017375
hsa-miR-125a	MIR125A	MI0000469
hsa-miR-140	MIR140	MI0000456
hsa-miR-195	MIR195	MI0000489
hsa-miR-204	MIR204	MI0000284

Genes

HGNC symbol	Full name accession	Accession ID
4EBP1	phosphorylated 4E-binding protein 1	ENSG00000258870
AKT	serine/threonine kinase 1	ENSG00000142208
β-catenin	β-catenin	ENSG00000168036
BRG1	brahma-related gene 1	ENST00000644760
CXCL5	C-X-C motif chemokine ligand 5	ENSG00000163735
CXCL8	C-X-C motif chemokine ligand 8	ENSG00000169429
EGFR	growth factor receptor	ENSG00000146648
ERK1	extracellular signal-regulated kinase 1	ENSCING00000007067
EZH2	enhancer of zeste homolog 2	ENSG00000106462
FOXA1	forkhead box A1	ENSG00000129514
IL6	interleukin 6	ENSG00000136244
JAK2	janus kinase 2	ENSG00000096968
MALAT1	metastasis-associated lung adenocarcinoma transcript 1	ENSG00000251562
MAPK	mitogen-activated protein kinase	ENSG00000100030
MMP2	matrix metallopeptidase 2	ENSG00000087245
MMP-9	matrix metallopeptidase 9	ENSG00000100985
MTOR	mechanistic target of rapamycin kinase	ENSG00000198793
PI3K	phosphatidylinositol-4,5-bisphosphate 3-kinase	ENSG00000121879
SIRT1	silent information regulator 1	ENSG00000096717
SNAI1	snail family transcriptional repressor 1	ENSG00000124216
SRPK1	serine/arginine protein kinases 1	ENSG00000096063
TCF7L2	transcription factor 7-like 2	ENSG00000148737
TGF-β	transforming growth factor β	ENSG00000105329
VEGFR2	vascular endothelial growth factor receptor 2	ENSG00000112715
WET2	wnt family member 2	ENSG00000105989
ZEB1	zinc finger E-box binding homeobox 1	ENSG00000148516

frontiers in Medicine

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fmed.2022.757212

Limited Genomics Training Among Physicians Remains a Barrier to Genomics-Based Implementation of Precision Medicine

*Valerie M. Schaibley[1,2,3], Irma N. Ramos[1,4], Raymond L. Woosley[5], Steven Curry[5], Sean Hays[6] and Kenneth S. Ramos[1,5,7]**

[1] Center for Applied Genetics and Genomic Medicine, The University of Arizona Health Sciences, Tucson, AZ, United States, [2] Genetic Counseling Graduate Program, University of Arizona College of Medicine – Tucson, Tucson, AZ, United States, [3] Department of Cellular and Molecular Medicine, University of Arizona College of Medicine – Tucson, Tucson, AZ, United States, [4] Department of Health Promotion Sciences, University of Arizona Mel and Enid Zuckerman College of Public Health, Tucson, AZ, United States, [5] Division of Clinical Data Analytics and Decision Support, Department of Medicine, University of Arizona College of Medicine – Phoenix, Phoenix, AZ, United States, [6] SciPinion, Bozeman, MT, United States, [7] Institute of Biosciences and Technology, Texas A&M Health Science Center, Houston, TX, United States

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute (NCI), United States

Reviewed by:
Darrel Waggoner, University of Chicago, United States
Maurice Godfrey, University of Nebraska Medical Center, United States

***Correspondence:**
Kenneth S. Ramos
kramos@tamu.edu

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: 11 August 2021
Accepted: 18 February 2022
Published: 18 March 2022

Citation:
Schaibley VM, Ramos IN, Woosley RL, Curry S, Hays S and Ramos KS (2022) Limited Genomics Training Among Physicians Remains a Barrier to Genomics-Based Implementation of Precision Medicine. Front. Med. 9:757212. doi: 10.3389/fmed.2022.757212

The field of precision medicine has undergone significant growth over the past 10 years. Despite increasing applications of clinical genetic and genomic testing, studies consistently report limited knowledge of genetics and genomics among healthcare providers. This study explored barriers to the implementation of precision medicine by surveying physicians working in a large academic medical center. We assessed prior training in genetics, use of genetic testing in the clinic, desire for additional resources in genetics and genomic medicine and perceived barriers to successful integration of precision medicine. Only 20% of respondents reported moderate or extensive training in genetics. Physicians with limited or no training in genetics were less likely to have ordered a genetic test for any purpose. Furthermore, 41% of physicians responded that their lack of training identifying appropriate genetic tests and how to interpret genetic testing results was the most significant barrier to ordering genetic testing for their patients. These findings suggest that future efforts to realize the promise of precision medicine should focus on the integration of training programs for non-genetics trained healthcare providers.

Keywords: precision medicine, genetics, genomics, clinical care, training

INTRODUCTION

Precision medicine is quickly expanding into mainstream clinical care. With the projected growth of precision medicine in the coming years, genetics and genomics will become an increasingly mainstream component of routine clinical care (1). As the taxonomy of disease is redefined on the basis of genetic and genomic insights, providers and healthcare systems will increasingly need to integrate genetic testing into patient care (2). Already, the rise in genetic testing has increased demand for clinical genetics services in the United States (3). This has collided with workforce

shortages in clinical genetics, with many practices unable to take new patients and significant job vacancies for medical geneticists and genetic counselor positions across the country (3).

With increasing recognition of the link between chronic diseases and genomics and an inadequate number of healthcare providers (HCP) trained in genetics, the care of patients and families will fall to non-genetics trained providers. However, research has shown that many HCPs are not familiar with genetic testing procedures and test interpretation, and lack confidence in providing genetics-informed care and discussing genomics and genetics topics with their patients (4). Limited physician knowledge of genomics and genetic testing is repeatedly cited as a barrier for the implementation of precision medicine (5–8). Furthermore, lack of HCP knowledge of technological advances in the area of genetics has the potential to exacerbate health disparities in access to genetic testing among traditionally underserved populations (6).

In order to assess HCP knowledge and opinions on precision medicine and the integration of genomic medicine into their clinical practice, we surveyed HCPs working in a large academic medical center in central Arizona. The survey focused on the use of genetic testing in clinical practice, comfort using this type of data, and perceived barriers to expanding the use of clinical genetic testing.

MATERIALS AND METHODS

Survey questions were developed to evaluate physicians' knowledge and opinions of precision medicine. Participants were asked 12 questions to assess their use of genetic testing and ability to capitalize on this data in the clinic (**Table 1**). The complete survey is provided as **Supplementary Material**. Participants were not required to complete all questions in the survey.

A letter requesting participation in the survey was distributed to the entire physician staff at Banner - University Medical Center in Phoenix, Arizona (BUMCP). BUMCP is the primary academic medical center located in downtown Phoenix and is affiliated with the University of Arizona College of Medicine – Phoenix.

In compliance with institutional policy for human subjects research, the protocol was reviewed by the Banner Office for Human Research and The University of Arizona Institutional Review Board and deemed exempt from full review. All survey responses were anonymous. Sixty-four individuals completed the survey between October 30, 2017 and November 7, 2017 and the survey was closed on November 10, 2017. The survey was distributed, and responses were collected using a proprietary platform developed by SciPinion[1].

Survey results were analyzed using the Pearson's chi-squared test for independence or a Chi-square goodness of fit test. Yates correction was used for all Chi-squared analyses to correct for the low sample size. The 95% confidence intervals shown in **Figure 1** were calculated using the Wilson method. Statistical analysis was conducted in R version 3.5.1 and RStudio version 1.2.1335.

[1]https://scipinion.com/

Responses to the single open-ended question in the survey ("What is your definition of precision medicine?") were coded into five categories: (1) tailoring medical care to an individual, (2) using genetics to guide medical care (diagnosis, therapy, etc.), (3) understanding the combination of factors that influence health, like genetics, environment, and lifestyle, (4) basis for precision medicine is only based on genetic information, and (5) don't know/no response. Reponses could fall into more than one category.

Responses to the question "What is your level of training in genomics?" that fell into the "Moderate" or "Extensive" categories were combined into a "Moderate or Extensive" category due to only a single respondent indicating they had extensive training in genomics. The response categories were not defined for this question, as we wanted respondents to self-assess their level of training and experience in genomics as "None," "Limited," "Moderate," or "Extensive."

RESULTS

Study Summary and Description of Data

A 12-question survey was offered to providers working at a large academic medical center with practices in Phoenix, Arizona. The study was designed to assess the level of knowledge of precision medicine. An email invitation to the survey was distributed to physician providers at BUMCP using an internal listserv, with one reminder sent mid-way between October 30 and November 7, 2017. Sixty-four physicians completed the survey.

The majority of participants (86%) responded to all survey questions and most respondents were male (77%). Respondents completed medical training between 1971 and 2010. There was no significant difference between year of completion of medical training among respondents (Pearson's $\chi^2 = 4.75$, $p = 0.31$). About half of respondents indicated that they had previously ordered a genetic test for a patient, with 52% of respondents reporting use of a genetic test for diagnosis in the past (**Table 1**). The survey showed that 86% of providers would welcome assistance with the interpretation of genetic testing results and that they would consult an expert in genomic medicine if the option was available (**Table 1**). Most providers (78%) indicated that they would attend training, if available, and that 85% believed precision medicine will define standards of care in the future (**Table 1**).

Physician Definitions of Precision Medicine

Survey participants were asked to define precision medicine in their own words. A total of 59 responses to this question were provided. Three of the responses were specific, including a description of precision medicine as incorporating not just genetics, but other aspects of a patient's health as well:

> *"A treatment in medicine designed to take in account the patient's illness in terms of variability in their lifestyle, genetic genome and environment. The treatment is aimed at defining particular group of patients with particular genetic genome that therapy maybe*

TABLE 1 | Summary statistics for binary response questions.

Question	N "Yes" responses	Percentage "Yes" responses
Have you ever ordered a genetic-based test for diagnostic purposes?	33	52%
Have you ever ordered a genetic test for one of your patients?	36	56%
Do you feel you need help interpreting clinical genomic data if these data was made available to you?	55	86%
Do you think precision medicine will help define standards of care in medicine?	53	85%
Would you place a consult to a physician with expertise in genomic medicine if available at your institution?	55	86%
Would you like to attend trainings on precision medicine and genomics based testing for diagnostic purposes?	50	78%

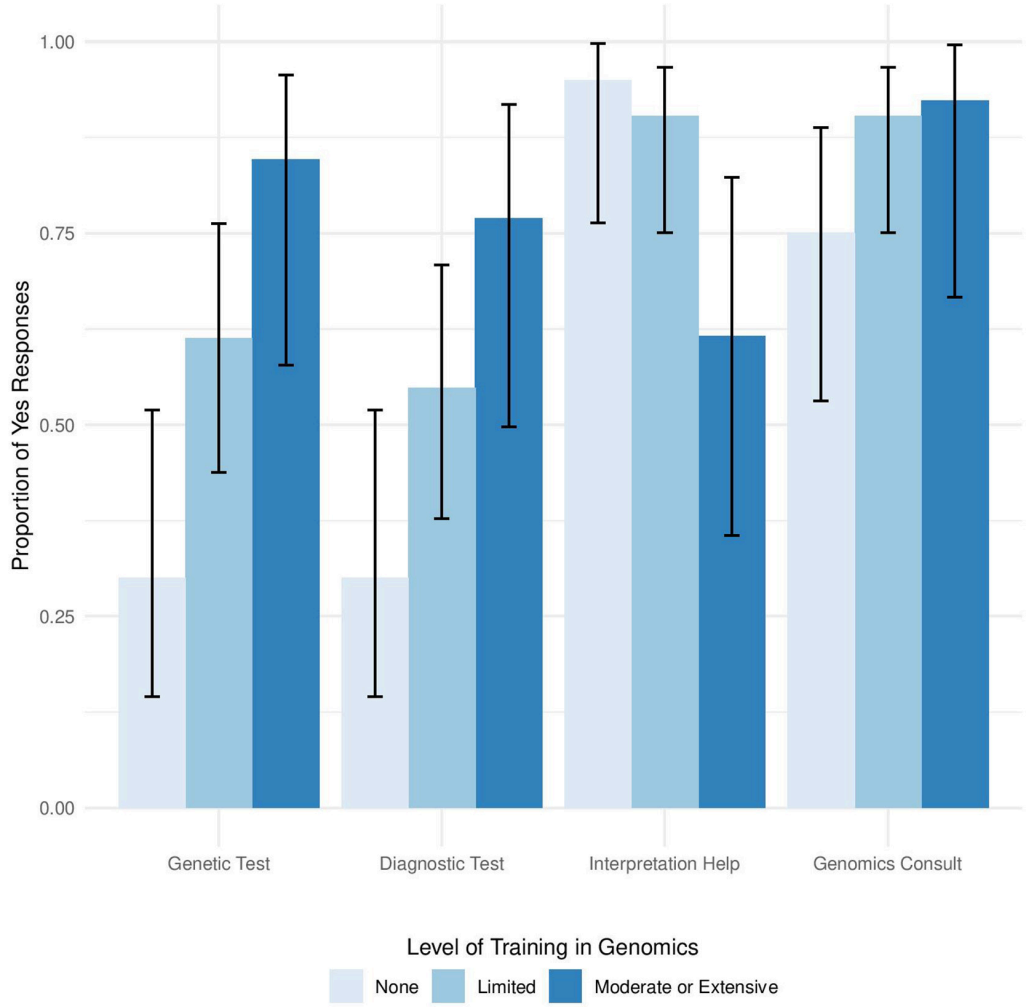

FIGURE 1 | Level of genomics training influences use of genetics testing in clinical settings. Proportion of providers who responded Yes to the questions, "Have you ever ordered a genetic test for one of your patients?," "Have you ever ordered a genetic-based test for diagnostic purposes?," "Do you feel you need help interpreting clinical genomic data if these data was made available to you?," and "Would you place a consult to a physician with expertise in genomic medicine if available at your institution" arranged from left to right stratified by self-reported level of training in genomics.

targeted for that group or prevention of illnesses associated that genome."

In contrast, about 10% of respondents replied to this question with, "I don't know," or "No idea," suggesting that providers' understanding of precision medicine varies widely (**Table 2**). Most respondents included a description of tailoring medical care to an individual, typically indicating that genetics was a primary driver of precision medicine (**Table 2**). Only 5% of respondents indicated that precision medicine encompasses more than just genetics (**Table 2**), suggesting that for these providers, genomics is the most significant element of precision-medicine based healthcare.

TABLE 2 | Categorized responses to open ended question*.

Category	*N*	Percentage
Tailoring medical care to an individual	45	76.3%
Using genetics to guide medical care (diagnosis, therapy, etc.)	42	71.2%
Combination of factors that influence health (genetics, environment, and/or lifestyle)	3	5.1%
Basis for precision medicine is only genetic information	31	52.5%
Don't know/no response	6	10.2%

**Responses to the question, "What is your definition of precision medicine?" were coded into five categories. Responses could fall into more than one category.*

Level of Training in Genomics Influences Use of Genomics in Clinical Settings

Chi-square goodness of fit analysis showed a significant difference between the level of training in genomics among respondents ($\chi^2 = 30.125$, $p = 1.299e\text{-}6$), with 80% of respondents reporting limited to no training in genomics. We found a statistically significant relationship between provider level of training in genetics and whether or not providers would order a genetic test (Pearson's $\chi^2 = 10.17$, $p = 0.006$) or use a genetic test for diagnosis purposes (Pearson's $\chi^2 = 7.2$, $p = 0.027$). Providers with Moderate or Extensive training in genetics more frequently ordered genetic tests than providers with limited training in genetics (**Figure 1**). Furthermore, providers with moderate or extensive training in genomics more frequently used genetic testing for diagnostic purposes than providers with limited or no advanced training in genomics (**Figure 1**). There was no significant relationship between respondents' level of genetics training and year of completing medical training (Pearson's $\chi^2 = 8.5$, $p = 0.7483$).

In addition to utilizing genetic testing in the clinic, level of training in genomics also influenced how providers would use that data in the clinic. There was a significant relationship between level of training in genomics and desire for assistance with interpretation of genetic test results (**Figure 1**, Pearson's $\chi^2 = 8.26$, $p = 0.016$), with providers reporting less genomics training indicating that they would like help with genetic test interpretation. In contrast, no significant relationship was found between the level of training in genomics and the desire to consult an expert in genomic medicine (**Figure 1**, $p = 0.3974$), suggesting that providers generally agree that they would benefit from consulting with an expert in the field, regardless of level of training in genomics.

Barriers to Implementation of Genetic Testing in the Clinic

Providers were asked to choose the most significant barrier for them when considering the option of ordering a genetic test for any given patient. The responses could be (1) availability of genetic tests, (2) personal training or knowledge of which genetic tests to order and how to integrate results from specific genetic tests, (3) medical guidelines, (4) cost, (5) lack of available therapies that are specific to genetic profiles, or (6) lack of confidence in therapy(ies) when they are available for specific genetic profiles.

Chi-square goodness of fit analysis comparing the reported barriers to implementing genetic testing in clinic found a significant difference from a uniform distribution (**Figure 2A**, $\chi^2 = 33.476$, $p = 3.027e\text{-}6$). In our survey, 26 providers (41%) indicated that personal training on which genetic tests to order and how to interpret results from specific genetic tests was the most significant barrier to ordering a genetic test for a patient. As expected, we identified a relationship between the level of training in genetics and perceived barriers to testing (**Figure 2B**, Pearson's $\chi^2 = 6.9959$, $p = 0.03$). For individuals with no training in genomics, lack of personal training was the most common response (**Figure 2B**), with over 60% of respondents who had no training in genomics indicating that this lack of training was the major barrier to ordering genetic tests for their patients. As respondents' level of training in genomics increased from limited to moderate or extensive, lack of personal training in genomics was still a common barrier to genetic testing. However, these groups also reported that lack of tailored therapies or availability of testing were additional barriers that restricted the use of clinical genetic testing (**Figure 2B**).

DISCUSSION

Lack of provider knowledge in genetics has long been recognized as a barrier to the large-scale adoption of genetic testing into mainstream clinical care (5–8). However, despite a growing trend of increasing educational programs in genetics and genomics oriented toward non-genetics HCPs (9), recent studies suggest that many primary care physicians continue to be lacking in knowledge and confidence in clinical genetics and genomics (10, 11). In our study, participants cited that a lack of personal training was the most significant barrier to the use of genomics data in the clinic. Furthermore, we found that providers with higher levels of training were more likely to have ordered a genetic test, suggesting that additional training in genomics made them more comfortable with integration of genetics and genomics data into their practice. Together, these data support the claim that lack of training in genetics remains a significant obstacle in the expansion of genetics and genomics into clinical care. As such, training programs that teach fundamental concepts in genetic testing utilization may help non-genetics healthcare providers become more comfortable using this type of testing.

While our study found that roughly 50% of respondents have used some form of genetic testing in the clinic, 86% would like to have help interpreting these data. Furthermore, respondents in all groups wanted a genetics consulting service, even the single provider reporting "Extensive" training in genetics and genomics. Of note, the category with the most respondents who answered "No" to the question about genetics consultation was the group who indicated they had no advanced training in genetics ($n = 5$). This suggests that this group may not use genetic testing enough to warrant a consulting service, are unaware of the potential benefit of a genetics consult to help order and interpret genetic tests and genomics data, or feel that there is insufficient data to support the costs of genetic testing.

As genetics and genomics knowledge becomes commonplace, primary care practice will be heavily impacted by a massive

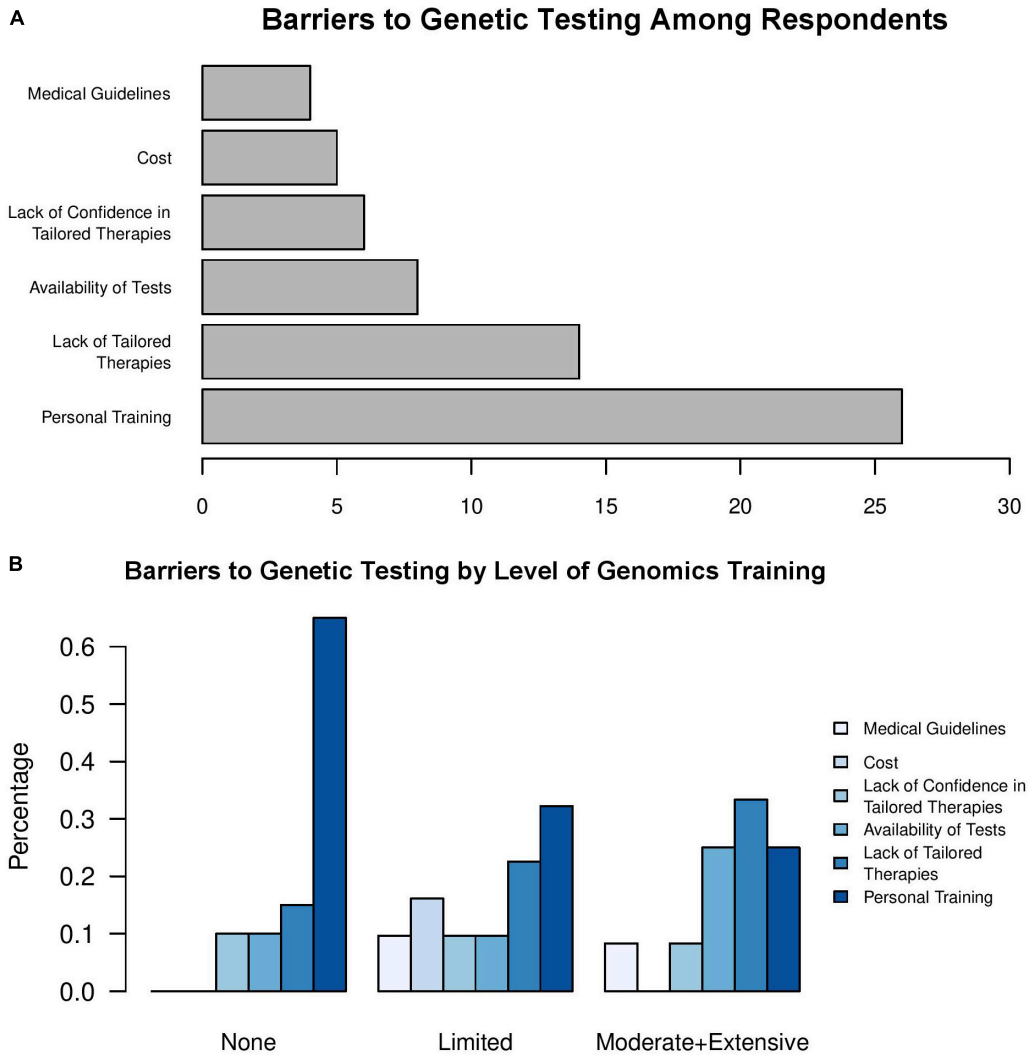

FIGURE 2 | Lack of personal training in genomics is a barrier to clinical genetic testing. Physician's responses to the question, "What is the major barrier for you to order a genetic test for a given patient?" among **(A)** all respondents and **(B)** stratified by level of training in genomics.

inflow of genetic and genomic data. This will be especially exacerbated in medically underserved populations (6) and will only grow as consumer-initiated genetic testing expands (12), and return of genetic testing results in research programs grows (13, 14). Our findings suggest that creation of referral clinics within large healthcare settings may be an accelerator to the adoption of genetics and genomics data into clinical practice. However, genetics specialists are limited. While there has been recent growth in the genetic counseling profession, with the expansion of genetic counseling training programs in the United States (15), there will likely continue to be high demand for medical geneticists until gaps in training programs are addressed (3). Novel approaches to referral clinics, such as leveraging genetic counselors' skills in providing information and interpretation of genetic testing to both patients and providers, while allowing medical geneticists to focus on complex diagnoses and management, could be a model to help address access to genetics specialty care while demand for medical geneticists continues.

Our study expands on existing literature showing that non-genetics trained healthcare providers are not comfortable implementing genetics into their clinical practice (4, 10, 11). Our study was unselected for physician specialty in order to broadly capture providers comfort and experience with genetic testing, outside of formal training and this may have led to discrepancies between individual responses. As such, future studies that examine the knowledge and comfort of using genetic testing among specialists and primary care providers would further clarify how well-equipped physicians in different areas of medicine are to incorporate genetic testing into their clinical practice. This could be further expanded by examining providers working in more diverse healthcare settings and geographic regions, where access to genetics specialists may be severely limited.

In our survey, most respondents indicated that precision medicine will define standards of care, with genetics having an increasingly prominent role in clinical practice. This calls for organized efforts by health care organizations to expand genetics and genomics education for both genetics and non-genetics providers to meet the future growth and demand for these types of services. Such expansion should include standardized interpretation resources, continuing education programs for providers and genomics consultation services.

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 the Banner Office for Human Research and The University of Arizona Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

IR, SH, and KR were responsible for the conception and design of the work. VS and KR wrote the manuscript. All authors participated in drafting and revising the manuscript critically for intellectual content, approved of the final version to be published, and agreed to be accountable for all aspects of the work by ensuring that questions related to accuracy and integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGMENTS

We would like to thank Kurt Zhang at the Texas A&M Institute of Biosciences and Technology for their review and feedback on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.757212/full#supplementary-material

REFERENCES

1. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. *Cell.* (2021) 184:1415–9. doi: 10.1016/j.cell.2021.01.015
2. Guttmacher AE, Porteous ME, McInerney JD. Educating health-xcare professionals about genetics and genomics. *Nat Rev Genet.* (2007) 8:151–7. doi: 10.1038/nrg2007
3. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M. Current conditions in medical genetics practice. *Genet Med.* (2019) 21:1874–7. doi: 10.1038/s41436-018-0417-6
4. Secretary's Advisory Committee on Genetics Health and Society. *Genetics Education and Training.* (2011). Available online at: https://osp.od.nih.gov/wp-content/uploads/2013/11/SACGHS_education_report_2011.pdf (accessed April 26, 2011).
5. Bonter K, Desjardins C, Currier N, Pun J, Ashbury FD. Personalised medicine in Canada: a survey of adoption and practice in oncology, cardiology and family medicine. *BMJ Open.* (2011) 1:e000110. doi: 10.1136/bmjopen-2011-000110
6. Chou AF, Duncan AR, Hallford G, Kelley DM, Dean LW. Barriers and strategies to integrate medical genetics and primary care in underserved populations: a scoping review. *J Community Genet.* (2021) 12:291–309. doi: 10.1007/s12687-021-00508-5
7. Najafzadeh M, Davis JC, Joshi P, Marra C. Barriers for integrating personalized medicine into clinical practice: a qualitative analysis. *Am J Med Genet A.* (2013) 161A:758–63. doi: 10.1002/ajmg.a.35811
8. Petersen KE, Prows CA, Martin LJ, Maglo KN. Personalized medicine, availability, and group disparity: an inquiry into how physicians perceive and rate the elements and barriers of personalized medicine. *Public Health Genomics.* (2014) 17:209–20. doi: 10.1159/000362359
9. Talwar D, Tseng TS, Foster M, Xu L, Chen LS. Genetics/genomics education for nongenetic health professionals: a systematic literature review. *Genet Med.* (2017) 19:725–32. doi: 10.1038/gim.2016.156
10. Carroll JC, Allanson J, Morrison S, Miller FA, Wilson BJ, Permaul JA, et al. Informing integration of genomic medicine into primary care: an assessment of current practice, attitudes, and desired resources. *Front Genet.* (2019) 10:1189. doi: 10.3389/fgene.2019.01189
11. Hauser D, Obeng AO, Fei K, Ramos MA, Horowitz CR. Views of primary care providers on testing patients for genetic risks for common chronic diseases. *Health Aff (Millwood).* (2018) 37:793–800. doi: 10.1377/hlthaff.2017.1548
12. Majumder MA, Guerrini CJ, McGuire AL. Direct-to-consumer genetic testing: value and risk. *Annu Rev Med.* (2021) 72:151–66. doi: 10.1146/annurev-med-070119-114727
13. Grzymski JJ, Elhanan G, Morales Rosado JA, Smith E, Schlauch KA, Read R, et al. Population genetic screening efficiently identifies carriers of autosomal dominant diseases. *Nat Med.* (2020) 26:1235–9. doi: 10.1038/s41591-020-0982-5
14. Denny JC, Rutter J, Goldstein DB, Philippakis A, Smoller JW, Jenkins G, et al. The "all of us" research program. *N Engl J Med.* (2019) 381:668–76. doi: 10.1056/NEJMsr1809937
15. Hoskovec JM, Bennett RL, Carey ME, DaVanzo JE, Dougherty M, Hahn SE, et al. Projecting the supply and demand for certified genetic counselors: a workforce study. *J Genet Couns.* (2018) 27:16–20. doi: 10.1007/s10897-017-0158-8

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

frontiers | Frontiers in Medicine

REVIEW
published: 08 April 2022
doi: 10.3389/fmed.2022.806696

Bioinformatic Challenges Detecting Genetic Variation in Precision Medicine Programs

Matt A. Field [1,2,3]*

[1] Centre for Tropical Bioinformatics and Molecular Biology, College of Public Health, Medical and Veterinary Science, James Cook University, Cairns, QLD, Australia, [2] Immunogenomics Lab, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia, [3] Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia

Precision medicine programs to identify clinically relevant genetic variation have been revolutionized by access to increasingly affordable high-throughput sequencing technologies. A decade of continual drops in per-base sequencing costs means it is now feasible to sequence an individual patient genome and interrogate all classes of genetic variation for <$1,000 USD. However, while advances in these technologies have greatly simplified the ability to obtain patient sequence information, the timely analysis and interpretation of variant information remains a challenge for the rollout of large-scale precision medicine programs. This review will examine the challenges and potential solutions that exist in identifying predictive genetic biomarkers and pharmacogenetic variants in a patient and discuss the larger bioinformatic challenges likely to emerge in the future. It will examine how both software and hardware development are aiming to overcome issues in short read mapping, variant detection and variant interpretation. It will discuss the current state of the art for genetic disease and the remaining challenges to overcome for complex disease. Success across all types of disease will require novel statistical models and software in order to ensure precision medicine programs realize their full potential now and into the future.

Keywords: precision medicine, variant detection, high-throughput sequencing, pathogenic variant, variant prioritization, FPGA—field-programmable gate array, GPU-accelerated

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute (NCI),
United States

Reviewed by:
Nurlan Dauletbayev,
McGill University, Canada
Raquel Cruz Guerrero,
Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Spain
Fernanda Rodrigues-Soares,
Universidade Federal do Triângulo Mineiro, Brazil

***Correspondence:**
Matt A. Field
matt.field@jcu.edu.au

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *01 November 2021*
Accepted: *21 March 2022*
Published: *08 April 2022*

Citation:
Field MA (2022) Bioinformatic Challenges Detecting Genetic Variation in Precision Medicine Programs. Front. Med. 9:806696. doi: 10.3389/fmed.2022.806696

INTRODUCTION

Precision medicine programs are increasingly being implemented worldwide with a goal of improving patient care for an individual (1). Largely enabled by access to increasingly affordable high quality sequence data, great strides have been made in the diagnosis and management of genetic disease. By considering a patients unique genetic, environmental and lifestyle factors precision medicine aims to develop customized patient-specific treatments. Increasingly important in precision medicine programs is the ability to utilize genetic information to stratify patients with regard to treatment options and outcomes. Such patient information can be broadly classified into predicative and prognostic biomarkers with prognostic biomarkers informing on patient outcome in contrast to predictive biomarkers which directly guides treatment (the focus of this review). Currently, diagnosis and treatment of cancer and rare diseases are the largest beneficiaries of precision medicine programs. In cancer, huge numbers of druggable molecular alterations have been described and cataloged in growing public repositories like Clinical Interpretation of Variants in Cancer (CIViC) (2). As of February 2022, CIViC contain an incredible 3,041 actionable variants

in 464 genes supported by 8,576 evidence items. Beyond cancer, the genetic cause of more than 80% of the roughly 6,000 known rare diseases has been elucidated in the last decade alone (3). While impressive, currently only ∼5% of these diseases have an accepted targeted treatment indicative of the work still required (3).

Despite progress in diagnosing and treating genetic diseases, a bottleneck persists in variant interpretation. The increase in sequencing capacity has identified huge numbers of new suspected pathogenic variants however there is often sparse or inconclusive supporting functional evidence. For example, cystic fibrosis (CF) is caused by up to ∼300 pathogenic variants in the *CTFR* gene however their impact is often heterogenous amongst individuals (4). Functional inference prediction tools are often run instead to access the likelihood of a mutation ablating protein function however such tools are known to have high false positive rates (5). Overall, substantial progress has been made in genetic disease however numerous challenges need to be addressed before precision medicine programs can be delivered at scale and for complex, polygenic disease.

Reliably identifying disease causing variants remains a challenge within the field particularly for complex disease. While great strides have been made for cancer and rare diseases, the diagnosis rates for complex diseases remain much lower (6). Despite these challenges, there are many examples of genetic traits in polygenic disease contributing to clinical manifestations [e.g., blood disease (7), autoimmune disease (8)]. To increase the diagnosis rates for complex diseases, previous approaches have employed a wide variety of strategies. For example, careful sample selection improve diagnosis rates by focusing on families with multiple affected individuals who exhibit extreme phenotypes and early onset of disease (9). Additionally, particular variant classes can be prioritized in different scenarios such as homozygous mutations for consanguineous pedigrees (10) and *de novo* mutations for trios with an affected child and unaffected parents (11). While these strategies are feasible in particular scenarios, in many cases only a single patient is available meaning prioritization strategies must consider all genetic variation detected in a patient.

An additional challenge in variant detection is the increased recognition of the importance of larger copy number and repeat variation in driving disease. These variant classes are harder to reliably detect than single nucleotide variants (SNVs) and small insertion/deletions (indels) particularly with short read sequencing technologies (12). Even for SNVs and small indels there are limitations with most precision medicine programs prioritizing variants disrupting gene function yet increasingly portions of the "missing heritability" in disease is being explained by small variants that either generate unexpected splicing errors or disrupt poorly annotated regulatory elements (13). These challenges are compounded within populations of non-European ancestry due the over representation of individuals of European ancestry within public variant databases. While this trend is improving, a 2016 study found 81% of all GWAS study samples were of European ancestry with only 4% of all samples being of African or Latin American ancestry or Indigenous (14).

Inherent to any successful precision medicine program is the timely and accurate detection of genetic variation and the prioritization of the variants most likely to be relevant to the patient's condition. Advances in software and hardware are playing an increasingly innovative role in delivering on these goals particularly for accurate variant detection and prioritization. Software-based approaches are varied and include developing new algorithms, increasing efficiencies of existing algorithms, increasing parallelization and improved standardization of common file formats (15). Hardware-based approaches are increasingly important and include increased availability of cluster and cloud based compute environments (16), field-programmable gate arrays (FPGA) devices (17) and graphical processing units (GPU) enabled bioinformatics algorithms (18).

Pharmacogenetic variants are also important in precision medicine with individual variability in drug response increasingly being attributed to genetic variation. An average individual is estimated to carry three clinically actionable pharmacogenetic variants with 97% of individuals carrying at least one such variant (19). Increasingly large repositories that aggregate and annotate pharmacogenetic variants [e.g., PharmGKB (20)] are being used in drug dosage decision making. While encouraging, the majority of known pharmacogenetic variants remain underutilized in precision medicine. This is largely due to a poor understanding of the underlying mechanisms and challenges in accurately identifying and annotating pharmacogenetic variants. For example, a recent study showed pharmacogenetic variants causing missense mutations and associated with off-target effects are incorrectly classified as benign by functional inference prediction software (21). Further software development is needed to account for this special class of variation (22).

Large-scale translation of research results into the clinic remain a significant bottleneck for the wide-spread implementation of precision medicine programs. While increasingly detailed annotation and prioritization workflows are being described and shared (23), most still remain siloed within individual institutions or are bound to specific hardware configurations. Improved containerization of workflows is helping to facilitate sharing of analysis pipelines (24) with initiatives like the Global Alliance for Genomics and Health (GA4GH) facilitating the timely sharing of large genetic data sets. While improving standardization and sharing of resources is critical, a larger challenge is the availability of accurate databases of clinically actionable variants. While many such repositories exist, studies have identified inaccuracies throughout (25). To illustrate, a recent study followed up 239 variants in the Human Gene Mutation Database (HGMD) classified as disease-causing and found only 7.5% of these variants met the criteria required to be called disease-causing (26). For precision medicine to succeed at scale, more accurate and detailed databases of clinically actionable variants are required.

Despite substantial progress, reliably detecting genetic variants within precision medicine programs has many challenges remaining. While solutions are actively being developed it is clear more improvements are needed if we are to realize the full potential of population-wide precision medicine

programs. In this review, I will describe the current and future challenges for identifying clinically relevant genetic variants in precision medicine programs with resources summarized in **Table 1**.

CURRENT CHALLENGES AND SOLUTIONS

A wide variety of strategies are being employed to detect clinically relevant genetic variation at scale. These approaches can be broadly classified as software-based or hardware-based (**Figure 1**).

Software Based

Software development and optimization play an important role in improving precision medicine programs by improving algorithm performance and reducing run time and memory requirements. This is occurring *via* a variety of mechanisms including the development of new algorithms, optimization of existing algorithms, increasing parallelization *via* job partitioning, and standardized file formats (**Table 2**).

New algorithms are being developed for a variety of analysis steps in variant detection workflows, particularly for variant prioritization. While the generation of either germline or somatic raw variant calls is increasingly routine [e.g., BWA for short read alignment (42) followed by GATK best practices (43)], development of algorithms to identify clinically relevant variants from raw variant calls remains an active area of software development. The increasing availability of variant annotation data has led to the development of annotation aggregator packages such as ENSEMBL Variant Effect Predictor (VEP) (31) or ANNOVAR (33). With external annotation sets and gene models rapidly updating, such tools are indispensable for applying the latest annotations to raw variant lists. Another area of active software development is predicting the functional impact of variant classes such as missense mutations. Heavily used tools such as PolyPhen2 (36) have been shown to exhibit high false positive rates (5) and newer tools are increasingly utilizing machine learning (38) and a consensus-based approach (44) to try to overcome these limitations however more work is needed to improve their accuracy. The most active area of development currently is disease-specific solutions with the increasing recognition that any disease requires tailored annotation / prioritization and may even require different types of sequence data. For example, with autoimmune disease T-cell receptor (TCR) and B-cell receptor (BCR) repertoires are often sequenced requiring custom software to identify the relevant clonotypes (45). Additionally, incorporating disease specific annotations [e.g., Immgen for autoimmune disease (46)] requires custom handling as disease-specific databases are generally not available within the annotation aggregation tools.

Ongoing development of many commonly used bioinformatics algorithms is reducing run time and memory requirements. For example, an update to the popular amplicon cluster software Swarm reduced memory usage by 50% and run time by 7X (47). These improvements are often driven by increasingly large data sets with many long-running software packages having been created when sequence data sets were smaller. Increasingly, individuals not involved in the original development of the software are finding ways to speed up and reduce memory usage of many commonly used algorithms. For example, an external group modified the popular Minimap2 (48) long read aligner by incorporating multi-index merging which reduced memory usage by an order of magnitude (49). While gains have been significant in many instances, further reductions in run time and memory usage will greatly facilitate the wide-spread uptake of precision medicine programs.

A common approach to reduce run time is increasing parallelization *via* programming models like MapReduce (50).

TABLE 1 | Resources for variant detection in precision medicine programs.

Database	Function	Web link
dbSNP (27)	Population level variation	http://www.ncbi.nlm.nih.gov/snp
gnomAD (28)	Population level variation	https://gnomad.broadinstitute.org
1000 Genomes Phase 3 (29)	Population level variation	http://phase3browser.1000genomes.org
Database of Genomic Variants (30)	Population level variation	http://dgv.tcag.ca/dgv/app/home
Variant Effect Predictor (31)	Variant annotation	https://ensembl.org/info/docs/tools/vep/index.html
dbNFSP (32)	Variant annotation	https://sites.google.com/site/jpopgen/dbNSFP
AnnoVar (33)	Variant annotation	http://annovar.openbioinformatics.org/en/latest/
ClinVar (34)	Clinical annotation	https://www.ncbi.nlm.nih.gov/clinvar
LOVD (35)	Clinical annotation	http://www.lovd.nl
PolyPhen2 (36)	Functional impact	http://genetics.bwh.harvard.edu/pph2/
SIFT (37)	Functional impact	https://sift.bii.a-star.edu.sg/
CADD (38)	Functional impact	https://cadd.gs.washington.edu/
GTEx (39)	Gene expression	https://gtexportal.org
Multi-symbol checker (40)	Gene naming	https://www.genenames.org/tools/multi-symbol-checker
OMIM (41)	Gene / disease annotation	https://www.omim.org

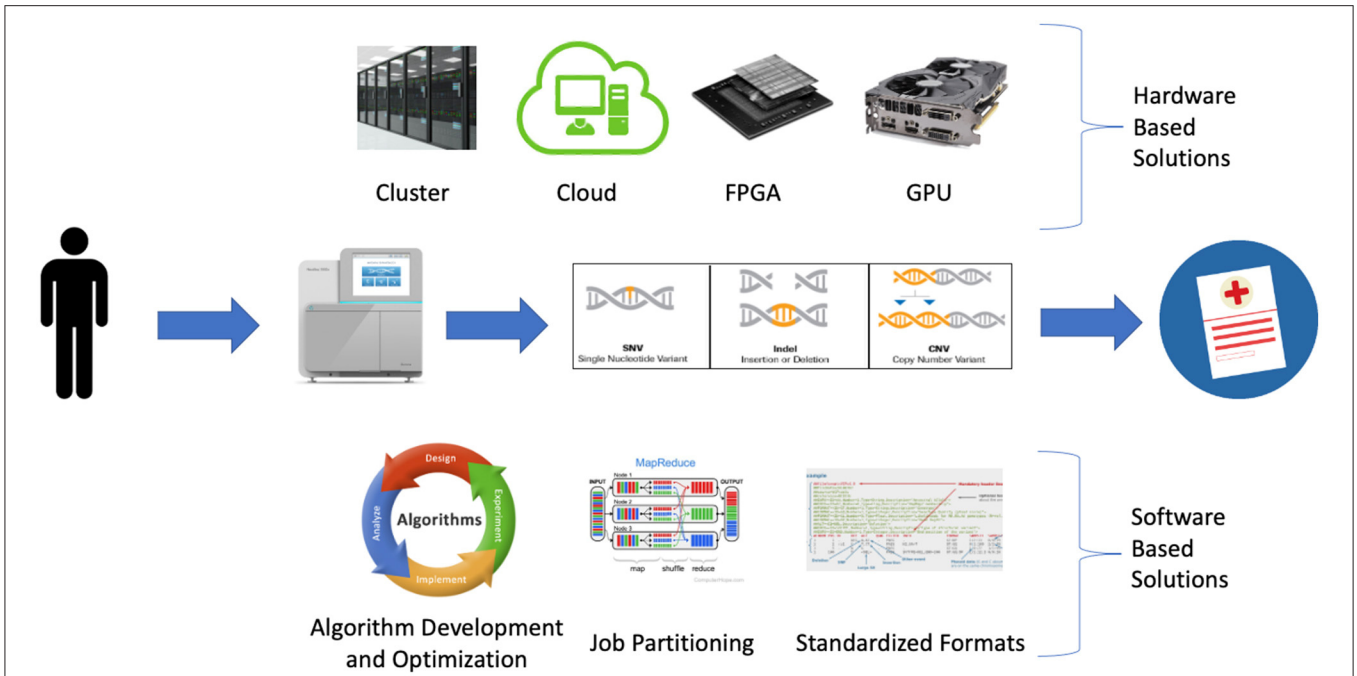

FIGURE 1 | Software and hardware-based strategies being employed to address bioinformatic bottlenecks in large scale precision medicine programs.

TABLE 2 | Software based solutions.

Strategy	Advantages	Disadvantages
Algorithm development	– Develop novel approaches	– Requires community uptake
	– Existing suite of tools available for benchmarking	– Challenging to significantly change existing workflows
Algorithm optimization	– Quicker to improve existing algorithms	– Gains are often minimal if software well-designed initially
	– Simple to benchmark versus previous releases	– Any changes in expected output requires verification
Job partitioning	– Increases parallelization and reduces serial run time	– Splitting and combining results adds software complexity
Standardized file formats	– Standardized formats allows easy algorithm benchmarking	– No flexibility for new data types or information

MapReduce is a general purpose model designed to run efficiently over large datasets on commodity compute clusters. The incorporation of MapReduce by Apache Hadoop has led to its incorporation throughout the bioinformatics landscape, now found in software such as GATK (43) and BLAST (51). In addition to using models like MapReduce, custom solutions are often employed such as partitioning long-running whole genome jobs into smaller genomic chunks, often at the level of chromosome (52). Using this approach, we can expect at least an order of magnitude reduction in run time as the largest single chromosome represents <10% of the total human reference genome size. It should be noted that while this approach is suitable for algorithms where each chromosome is analyzed independently, this approach won't work when information from multiple chromosomes is required for an analysis (e.g., genome wide stats, detecting inter-chromosomal translocations). Another issue with this approach is the increased complexity required to manage the jobs and merge the per-chromosome output files.

File formats are increasingly being standardized to improve reproducibility and data sharing. For example, virtually all short read aligners now generate SAM alignment files while most variant detection software outputs variant call format (vcf) files. Standardizing file types can reduce ongoing storage requirements *via* improved compression which allows algorithms to work entirely with compressed data such as gzipped FASTQ files or compressed SAM files (BAM/CRAM). For example, read alignment generates an extremely large SAM file containing one row per read pair. Given whole genome datasets routinely contain >1 billion read pairs SAM files quickly become large and unmanageable for manipulation. To address this, a lossless binary version of the SAM file was created that reduces the file size by up to 75%. The resultant BAM file is significantly smaller and can be effortlessly queried and manipulated *via* bioinformatic packages such as SAMTools (53). Despite the improvements with BAM files, a more compressed format called CRAM was subsequently developed, resulting in a further reduction of 40–70% in size relative to BAM (54). While promising, a limitation

of the CRAM format is that compression is not a lossless conversion whereas BAM compression is lossless. Overall file standardization has made significant improvements to variant detection workflow efficiency and portability however challenges do exist with frequent version changes often failing to maintain backwards compatibility.

Hardware Based

Hardware developments are making significant contributions to precision medicine programs *via* increasingly large and accessible compute infrastructure and hardware accelerated solutions designed to address software bottlenecks. The increase in available computational resources is primarily driven by increasingly large and accessible cluster and cloud compute resources while the hardware accelerated solutions consist largely of new FPGA devices and GPU-enabled algorithms. Collectively the increasing uptake of these hardware-based solutions is easing existing computational bottlenecks within precision medicine programs (55). However while these hardware solutions are often designed to address the same bottlenecks, they differ with regard to ease of use, cost, performance and scalability (**Table 3**).

Most high throughput genome analysis workflows were originally designed to run on commodity clusters due to their affordability, scalability and relative ease of use. From small clusters running on local infrastructure to enterprise-level systems with thousands of readily-available cores, their design follows the same model with scheduler software responsible for managing jobs and resources across a distributed system of linked computers. This setup enables efficient parallelization of jobs using commodity hardware with minimal overhead. As such systems grow with more users and resources however, increasing levels of expertise are required for seamless operation. With such expertise, clusters are able to process huge numbers of jobs in parallel making this infrastructure critical to many project requiring efficient and timely data processing.

In addition to increasingly large compute clusters, accessible and expandable cloud-based compute resources are driving an increasing number of precision medicine programs (56). In contrast to cluster based solutions, cloud solutions perform all analyses on remote systems across a network connection. In a cloud based model, storage and compute resources are commodities that can either be borrowed or rented from a provider such as Amazon Web Services or Microsoft Azure. The greatest advantage of cloud compute is its flexibility; users can access exactly the resources required for virtually any job. This flexibility enables users of any size to utilize cloud resources providing the appropriate compute environment is available. Setting up custom cloud-ready workflows requires a significant effort initially although increasingly the most common genomics workflows are being made available [e.g., nf-core (57)]. Potential issues with public cloud resources include issues handling sensitive patient information and challenges moving large genomic data sets. To address these, some groups are opting for a hybrid solution by creating private cloud infrastructure potentially getting the benefits of both cluster and cloud approaches. Regardless of the approach, it is clear cloud compute infrastructure will play an increasingly large role in precision medicine programs (16).

Beyond increasingly large and flexible compute infrastructure, hardware accelerated solutions such as GPU-enabled algorithms and FPGA devices are now being used to reduce run time in precision medicine programs (15). While GPU-enabled versions of many popular bioinformatics algorithms have existed for a long time [e.g., GPU-BLAST is 10 years old (58)], it is only recently that we are beginning to see wide-spread uptake of these algorithms. Algorithms able to utilize GPUs can significantly increase parallelization by taking advantage of the large number of specialized cores on a single graphics card. In contrast with sequential CPU processing, GPUs offer superior scalability and reduced costs per unit however the biggest challenge is creating the specialized code required to utilize GPUs. Further, portability is a challenge as any GPU code developed is vendor-specific meaning it cannot run on another vendors GPUs. In reality, most GPU-enabled bioinformatics algorithms are currently written using NVIDIA's Compute Unified Device Architecture (CUDA) with examples from variant detection workflows focused on the short read alignment step [e.g., SOAP3 (59)]. However, with the increasing availability of GPU-enabled algorithms across the whole research spectrum more options relevant to precision medicine are likely forthcoming.

In addition to the potential of GPU-enabled algorithms, an increasing number of FPGA devices are available for precision medicine variant detection (60). FPGAs are integrated circuits designed to be configured for specific software applications. FPGAs offer many advantages in that they are flexible, inherently

TABLE 3 | Hardware based solutions.

Resource	Advantages	Disadvantages
Compute cluster	– Low cost entry	– Controller is single point of failure
	– Uses commodity hardware	– Technical expertise required
Cloud compute	– Highly scalable	– Data transfer and cost
	– No local installation	– Privacy concerns for sensitive data
FPGA	– Direct hardware / software link	– Challenging to program/re-program
	– Relatively low cost	– Integration requires technical expertise
GPU	– Cheaper than CPUs	– Chipset specific coding required
	– High parallelization possible	– Higher power usage than FPGAs

parallel, re-programable and relatively low cost. The greatest limitation of FPGA is they are very difficult to program compared to GPUs (15) however devices exist for both short read alignment [Bowtie (61)] and even entire precision medicine workflows (DRAGEN). Developed by Edico Genome and now owned by Illumina, DRAGEN can reduce already parallelized variant detection workflows by up to an order of magnitude (62). DRAGEN has also been deployed at scale in partnership with Genomics England for their rare disease analysis platform. It is clear FPGA devices have a significant role to play in precision medicine.

While all the software and hardware solutions are described in isolation, in reality various hardware and software combinations are being tested in new precision medicine workflows.

Variant Detection

Detecting small genetic variants within sequenced human genomes is a relatively mature high-through sequencing application. Despite this progress, challenges remain to comprehensively characterize all variation. Variant detection challenges include an incomplete human reference genome, a limited number of robust validated variant truth sets and no clear best performing algorithm; challenges which are amplified for less well characterized variant classes such as repeat and copy number variation which are increasingly being implicated in human disease (63).

Since the initial human genome assembly in 2001, improvements in both software and long read sequencing technology have improved the genome assembly to the point where we now have the first telomere-to-telomere genome assembly for most chromosomes (64). While promising for the future, most precision medicine programs currently utilize the GRCh38 assembly and will likely continue to do so for the near future largely due to the abundance of well characterized annotations data reported relative to these genomic coordinates. For example, one of the most important annotation sets GNOMAD (28) only converted to GRCh38 in October 2019, almost five full years after the initial GRCh38 assembly was released in December 2013. A similar period of time will likely be required to convert existing workflows and annotations to the improved telomere-to-telomere assembly following wide-spread acceptance within the community. For context, the GRCh38 assembly still contains 850 sequence gaps with numerous mis-assembled regions reported over the years.

Improving variant detection workflows requires robust validated variant truth sets for benchmarking both new algorithms and updated versions of existing algorithms. Until quite recently a single reference dataset (NA12878) was available for benchmarking which was limited by ∼30% of the reported variants being classified as low confidence due to either low coverage, local alignment problems, or systematic sequencing errors (65). The wide-spread availability of high quality long read sequence data and the increased number of samples available within consortiums like Genome in a Bottle mean an increasing number of relatively complete high quality variant truth sets are available for benchmarking.

While the algorithms for detecting SNVs and small indels are increasingly accurate and reliable, the algorithms for detecting other types of variation such as repeat, copy number and structural variation remain an active area of development. To illustrate, a recent review reported SNV and small indel F-scores of >0.975 and >0.85, respectively, (12) while a review of copy number and structural variant detection algorithms reported precision values of between 0.40 and 0.91 and recall values from 0.07 to 0.28 depending on the type of variant being detected (66). Limited data is available reporting the true accuracy of repeat variation detection algorithms due to lack of a gold standard reference validation set with most tools instead relying on analyses using *in silico* data. It should be noted that despite the highly precision and recall reporting for SNV calling, studies have shown that recurrent false positive variants are routinely called and exist within variant repositories (67).

Central to any analysis step is the selection of the algorithm(s) to run. While for many analysis steps a single algorithm is determined to perform sufficiently, for many variant detection applications leading algorithms generate highly discordant results with no single algorithm performing optimally under all conditions (52). To address this, an increasingly popular approach is to run multiple algorithms and apply a consensus approach in order to minimize the effect of any potential biases within a single algorithm [e.g., DNA (52)/RNA (68)]. This approach has been shown to generate the highest quality variant data sets for either specificity or sensitivity depending on whether the intersection or the union of the variant calls is taken, respectively.

Variant Interpretation

Whole genome sequencing (WGS) generates millions of raw variant calls, the large majority of which are not relevant to disease. While targeted sequencing experiments such as exome or gene panel sequencing reduce the number of raw variant calls, the challenge of variant filtering and interpretation to identify clinically relevant variants remains. Beginning with raw variant calls, the most common filtering strategy is to apply a series of successive annotation and prioritization steps in order to reduce the genomic search space for clinically relevant variants. Such strategies include stratifying variants by impact on genes, running functional inference prediction software for missense mutations, overlapping to both ethnically matched population-level and disease-specific variant repositories, and sequencing pedigrees for germline disease and paired tumor/normal samples for cancer (**Table 4**). Overall, each step reduces the genomic search space with an overarching goal of reducing the final list of candidate variants down to a size suitable for in-depth manual interrogation.

Often the first annotation step is to stratify variants based on their impact on genes. For example, SNVs causing non-synonomous/nonsense mutations or small indels situated within exons causing a frameshift are prioritized. Determining this impact can be challenging however due to factors such as differences in gene models or multiple isoforms reported within a single gene model. For example, a recent study aligned RNA-Seq data to three popular gene models (ENSEMBL, RefSeq,

TABLE 4 | Strategies for variant prioritization.

Strategy	Strengths	Limitations
Consensus-approach running multiple algorithms	– Minimize algorithm biases – Reduce specificity or sensitivity by taking intersection or union	– Adds computational complexity – Longer run time
Stratify by impact on genes	– Prioritize disease enriched variant sets (e.g., missense or splice-site variants)	– Changes reported relevant to specific version of gene model – Multiple isoforms often available
Functional inference prediction software	– Prioritize mutations likely to disrupt protein	– Tools have known high false positive rates
Overlap population-level variant databases	– Allows filtering of common population-level variation	– Contains errors and incomplete records due to lack of curation
Overlap disease-specific databases	– Identify variants or genes previously implicated in disease	– Large numbers of non-causal variants often included
Pedigree sequencing	– Generate pedigree-wide annotation (disease inheritance, compound heterozygosity, etc)	– Obtaining samples for larger family
Paired cancer sequencing	– Matched tumor/normal samples can detect somatic variation	– Sample purity – Tumor heterogeneity

and UCSC) and found 95% of non-junction read alignments were identical across the three gene sets however only 53% of junction spanning read alignments were identical (69). Such studies illustrate the importance of careful gene model selection. Even within a single gene model multiple isoforms are often reported, meaning the choice of isoform can alter the expected impact on the gene. Many workflows opt to compare the impact across all isoforms and report the most severe outcome while others report the impact relative to the annotated "canonical" transcript as reported by gene models such as ENSEMBL, RefSeq, and UCSC.

Another challenge in variant interpretation is the identification of missense mutations most likely to disrupt protein function. With hundreds or thousands of missense mutation calls per patient, a large number of computational tools have been developed to prioritize these variants. Such tools are generally trained on validated disease mutations as a positive set and common polymorphisms as a negative set and consider three main types of evidence; sequence conservation, protein structure, and protein annotations. These tools however are untested against the full spectrum of random *de novo* mutations and validation studies have reported consistently high false positive rates for both candidate disease-causing (5) and pharmacogenetic variants (21). Increasing gains in performance are reported by tools that apply a consensus approach by incorporating scores from other algorithms into their own scoring (e.g., CADD (38). Additional gains have recently been reported in algorithms applying machine learning approaches trained on increasing large data sets (70) however wide spread validation studies are required to validate these claims.

Databases of population-level variation are extremely valuable for reducing the search space *via* the removal of common variants as candidates. Databases like dbSNP (27) and GNOMAD (71) contain increasingly detailed population-level variant frequency information which allows both the de-prioritization of common variants as well as the prioritization of rare or *de novo* variants. It is critical when applying such filters to use ethnically matched allele frequencies using the increasingly granular variant information available within the variant repositories. Without ethnic matching, many variants are incorrectly characterized as novel or rare due to under-sampling in the repository of the patient's ethnic group. Despite efforts in recent year to increase numbers of under-represented ethnicities in such databases, much work is needed to include all groups such as Indigenous populations (72).

Equally important to population-level databases are human disease databases which allow previously implicated variants and/or genes to be prioritized. Databases of clinically relevant variants are numerous and growing rapidly in size (e.g., ClinVar (34) for germline and CIVIC (2) for cancer). Importantly, these databases follow standardized Human Genome Variation Society (HGVS) approved nomenclature for DNA and RNA variants allowing direct comparison across disparate data sets. In addition to comprehensive generic disease databases, increasingly disease-specific databases are being developed such as Infevers (73) for auto-inflammatory disorders or IARC TP53 (74) for TP53 specific mutations. While disease databases are an extremely valuable resource, most have been shown to contain high numbers of false positive due to manual curations being made with incomplete functional data. For example, one study found 27% of reported recessive disease-causing variants were false positives and were actually either common polymorphisms or mis-annotated (25). Such studies highlight the need to improve such databases *via* increasingly rigorous functional validation studies.

A powerful approach for reducing the search space for disease-causing variants in rare disease is the sequencing of families or pedigrees. Using this approach there are two main applications; sequencing trios with an affected child and two unaffected parents or sequencing multiple members of larger pedigrees containing multiple affected members. In both instances custom software is required to identify the variants most likely to be causal; namely *de novo* mutations in the trios and variants shared between affected and missing in unaffected members in the larger pedigrees. With pedigrees, specialized software is required to concurrently consider all

variation and provide pedigree-specific annotation such as disease inheritance patterns, phasing information, and potential compound heterozygosity (75). While such tools are increasingly mature, more is needed to incorporate their results into precision medicine workflows.

For detecting somatic mutations in cancer, the most effective strategy is sequencing paired tumor and normal samples and analyzing them simultaneously with cancer-specific software to identify candidate driver mutations (76). The presence of a matched control sample facilitates the identification of somatic variants however issues such as sample cross-contamination and tumor heterogeneity ensure cancer-specific software is required for reliable somatic variant detection. In this space, single cell sequencing has the potential to mitigate some of the issues around sample heterogeneity (77).

While currently most precision medicine programs run some combination of the above annotation steps in series, increasingly machine-learning based approaches are being developed to identify clinically relevant variants directly from raw variant lists (78). While much work is required to achieve this lofty goal, machine-learning based approaches are already being used successfully for more specific applications within the larger workflows such as detecting variant pairs causing disease (79), prioritizing non-coding variants (80) and identifying new pharmacogenetic variants (22). While these applications show promise, to date there are limited examples of large machine learning approaches being utilized at scale in precision medicine programs (81). In fact, a recent review could identify only a few examples of machine learning methods impacting clinical practice; an observation they largely attributing to the poor performance of the predictive models, difficulties interpreting complex model predictions and lack of validation in clinical trials sufficiently demonstrating improvements to current standard of care (82).

DISCUSSION

Precision medicine programs continue to mature and expand around the world (1). One of the most common application in such programs is detecting genetic variation relevant to a patient's condition. Significant improvements in both software and hardware over the last few years have made the detection of small genetic variation from patient sequence data an increasingly routine process. To improve the success of existing programs, work is required both with regard to detecting large and repetitive genetic variation routinely and with improving the automation of variant prioritization. In the near-future, it will also be critical to synthesize patient clinical data with a variety of sequence data types.

Repeat variation is broadly classified as mobile elements and tandem repeats which are further divided by size in short tandem repeats and satellites. Due to challenges detecting repeat variation using short read sequencing their frequency is largely unknown but current estimates are ∼10,000 tandem repeats and ∼2,000 mobile elements per human genome (83). Repeat variants are important as they are increasingly being implicated in driving human disease, particularly neuropathological disorders like autism (84). Similarly larger structural and copy number variation (generally defined as deletions, insertions, duplications, inversions and translocations >50bp) are increasingly being cataloged and implicated in driving disease, particularly in cancer (85). Despite the importance of these variant classes to human disease, they are largely not being interrogated in current precision medicine programs due to challenges detecting them. To address this, substantial work is needed in several areas including improved detection algorithms, better validation truth sets and repositories of both population-level and clinically-relevant variation. Long read sequencing will play a critical role in generating these improved repositories and truth sets.

While variant interpretation and prioritization workflows continue to improve, greater automation of the process is required to alleviate this current bottleneck. While annotation aggregators like VEP are continually incorporating additional external data sets, custom workflows are typically still required to collate and rank variants most likely to be clinically relevant. The desired output of such a workflow is a small list of candidate variants suitable for manual interrogation which will undergo an in-depth investigation for potential inclusion in the final clinical report. This manual process is extremely time-consuming however and requires further automation. While challenging to automate, software is urgently needed which inputs a raw vcf file and the relevant clinical information and outputs a small lists of likely causal variants suitably annotated for a clinical report. While an increasing number of groups are tackling these problems, more work is needed.

While currently most programs focus on detecting genetic variants using short-read DNA-based sequencing (e.g., targeted gene panels, exomes or WGS) increasingly other patient sequence data is being generated including transcriptome, long read, microbiome and single cell sequencing. For example, sequencing the transcriptome from a patient can be used to identify transcriptional changes likely caused by genetic mutations. A recent study used this strategy to improve diagnosis rates by 35% over genome sequencing alone by identifying deep intronic variants which altered splicing (13). Long read sequencing is increasingly being employed to detect complex variation unable to be easily detected with short read technologies (86). If the cost and quality of long read sequencing continues to improve it is feasible that long reads can be used routinely in precision medicine programs in the future. Microbiome is likely to be important in future programs as well. Dysbiosis of the microbiome is increasingly linked to human disease and the ability to examine differential abundance of metagenomic data (87) before and after treatment represents a new avenue for exploration (88). Finally, single cell sequencing technologies will have an increasingly large role to play given their ability to detect disease causing variants at single cell resolution over time (77). While such possibilities are exciting, it is clear current workflows are unable to work with complex multi-omics patient data sets and that substantial developments in software and hardware are required to support this in the future.

FUTURE CHALLENGES

The ongoing success of precision medicine programs for genetic disease has led to increasingly large and diverse sequence information being generated per patient. Programs are expanding in terms of number of patients sequenced, the sequencing technology employed and the type of diseases being examined. Scaling up and standardizing existing programs to population level numbers requires significant improvements in the throughput and interoperability of the systems. The other significant challenge will be the incorporation of information from additional sequencing applications including transcriptome, long read, microbiome, and single cell sequencing. The next generation of supporting software and hardware needs to be flexible and robust to manage the coming deluge of data.

CONCLUSION

Identifying clinically relevant genetic variation is one of the hallmarks of successful precision medicine programs. This review discusses the wide variety of strategies being employed to both speed up and improve the detection of clinically relevant variants. While challenging today, increasingly complex patient data sets will be generated in the near future which will require sophisticated hardware and software solutions. To support this, substantial new methodologies able to synthesize large volumes of disparate data types will be needed. These new tools will allow precision medicine programs to realize their full potential both now and into the future.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and conceived and wrote the manuscript. The author approved it for publication.

FUNDING

MF is funded by NHMRC APP5121190.

ACKNOWLEDGMENTS

I thank the National Computational Infrastructure (Australia) for continued access to significant computation resources and technical expertise.

REFERENCES

1. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. *Cell.* (2021) 184:1415–9. doi: 10.1016/j.cell.2021.01.015
2. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet.* (2017) 49:170–4. doi: 10.1038/ng.3774
3. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet.* (2020) 28:165–73. doi: 10.1038/s41431-019-0508-0
4. Pereira SV, Ribeiro JD, Ribeiro AF, Bertuzzo CS, Marson FAL. Novel, rare and common pathogenic variants in the CFTR gene screened by high-throughput sequencing technology and predicted by in silico tools. *Sci Rep.* (2019) 9:6234. doi: 10.1038/s41598-019-42404-6
5. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, et al. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci USA.* (2015) 112:E5189–98. doi: 10.1073/pnas.1511585112
6. Field MA. Detecting pathogenic variants in autoimmune diseases using high-throughput sequencing. *Immunol Cell Biol.* (2020) 99:146–56. doi: 10.1111/imcb.12372
7. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The polygenic and monogenic basis of blood traits and diseases. *Cell.* (2020) 182:1214–31 e11. doi: 10.1016/j.cell.2020.08.008
8. Jiang SH, Athanasopoulos V, Ellyard JI, Chuah A, Cappello J, Cook A, et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. *Nat Commun.* (2019) 10:2201. doi: 10.1038/s41467-019-10242-9
9. Johar AS, Anaya JM, Andrews D, Patel HR, Field M, Goodnow C, et al. Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. *Autoimmunity Rev.* (2014) 14:204–9. doi: 10.1016/j.autrev.2014.10.021
10. Al Sukaiti N, AbdelRahman K, AlShekaili J, Al Oraimi S, Al Sinani A, Al Rahbi N, et al. Agammaglobulinaemia despite terminal B-cell differentiation in a patient with a novel LRBA mutation. *Clin Transl Immunol.* (2017) 6:e144. doi: 10.1038/cti.2017.20
11. Dunkerton S, Field M, Cho V, Bertram E, Whittle B, Groves A, et al. A de novo mutation in KMT2A (MLL) in monozygotic twins with Wiedemann-Steiner syndrome. *Am J Med Genet A.* (2015) 167A:2182–7. doi: 10.1002/ajmg.a.37130
12. Chen J, Li X, Zhong H, Meng Y, Du H. Systematic comparison of germline variant calling pipelines cross multiple next-generation sequencers. *Sci Rep.* (2019) 9:9345. doi: 10.1038/s41598-019-45835-3
13. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med.* (2017) 9:eaal5209. doi: 10.1126/scitranslmed.aal5209
14. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* (2016) 538:161–4. doi: 10.1038/538161a
15. Lightbody G, Haberland V, Browne F, Taggart L, Zheng H, Parkes E, et al. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform.* (2019) 20:1795–811. doi: 10.1093/bib/bby051
16. Langmead B, Nellore A. Cloud computing for genomic data analysis and collaboration. *Nat Rev Genet.* (2018) 19:208–19. doi: 10.1038/nrg.2017.113
17. Sanaullah A, Yang C, Alexeev Y, Yoshii K, Herbordt MC. Real-time data analysis for medical diagnosis using FPGA-accelerated neural networks. *BMC Bioinform.* (2018) 19:490. doi: 10.1186/s12859-018-2505-7
18. Nobile MS, Cazzaniga P, Tangherloni A, Besozzi D. Graphics processing units in bioinformatics, computational biology and systems biology. *Brief Bioinform.* (2017) 18:870–85. doi: 10.1093/bib/bbw058
19. Wright GEB, Carleton B, Hayden MR, Ross CJD. The global spectrum of protein-coding pharmacogenomic diversity. *Pharmacogenomics J.* (2018) 18:187–95. doi: 10.1038/tpj.2016.77
20. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther.* (2012) 92:414–7. doi: 10.1038/clpt.2012.96
21. McConnell H, Andrews TD, Field MA. Efficacy of computational predictions of the functional effect of idiosyncratic pharmacogenetic variants. *PeerJ.* (2021) 9:e11774. doi: 10.7717/peerj.11774
22. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* (2018). doi: 10.1038/s41397-018-0044-2

23. Hamzeh AR, Andrews TD, Field MA. Detecting causal variants in mendelian disorders using whole-genome sequencing. *Meth Mol Biol.* (2021) 2243:1–25. doi: 10.1007/978-1-0716-1103-6_1
24. Wratten L, Wilm A, Goke J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat Meth.* (2021) 18:1161–8. doi: 10.1038/s41592-021-01254-9
25. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med.* (2011) 3:65ra4. doi: 10.1126/scitranslmed.3001756
26. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet.* (2013) 93:631–40. doi: 10.1016/j.ajhg.2013.08.006
27. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* (2001) 29:308–11. doi: 10.1093/nar/29.1.308
28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* (2020) 581:434–43. doi: 10.1038/s41586-020-2308-7
29. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* (2015) 526:68–74. doi: 10.1038/nature15393
30. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* (2014) 42:D986–92. doi: 10.1093/nar/gkt958
31. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics.* (2010) 26:2069–70. doi: 10.1093/bioinformatics/btq330
32. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat.* (2011) 32:894–9. doi: 10.1002/humu.21517
33. Wang K, Li M, Hakonarson H, ANNOVAR. functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010) 38:e164. doi: 10.1093/nar/gkq603
34. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* (2014) 42:D980–5. doi: 10.1093/nar/gkt1113
35. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF. den Dunnen JT. LOVD v20: the next generation in gene variant databases. *Human Mutation.* (2011) 32:557–63. doi: 10.1002/humu.21438
36. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Meth.* (2010) 7:248–9. doi: 10.1038/nmeth0410-248
37. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC, et al. web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* (2012) 40:W452–7. doi: 10.1093/nar/gks539
38. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J, et al. general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* (2014) 46:310–5. doi: 10.1038/ng.2892
39. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* (2013) 45:580–5. doi: 10.1038/ng.2653
40. Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* (2019) 47:D786–D92. doi: 10.1093/nar/gky930
41. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* (2005) 33:D514–7. doi: 10.1093/nar/gki033
42. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
43. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010) 20:1297–303. doi: 10.1101/gr.107524.110
44. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. Revel: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet.* (2016) 99:877–85. doi: 10.1016/j.ajhg.2016.08.016
45. McGuire HM, Watkins TS, Field M, Taylor S, Yasuyama N, Farmer A, et al. TCR deep sequencing of transgenic RAG-1-deficient mice reveals endogenous TCR recombination: a cause for caution. *Immunol Cell Biol.* (2018). doi: 10.1111/imcb.12033
46. Shay T, Kang J. Immunological Genome Project and systems immunology. *Trends Immunol.* (2013) 34:602–9. doi: 10.1016/j.it.2013.03.004
47. Mahe F, Czech L, Stamatakis A, Quince C, de Vargas C, Dunthorn M, et al. Swarm v3: towards tera-scale amplicon clustering. *Bioinformatics.* (2021). doi: 10.1093/bioinformatics/btab493
48. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191
49. Gamaarachchi H, Parameswaran S, Smith MA. Featherweight long read alignment using partitioned reference indexes. *Sci Rep.* (2019) 9:4318. doi: 10.1038/s41598-019-40739-8
50. Dean J GS. MapReduce: simplified data processing on large clusters. *Commun ACM.* (2008) 51:107–13. doi: 10.1145/1327452.1327492
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2
52. Field MA, Cho V, Andrews TD, Goodnow CC. Reliably detecting clinically important variants requires both combined variant calls and optimized filtering strategies. *PLoS ONE.* (2015) 10:e0143199. doi: 10.1371/journal.pone.0143199
53. Li H, A. statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* (2011) 27:2987–93. doi: 10.1093/bioinformatics/btr509
54. Hsi-Yang Fritz M, Leinonen R, Cochrane G, Birney E. Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Res.* (2011) 21:734–40. doi: 10.1101/gr.114819.110
55. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics.* (2011) 27:1741–8. doi: 10.1093/bioinformatics/btr295
56. Vogt H, Green S, Broderson J. Precision medicine in the clouds. *Nat Biotechnol.* (2018) 36:678–80. doi: 10.1038/nbt.4210
57. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol.* (2020) 38:276–8. doi: 10.1038/s41587-020-0439-x
58. Vouzis PD, Sahinidis NV, GPU-BLAST. using graphics processors to accelerate protein sequence alignment. *Bioinformatics.* (2011) 27:182–8. doi: 10.1093/bioinformatics/btq644
59. Liu CM, Wong T, Wu E, Luo R, Yiu SM Li Y, et al. SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics.* (2012) 28:878–9. doi: 10.1093/bioinformatics/bts061
60. Robinson T, Harkin J, Shukla P. Hardware acceleration of genomics data analysis: challenges and opportunities. *Bioinformatics.* (2021). doi: 10.1093/bioinformatics/btab017
61. Fernandez EB, Villarreal J, Lonardi S, Najjar WA, FHAST. FPGA-based acceleration of bowtie in hardware. *IEEE/ACM Trans Comput Biol Bioinform.* (2015) 12:973–81. doi: 10.1109/TCBB.2015.2405333
62. Miller NA, Farrow EG, Gibson M, Willig LK, Twist G, Yoo B, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* (2015) 7:100. doi: 10.1186/s13073-015-0221-8
63. Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell.* (2019) 177:70–84. doi: 10.1016/j.cell.2019.02.032
64. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature.* (2020) 585:79–84. doi: 10.1038/s41586-020-2547-7
65. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, et al. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol.* (2014) 32:246–51. doi: 10.1038/nbt.2835
66. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection

algorithms for whole genome sequencing. *Genome Biol.* (2019) 20:117. doi: 10.1186/s13059-019-1720-5
67. Field MA, Burgio G, Chuah A, Al Shekaili J, Hassan B, Al Sukaiti N, et al. Recurrent miscalling of missense variation from short-read genome sequence data. *BMC Genom.* (2019) 20:546. doi: 10.1186/s12864-019-5863-2
68. Waardenberg AJ, Field MA. consensusDE: an R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. *PeerJ.* (2019) 7:e8206. doi: 10.7717/peerj.8206
69. Zhao S, Zhang B, A. comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genom.* (2015) 16:97. doi: 10.1186/s12864-015-1308-8
70. Qi H, Zhang H, Zhao Y, Chen C, Long JJ, Chung WK, et al. MVP predicts the pathogenicity of missense variants by deep learning. *Nat Commun.* (2021) 12:510. doi: 10.1038/s41467-020-20847-0
71. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. (2019) 531210.
72. Caron NR, Chongo M, Hudson M, Arbour L, Wasserman WW, Robertson S, et al. Indigenous genomic databases: pragmatic considerations and cultural contexts. *Front Public Health.* (2020) 8:111. doi: 10.3389/fpubh.2020.00111
73. Sarrauste de. Menthiere C, Terriere S, Pugnere D, Ruiz M, Demaille J, Touitou I. *Infevers: the registry for FMF and hereditary inflammatory disorders mutations. Nucleic Acids Res.* (2003) 31:282–5. doi: 10.1093/nar/gkg031
74. Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat.* (2002) 19:607–14. doi: 10.1002/humu.10081
75. Field MA, Cho V, Cook MC, Enders A, Vinuesa C, Whittle B, et al. Reducing the search space for causal genetic variants with VASP: variant analysis of sequenced pedigrees. *Bioinformatics.* (2015). doi: 10.1093/bioinformatics/btv135
76. Wilmott JS, Field MA, Johansson PA, Kakavand H, Shang P, De Paoli-Iseppi R, et al. Tumour procurement, DNA extraction, coverage analysis and optimisation of mutation-detection algorithms for human melanoma genomes. *Pathology.* (2015). doi: 10.1097/PAT.0000000000000324
77. Singh M, Jackson KJL, Wang JJ, Schofield P, Field MA, Koppstein D, et al. Lymphoma driver mutations in the pathogenic evolution of an iconic human autoantibody. *Cell.* (2020) 180:878–94.e19. doi: 10.1016/j.cell.2020.01.029
78. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* (2019) 19:281. doi: 10.1186/s12911-019-1004-8
79. Papadimitriou S, Gazzo A, Versbraegen N, Nachtegael C, Aerts J, Moreau Y, et al. Predicting disease-causing variant combinations. *Proc Natl Acad Sci USA.* (2019) 116:11878–87. doi: 10.1073/pnas.1815601116
80. Smedley D, Schubach M, Jacobsen JOB, Kohler S, Zemojtel T, Spielmann M, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *Am J Hum Genet.* (2016) 99:595–606. doi: 10.1016/j.ajhg.2016.07.005
81. Plant D, Barton A. Machine learning in precision medicine: lessons to learn. *Nat Rev Rheumatol.* (2021) 17:5–6. doi: 10.1038/s41584-020-00538-2
82. Frohlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. *BMC Med.* (2018) 16:150. doi: 10.1186/s12916-018-1122-7
83. Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. *Science*. (2015) 349:aab3761. doi: 10.1126/science.aab3761
84. Trost B, Engchuan W, Nguyen CM, Thiruvahindrapuram B, Dolzhenko E, Backstrom I, et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature.* (2020) 586:80–6. doi: 10.1038/s41586-020-2579-z
85. Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, et al. Whole-genome landscapes of major melanoma subtypes. *Nature.* (2017) 545:175–80. doi: 10.1038/nature22071
86. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* (2018) 20:159–63. doi: 10.1038/gim.2017.86
87. Thang MWC, Chua XY, Price G, Gorse D, Field MA. MetaDEGalaxy: Galaxy workflow for differential abundance analysis of 16s metagenomic data. *F1000Res*. (2019) 8:726. doi: 10.12688/f1000research.18866.2
88. Petrosino JF. The microbiome in precision medicine: the way forward. *Genome Med.* (2018) 10:12. doi: 10.1186/s13073-018-0525-6

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

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fmed.2022.808378

A Somatic Mutation Signature Predicts the Best Overall Response to Anti-programmed Cell Death Protein-1 Treatment in Epidermal Growth Factor Receptor/Anaplastic Lymphoma Kinase-Negative Non-squamous Non-small Cell Lung Cancer

Jie Peng[1]†, Lushan Xiao[2†], Dan Zou[1†] and Lijie Han[3]*

[1] Department of Medical Oncology, The Second Affiliated Hospital, Guizhou Medical University, Kaili City, China, [2] Hepatology Unit, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, China, [3] Department of Hematology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:
Alice Chen,
Division of Cancer Treatment and Diagnosis, National Cancer Institute, United States

Reviewed by:
Deepika Sirohi,
The University of Utah, United States
Raquel Cruz Guerrero,
Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Spain
E. Srinivasan,
Saveetha University, India

***Correspondence:**
Jie Peng
sank44@sina.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: 03 November 2021
Accepted: 25 March 2022
Published: 03 May 2022

Citation:
Peng J, Xiao L, Zou D and Han L (2022) A Somatic Mutation Signature Predicts the Best Overall Response to Anti-programmed Cell Death Protein-1 Treatment in Epidermal Growth Factor Receptor/Anaplastic Lymphoma Kinase-Negative Non-squamous Non-small Cell Lung Cancer. Front. Med. 9:808378. doi: 10.3389/fmed.2022.808378

Background: We aimed to exploit a somatic mutation signature (SMS) to predict the best overall response to anti-programmed cell death protein-1 (PD-1) therapy in non-small cell lung cancer (NSCLC).

Methods: Tumor samples of 248 patients with epidermal growth factor receptor (EGFR)/anaplastic lymphoma kinase (ALK)-negative non-squamous NSCLC treated with anti-PD-1 were molecularly tested by targeted next-generation sequencing or whole exome sequencing. On the basis of machine learning, we developed and validated a predictive model named SMS using the training (n = 83) and validation (n = 165) cohorts.

Results: The SMS model comprising a panel of 15 genes (*TP53, PTPRD, SMARCA4, FAT1, MGA, NOTCH1, NTRK3, INPP4B, KMT2A, PAK1, ATRX, BCOR, KDM5C, DDR2,* and *ARID1B*) was built to predict best overall response in the training cohort. The areas under the curves of the training and validation cohorts were higher than those of tumor mutational burden and PD-L1 expression. Patients with SMS-high in the training and validation cohorts had poorer progression-free survival [hazard ratio (HR) = 6.01, $P < 0.001$; $HR = 3.89$, $P < 0.001$] and overall survival ($HR = 7.60$, $P < 0.001$; $HR = 2.82$, $P < 0.001$) than patients with SMS-low. SMS was an independent factor in multivariate analyses of progression-free survival and overall survival ($HR = 4.32$, $P < 0.001$; $HR = 3.07$, $P < 0.001$, respectively).

Conclusion: This study revealed the predictive value of SMS for immunotherapy best overall response and prognosis in EGFR/ALK-negative non-squamous NSCLC as a potential biomarker in anti-PD-1 therapy.

Keywords: non-small cell lung cancer, anti-PD-1, best overall response, somatic mutations, EGFR/ALK negative

INTRODUCTION

Although immune checkpoint inhibitors (ICIs) have shown considerable success in patients with non-small cell lung cancer (NSCLC) in recent years, unsatisfactory response rates are still a limitation (1, 2). In previous studies, several predictive biomarkers for successful immunotherapy including tumor mutational burden (TMB) (3, 4), gene expression profile (5), PD-L1 expression (6), tumor-infiltrating lymphocytes (7), and high microsatellite instability (8) have been reported. However, there were some prevalent limitations to these biomarkers. First, the different cut-off values for TMB or PD-L1 expression remain controversial. Second, tumor heterogeneity and unsatisfactory predictive accuracy have restricted the real-world clinical practice of current signatures, suggesting value in finding more useful and precise biomarkers.

An increasing number of studies have revealed that specific somatic variants are significantly associated with the immunotherapy response or resistance. For example, epidermal growth factor receptor (EGFR) mutation or *MDM2* amplification have been reported to be related to hyper-progressive disease (9). Furthermore, *STK11* and *B2M* are negatively associated with programmed cell death protein-1 (PD-1)/PD-L1 inhibitor resistance, resulting in poor responsiveness (10). Additionally, *TP53*, *KRAS*, and *POLE* mutations can predict the PD-1/PD-L1 blockade response in advanced NSCLC (11). Moreover, different commutations, including KL (*KRAS* and *STK11*) and KP (*KRAS* and *TP53*) molecular subtypes, showed diverse responses to ICIs in NSCLC (12, 13). Finally, multiple mutations in DDR or Notch1/2/3 pathways indicate favorable clinical prognosis and response in NSCLC patients receiving ICI therapy (14, 15). Thus, we speculated that a panel of somatic mutations could be exploited to build a robust predictive model for identifying patients who would acquire a favorable or poor response to immunotherapy in advanced NSCLC.

Considering the data mining of next-generation sequencing (NGS) and whole exome sequencing (WES) in EGFR/anaplastic lymphoma kinase (ALK) non-squamous NSCLC patients treated with ICIs, we used a routine machine learning model based on somatic mutation profiling to develop a genomic signature for predicting the best overall response (BOR) and the prognosis of immunotherapy. Such a classification pattern of genomic panels could serve as a useful and robust tool for selecting patients who would benefit from ICIs.

MATERIALS AND METHODS

Immunotherapeutic Patients

Our databases were derived from three cohorts of patients with advanced NSCLC receiving ICI therapy. From the first cohort, 75 patients with stage IV NSCLC were treated with a combination of nivolumab and ipilimumab in the clinical trial CheckMate-012 (NCT01454102) between February 2013 and March 2015 (16). Sixteen patients with squamous cell lung cancer and 12 with EGFR/ALK-positive mutations were excluded. In total, 47 eligible patients were included in the current study from cohort 1. For the second cohort, there were 56 patients from the Dana-Farber Cancer Institute (DFCI) with metastatic NSCLC treated with anti-PD-1 treatment (8). Seven patients with squamous cell lung cancer and thirteen with EGFR/ALK-positive mutations were excluded. As a result, a total of 36 eligible patients from cohort 2 were included in the current study. From the third cohort, 240 patients receiving only anti-PD-1 or a combination of anti-CTLA-4 and anti-PD-1 treatments were retrospectively collected from the Memorial Sloan Kettering Cancer Center (MSKCC) between April 2011 and January 2017 (10). From this cohort, 38 patients with squamous cell lung cancer and 37 with EGFR/ALK-positive mutations were excluded. A total of 165 eligible patients were selected from the MSKCC. Finally, a total of 83 patients from the first and second cohorts were included in the training cohort, and the remaining 165 from the third cohort were included in the validation cohort. The institutional review board of the Second Affiliated Hospital of Guizhou Medical University approved our clinical research design. We have been conducted in accordance with the World Medical Association's Declaration of Helsinki.

Study Design

In this study, a three-step approach was used to develop and validate the somatic mutation signature (SMS) in advanced NSCLC patients undergoing immunotherapy. First, on the basis of the least absolute shrinkage and selection operator (LASSO) method, we used the somatic mutation profiles of the WES database to select the optimal gene panel for predicting BOR. Second, we used the pattern classification of the support vector machine (SVM) algorithm to build a predictive model of the SMS according to clinical treatment response and genomic mutation profiling in the training cohort. The somatic mutations could be computationally evaluated using the various mutational databases and the severity of the disease could also be predicted using the SVM algorithms. The SMS model was validated in the independent MSKCC cohort. All patients were divided into two groups (SMS-low and SMS-high) on the basis of the optimal cut-off of the receiver operating characteristic (ROC) and analyzed to predict progression-free survival (PFS) and overall survival (OS). Furthermore, the SMS model was analyzed in multivariate analyses of PFS and OS and the application of different clinical variables in the combination set.

Best Overall Response, Progression-Free Survival, and Overall Survival

This research study aimed to examine the BOR, PFS, and OS. BOR was defined as a record of the best outcomes from the beginning of the study to the end of the treatment, which was confirmed after considering a variety of factors. Furthermore, it was evaluated using Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1, including complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). PFS was defined as the time from the start of ICI treatment to the first confirmation of PD with RECIST version 1.1 or death. OS is defined as the time from the start of ICI treatment until death or the last contact.

Whole Exome Sequencing and Targeted Next-Generation Sequencing

Tumor and blood samples from WES and targeted NGS profiles were collected before immunotherapy. DNA was extracted from formalin-fixed paraffin-embedded tumor masses and blood samples from patients. The 83 samples from the CheckMate-012 study (n = 36) and DFCI (n = 47) were tested using WES profiling and the 165 from the MSKCC using targeted NGS profiling as follows.

WES: Illumina Rapid Capture Exome Target Bait Kit (38 MB), Agilent Sure-Select Human All Exon v2.0 (44 MB)/v4.0 (51 MB) created a complete library for capturing exons. To do so, the HiSeq 2000, 2500, or 4000 platform (Illumina, San Diego, California) was used to provide 150 times the average target coverage, extensive exon library, and paired reads (2 × 76 bp). For each sample, the Burrows-Wheeler Aligner was used to generate a normal BAM file, and the tumor sequence was compared to the human hg19 genome construct. The Genome Analysis Toolkit was used to analyze basic quality factor recalibration, indel recombination, and duplicate deletion. Indel calls were generated using Indelocator software.[1] The TMB in each sample was defined as the total number of non-synonymous mutations, including SNVs and indels.

Targeted NGS (MSK-IMPACT): Targeted NGS analysis was performed as described in a previous study (10). After amplification and sequencing, the barcode library targeted exons and chose introns of 468 (13 patients, version 3), 410 (116 patients; version 2), or 341 (36 patients; version 1) genes. In all tumor samples, the average sequence index was 7,443, and the minimum coverage depth was 913. A custom pipeline was used to identify the somatic alterations in the tumor samples. To normalize somatic non-synonymous TMB on panels of different sizes, we divided the detected coding regions in each panel by the total number of mutations, which covered 0.98 megabases (Mb) of the 341, 1.06 Mb of the 410, and 1.22 Mb of the 468 gene panels.

[1]http://archive.broadinstitute.org/

TABLE 1 | Characteristics of patients in the training and validation cohorts.

Variable	Training cohort (n = 83)	Validation cohort (n = 165)	P-value
Age (years)			0.230
≤ 60	36 (43.37%)	59 (35.75%)	
> 60	47 (56.63%)	106 (64.25%)	
Sex			0.560
Female	48 (57.83%)	89 (53.93%)	
Male	35 (42.17%)	76 (46.07%)	
Smoking status			0.828
Never	15 (18.07%)	28 (16.97%)	
Ever	68 (81.93%)	137 (83.03%)	
PD-L1 expression (%)			<0.001*
NA	0	104 (63.03%)	
≤ 50	49 (59.03%)	45 (27.27)	
> 50	34 (40.97%)	16 (9.7%)	
TMB			0.240
High	35 (42.17%)	57 (34.55%)	
Low	48 (57.83%)	108 (65.45%)	
Best overall response			0.014*
CR/PR	32 (38.55%)	39 (23.64%)	
SD/PD	51 (61.45%)	126 (76.36%)	

P-value is derived from the difference between the training and validation cohorts in either of the clinical characteristics.
NA, not available; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; TMB, tumor mutation burden.
**P-value < 0.05.*

Tumor Mutational Burden and Programmed Cell Death-Ligand 1 Testing Analysis

On the basis of the results of WES and targeted NGS profiling, we defined a high TMB as > 10/Mb or total somatic non-synonymous as ≥ 200 and low TMB as ≤ 10/Mb or total somatic non-synonymous < 200. The 22C3 (DAKO), 28-8 (DAKO), and E1L3N (Cell Signaling, Danvers, MA) were performed according to the manufacturer's instructions of three antibodies. We used the percentage of membranous staining in the tumor cells to evaluate the PD-L1 expression. In this study, high PD-L1 expression was defined as > 50% of tumor staining.

Statistical Analyses

LASSO is used to select the optimal gene panel and build a predictive model by choosing the most important variables from high-dimensional features (17, 18). The purpose of LASSO is to construct a first-order penalty function to obtain a refined model *via* the final determination of some variable coefficient 0 for feature screening. In this study, a LASSO method based on fivefold cross-validation was used to select 15 non-zero coefficients. Then, the SMS based on an SVM algorithm was built to predict the immunotherapy response. The performance of the SMS model was then evaluated in the training and validation cohorts by using ROC analysis. The optimal cut-off value for predicting BOR was defined using the maximum Youden index.

A X^2-test was used to compare the SMS score with the BOR in MSK cohorts. The "pROC" package was used to plot the ROC curves and evaluate the accuracy. A confidence interval (CI) of 95% for the area under the curve (AUC) was computed for the training and validation cohorts. The Kaplan–Meier curves of PFS and OS were analyzed and plotted using the "survminer" package. Additionally, multivariate Cox regression analysis was performed on five variables: SMS, age, sex, smoker status, PD-L1, and TMB using the "rms" package. The "Forestplot" software package was used to analyze and visualize hazard ratios (HRs) for PFS and OS in the SMS-low and SMS-high subgroups. All statistical analyses were performed using GraphPad Prism (version 7.01) and R software (version 3.5.1). Statistical significance was set at $P < 0.05$.

RESULTS

Characteristics of Patients

The basic characteristics of the patients in the training and validation cohorts are shown in **Table 1**. Of the patients,

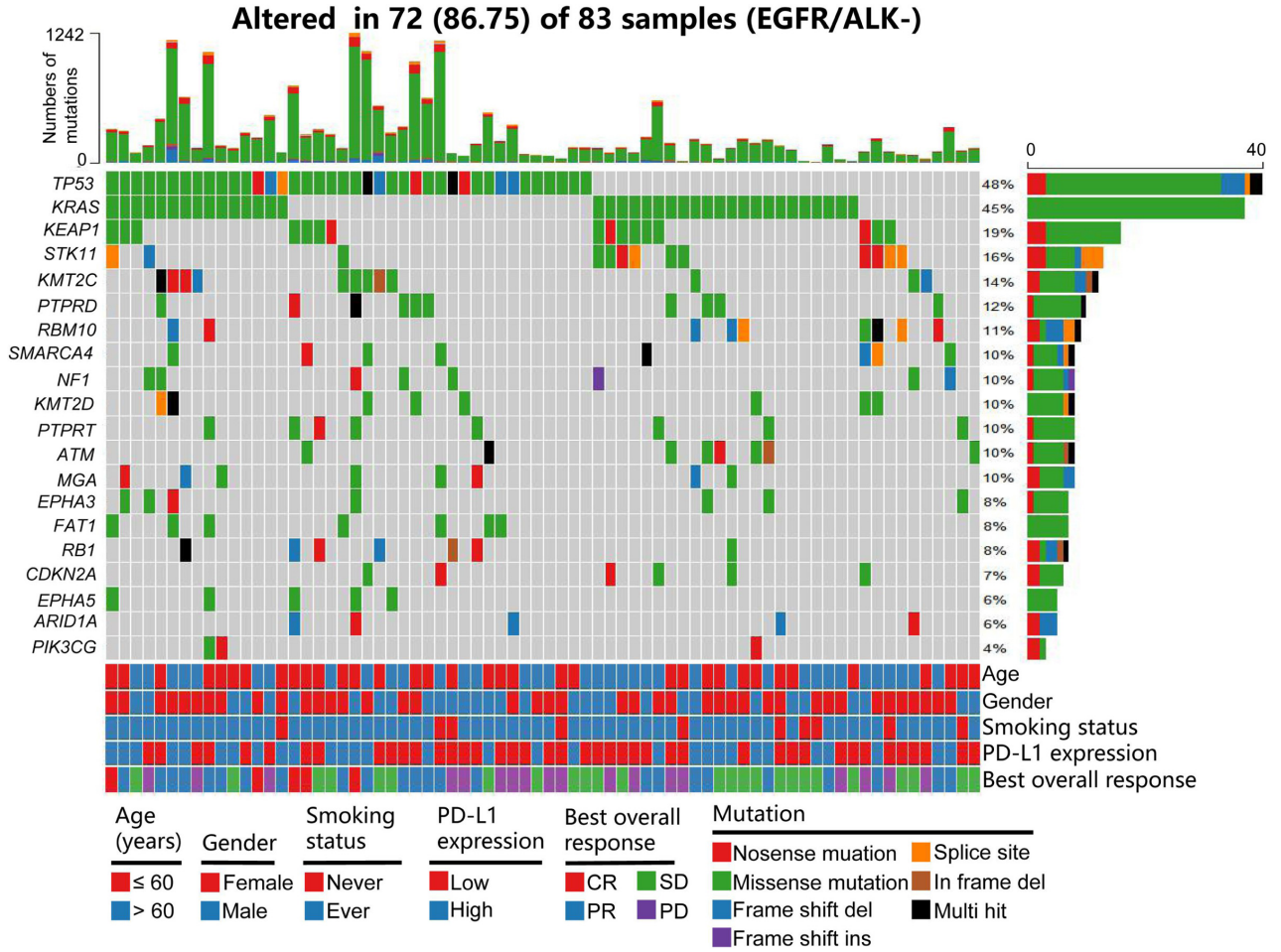

FIGURE 1 | Summary of clinical and molecular features associated with response of ICI-based therapy in the training cohort with patients with EGFR/ALK-negative non-squamous NSCLC. Individual patients are represented in each column. Age is stratified as ≤ 60 years or > 6 years; sex as female and male; smoking status as ever and never; PD-L1 expression as 0–49% or ≥ 50%; and BOR as CR, PR, SD, and PD. Mutations include 7 mutational subtypes, and the TMB of each patient is calculated. The occurrences of top 20 genes in each case are represented in the OncoPrint. PD-L1, programmed cell death-ligand 1; NSCLC, non-small cell lung cancer; BOR, best overall response; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; TMB, tumor mutation burden.

47 (56.63%) in the training cohort and 106 (64.25%) in the validation cohort were more than 60 years old. A total of 48 (57.83%) patients in the training cohort and 89 (53.93%) in the validation cohort were female. Most patients in the training and validation cohorts (81.93 and 83.03%, respectively) had a history of smoking. Thirty-four (40.97%) patients in the training cohort and sixteen (9.7%) in the validation cohort had high PD-L1 expression (>50%). The samples of 104 (63.03%) patients were not tested for PD-L1 expression in the validation cohort. Patients with high TMB (>10/Mb or ≥ 200) accounted for 42.17% of the training and 34.55% of the validation cohorts. In the evaluation of BOR, 32 (38.55%) patients in the training and 39 (23.64%) in the validation cohorts achieved CR/PR. We found that the frequencies of PD-L1 expression and BOR were inconsistent between the training and validation cohorts ($P < 0.001$ and $P = 0.014$), but there were no significant differences in the other clinical characteristics between the two cohorts. **Figure 1** shows a summary of the clinical and molecular features associated with the response of ICI-based therapy in the training cohort with EGFR/ALK-negative non-squamous NSCLC. Mutations include 7 mutational subtypes, and the occurrences of top 20 genes in each case are represented in OncoPrint.

Development and Validation of Somatic Mutation Signature for Best Overall Response

On the basis of 5-fold cross-validation of LASSO, 15 genes with somatic mutations were selected and used to build a model of the SMS (**Figures 2A,B**). We used a gene panel of 15 somatic mutations to train the model by using the SVM method, and the SMS was built to predict the BOR after fine tuning. The SMS was compared with the *TP53/KRAS/KEAP1/STK11* driver genes, TMB, and PD-L1 expression in both cohorts (**Figures 2C,D**). We found that the SMS showed high AUC in the training and validation cohorts (AUC = 0.859, 95% CI: 0.767–0.951,

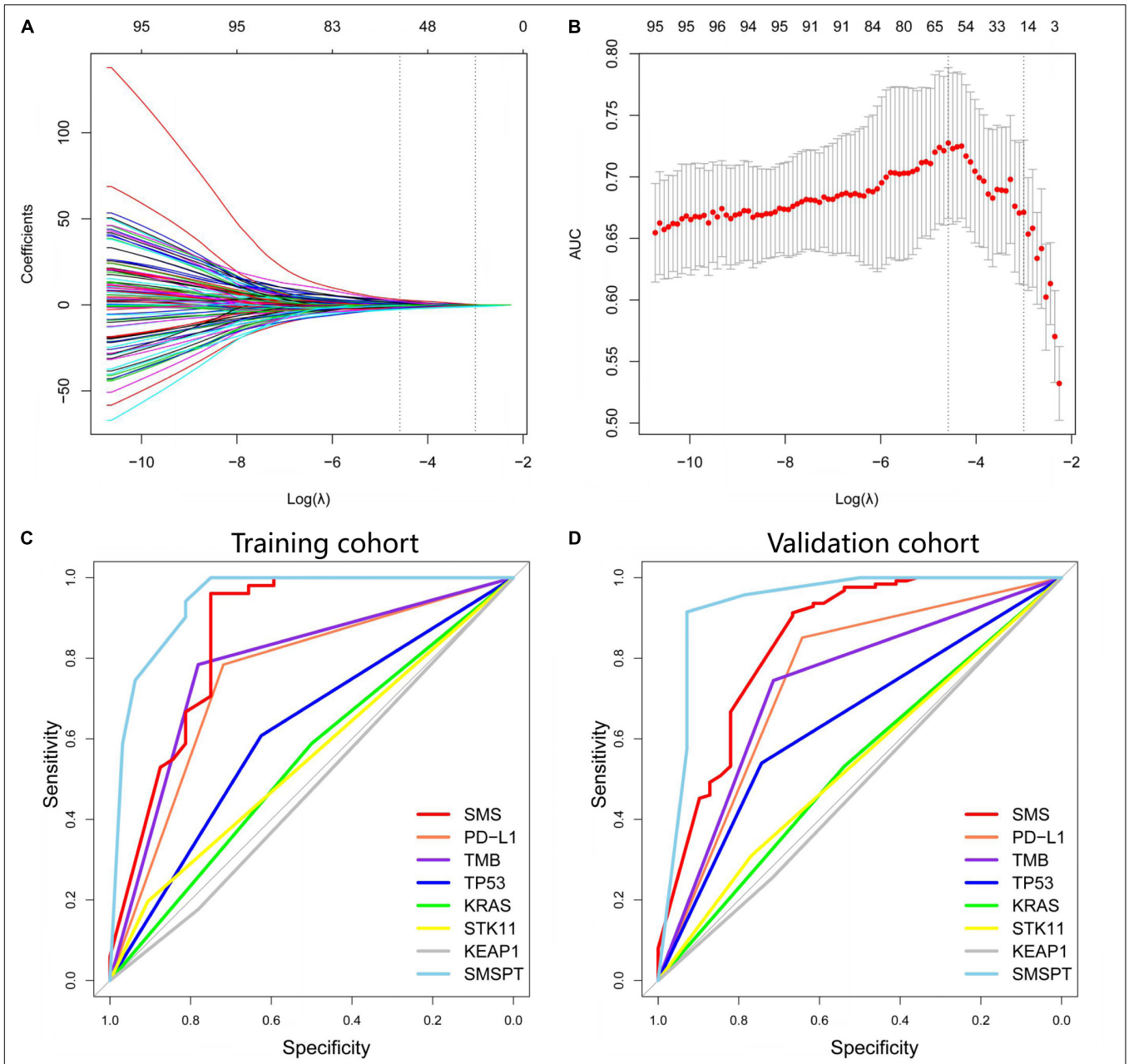

FIGURE 2 | Selection of somatic mutation genes using the LASSO method and building the SMS model for BOR. **(A)** Fifteen non-zero coefficients are selected, and the coefficients are plotted against the log (λ) sequence. **(B)** LASSO coefficient analysis of the somatic mutation genes is shown by 10-fold cross-validation. The minimum value of log (λ) is −1.44 based on the 1-SE criteria. **(C,D)** Based on AUCs plotting, performances of SMS, PD-L1, TMB, *TP53, KRAS, STK11, KEAP1,* and SMSPT in the training and validation cohorts are presented. LASSO, least absolute shrinkage and selection operator; SMS, somatic mutation signature; BOR, best overall response; PD-L1, programmed cell death-ligand 1; TMB, tumor mutation burden; SMSPT, integrating SMS, PD-L1, and TMB; AUC, area under the curve.

sensitivity = 96.08%, 95% CI: 86.54–99.52%, specificity = 75.00%, 95% CI: 56.60–88.54%, $P < 0.001$; AUC = 0.841, 95% CI: 0.761–0.922, sensitivity = 91.27%, 95% CI: 84.92–95.56%, specificity = 61.67%, 95% CI: 49.78–80.91%, $P < 0.001$, respectively) (**Supplementary Table 1**). *TP53* mutation positively correlated with BOR in the training and validation cohorts ($P = 0.075$ and $P < 0.001$, respectively), and we did not find an association between *KRAS/KEAP1/STK11* mutations and BOR ($P > 0.05$ each). PD-L1 expression and TMB were also significantly associated with BOR in the training (AUC = 0.751, 95% CI: 0.639–0.863, sensitivity = 78.43%, 95% CI: 64.68–88.71%, specificity = 71.88%, 95% CI: 53.25–86.25%, $P < 0.001$; AUC = 0.817, 95% CI: 0.721–0.913, sensitivity = 78.43%, 95% CI: 64.68–88.71%, specificity = 78.13%, 95% CI: 60.03–90.72%, $P < 0.001$) and validation cohorts (AUC = 0.747, 95% CI: 0.585–0.908, sensitivity = 85.11%, 95% CI: 71.69–93.80%,

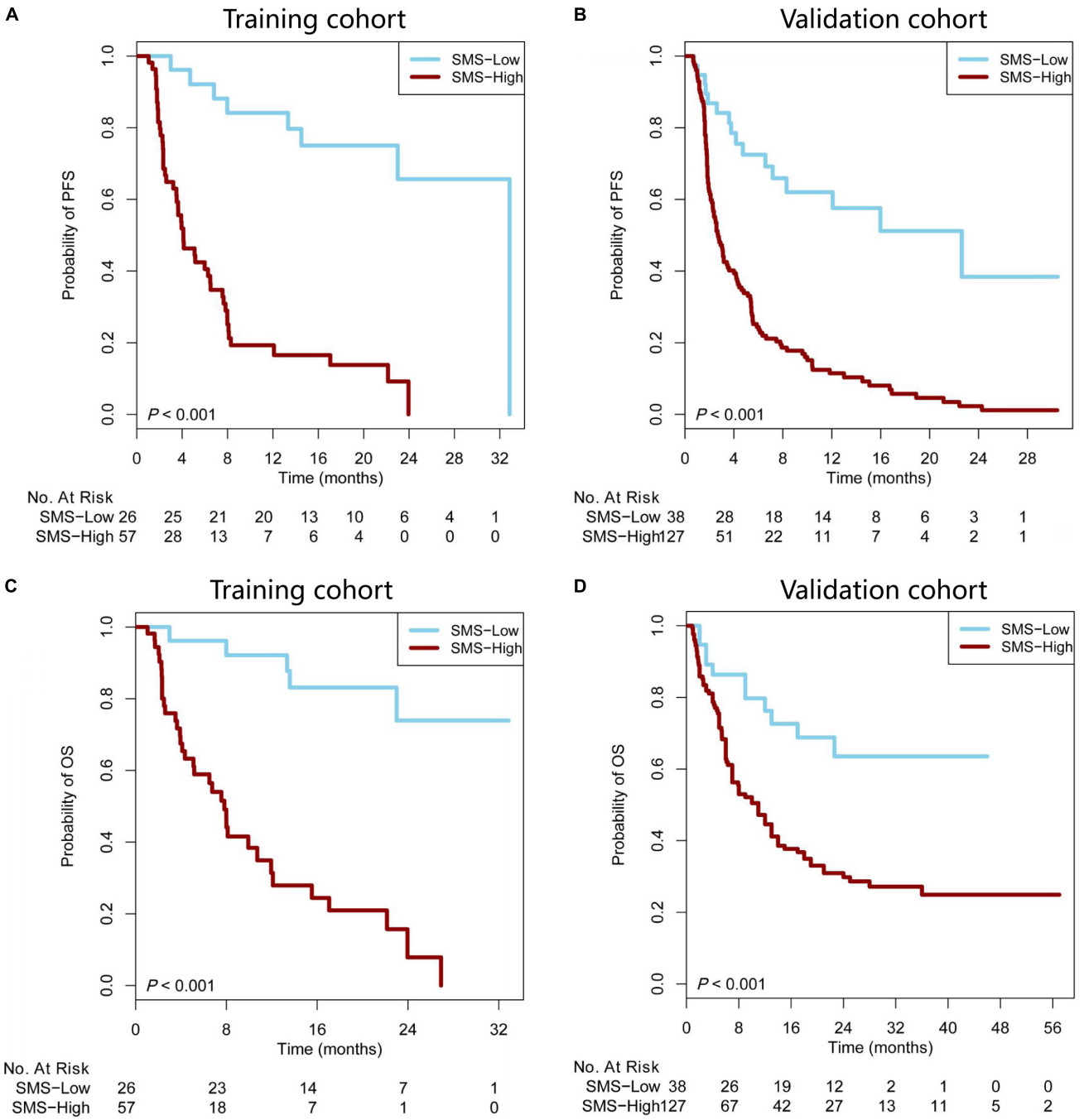

FIGURE 3 | The SMS predicts PFS and OS in patients with immunotherapy. **(A,B)** Kaplan–Meier survival curves showing PFS between the SMS-low and -high groups in patients from the training and validation cohorts treated with anti-PD-1 therapy. **(C,D)** Kaplan–Meier survival curves showing OS between the SMS-low and -high groups in patients from the training and validation cohorts treated with anti-PD-1 therapy. SMS, somatic mutation signature; PFS, progression-free survival; OS, overall survival; NR, not reached; PD-1, programmed cell death 1.

specificity = 64.29%, 95% CI: 35.14–87.24%, $P < 0.001$; AUC = 0.657, 95% CI: 0.558–0.757, sensitivity = 67.46%, 95% CI: 58.54–75.54%, specificity = 64.10%, 95% CI: 47.18–78.8%, $P = 0.002$). Furthermore, we used a logistic model integrating the SMS, PD-L1, and TMB to build a combination model named SMSPT, which showed high AUCs in the training and validation cohorts (AUC = 0.937, 95% CI: 0.886–0.988, sensitivity = 94.12%, 95% CI: 83.76–98.77%, specificity = 81.25%, 95% CI: 63.56–92.79%, $P < 0.001$; AUC = 0.933, 95% CI: 0.833–1.000, sensitivity = 91.49%, 95% CI: 79.62–97.63%, specificity = 92.86%, 95% CI: 66.13–99.82%, $P < 0.001$, respectively) (**Figures 2C,D** and **Supplementary Table 1**).

Somatic Mutation Signature Predicts Progression-Free Survival and Overall Survival in Patients With Immunotherapy

According to the cut-off value (SMS scores = 1.95), our patients were stratified into SMS-high (>1.95) or SMS-low (≤ 1.95) groups. Compared with the SMS-low group in the training cohort with anti-PD-1 therapy, the SMS-high group showed a poorer median PFS (mPFS: 4.11 vs. 32.86 months) and OS [mOS: 7.81 months vs. not reached (NR)] [HR = 6.01 (3.54–10.20), $P < 0.001$; HR = 7.60 (4.12–14.03), $P < 0.001$, respectively] (**Figures 3A,B**). We then tested the SMS model in the validation cohort and found that the SMS-high group also presented a poorer median PFS (mPFS: 2.70 vs. 22.63 months) and OS (mOS: 11.00 months vs. NR) [HR = 3.89 (2.72–5.54), $P < 0.001$; HR = 2.82 (1.80–4.41), $P < 0.001$, respectively] than did the SMS-low group (**Figures 3C,D**). Considering 104 patients without PD-L1 expression tested in the validation cohort, we combined 83 from the training cohort and 61 from the validation cohort to build a combination cohort (n = 144) and performed a multivariate analysis of PFS and OS. We found that PD-L1, TMB, and the SMS were independent predictors of PFS in anti-PD-1 therapy (**Table 2**). Moreover, smoking status, PD-L1, TMB, and the SMS were also independent predictors of OS in this study.

Applicability of Somatic Mutation Signature in Epidermal Growth Factor Receptor/Anaplastic Lymphoma Kinase-Negative Non-small Cell Lung Cancer Patients With Different Clinical Variables

We further analyzed whether the SMS predictive model was feasible in specific groups of all EGFR/ALK-negative NSCLC patients. According to the basic clinical characteristics in the combination cohort, a univariate subgroup of PFS and OS was analyzed using the SMS (**Figures 4A,B**). The patients with SMS-high had significantly shorter PFS and OS regardless of age (≤ 60 vs. > 60 years) and sex (male vs. female). However, the SMS showed differentiated predictive values for anti-PD-L1 therapy in smokers. The SMS predicted the PFS and OS better in patients who are ever smoker than in patients who are never smokers. But the number of patients who never smoked was small (n = 27). Interestingly, we found that the SMS had good predictive ability in the subgroups of TMB and PD-L1 expression (**Figures 4A,B**). The SMS in the high TMB subgroup had better predictive ability for PFS and OS than that in the low TMB subgroup. In addition, the SMS had better predictive ability for PFS and OS in the low PD-L1 expression subgroup than in the high PD-L1 expression subgroup.

DISCUSSION

In this study, we found that the SVM classification of somatic mutations could predict BOR in patients with EGFR/ALK-negative NSCLC treated with anti-PD-1. In two independent cohorts, we found that the accuracy of the SVM model was greater than that of PD-L1 or TMB expression. In the two groups, patients treated with ICIs in the SMS-low group had better OS and PFS than those in the SMS-high group. We also found that the SMS model could predict prognosis in several clinical subgroups.

Previous studies have used sequencing techniques, including WES, WGS, and NGS, to analyze the association between genomic variants and prognosis (19–21). In most studies of immunotherapy (22, 23), specific genes were studied, and we found that it made it difficult to realize the precise predictive value for benefits from anti-PD-1 therapy in patients. In the current study, *TP53*, but not *KRAS*, mutations were positively associated with immunotherapy BOR in driver mutations, and *STK11* or *KEAP1* mutations were not significantly related to BOR. These results revealed that the single genomic mutation had weak predictive abilities for different molecular statuses, potentially resulting from tumor heterogeneity. Currently, WES and NGS testing of tumor tissue and blood samples have been used to quantify TMB in various solid tumors (24, 25). TMB is frequently calculated from the accumulation number of non-synonymous mutations, and high TMB has been reported to be correlated with a good response to anti-PD-1 therapy (3, 4, 26). However, the predictive value of TMB is controversial, and the accuracy is not satisfactory. Thus, we used the SMS model based on somatic mutations derived from targeted WES and NGS to determine BOR for anti-PD-1 therapy in patients with EGFR/ALK-negative NSCLC. The SMS classifications could precisely predict BOR and No-BOR in patients. We found that the SMS had a more accurate

TABLE 2 | Multivariate analyses of PFS and OS in combination cohort (n = 144).

Variable	PFS		OS	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age (≤ 60 vs. > 60)	0.80 (0.52–1.21)	0.291	1.02 (0.68–1.54)	0.096
Sex (female vs. male)	0.84 (0.56–1.26)	0.417	0.70 (0.46–1.06)	0.910
Smoker status (never vs. ever)	1.24 (0.74–2.06)	0.397	1.958 (1.14–3.35)	0.014*
TMB (high vs. low)	0.44 (0.26–0.74)	0.002*	0.57 (0.35–0.92)	0.023*
PD-L1 (high vs. low)	0.28 (0.17–0.46)	<0.001*	0.36 (0.22–0.59)	<0.001*
SMS (high vs. low)	4.32 (2.32–8.06)	<0.001*	3.07 (1.71–5.49)	<0.001*

PFS, progression free survival; OS, overall survival; HR, hazard ratio; CI, confidence interval; TMB, tumor mutation burden; SMS, somatic mutation signature.
**P-value < 0.05.*

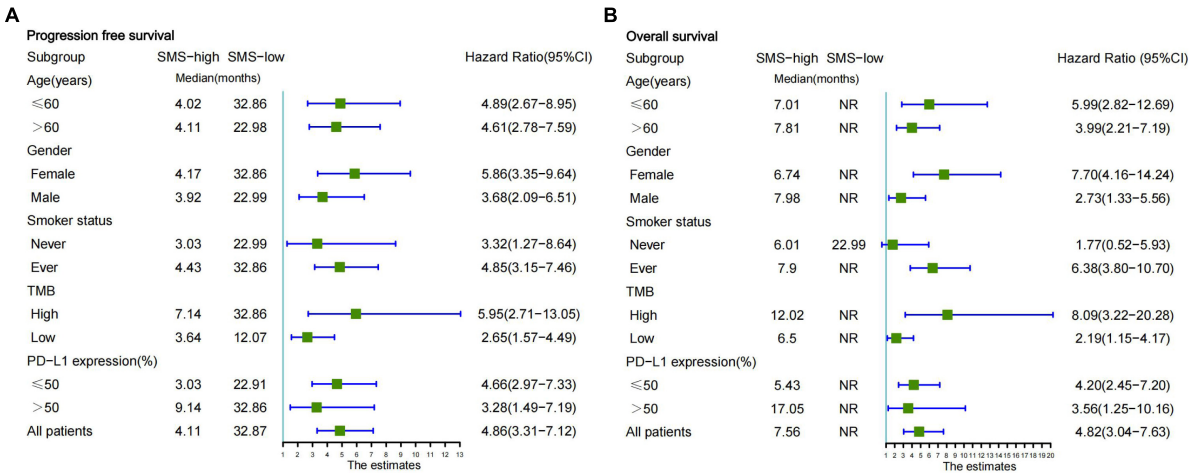

FIGURE 4 | Subgroup analysis of the SMS for PFS and OS from the combination cohort according to basic clinical variables. **(A)** Forestplot of the SMS for PFS is presented in the combination cohort. **(B)** Forestplot of the SMS for OS is presented in the combination cohort. Each subgroup HR is computed from univariate analysis. Fixed-effects model is used to calculate pooled HRs for each subgroup. The bars indicate 95% CI. SMS, somatic mutation signature; PFS, progression-free survival; OS, overall survival; NR, not reached; PD-L1, programmed cell death-ligand 1; HR, hazard ratio; CI, confidence interval.

prediction than TMB and PD-L1 expression. Interestingly, the comprehensive model integrating the SMS, TMB, and PD-L1 expression had high predictive accuracy in both the training and validation cohorts. This indicates that not only TMB and PD-L1 expression testing but also fully mining mutation features based on machine learning is helpful in improving prediction ability. To help clinical practice, our model can be freely used online on a computer or mobile phone.[2] Thus, the novel method is easy to use and could potentially screen patients with NSCLC for benefits from immunotherapy.

Previous studies show that through the use of immunotherapy for various cancers, patients who attained CR/PR frequently have better prognosis (27–29). We further analyzed the association between the SMS and prognosis and found that patients with SMS-low showed significantly longer PFS and OS than patients with SMS-high in both cohorts. This result suggests that the SMS model based on predicting BOR could effectively evaluate the clinical outcome of immunotherapy in the molecular subgroup of EGFR/ALK-negative NSCLC. In multivariate analyses of PFS and OS in the combination cohort, we found that the SMS, PD-L1 expression, and TMB were independent predictive factors, suggesting that the SMS model based on somatic mutations could be considered a novel biomarker for predicting prognosis. We also found that smoking status was an independent factor for OS. Patients who have smoked might have more mutant antigens causing lung cancer, which has been revealed in previous studies of immunotherapy (30, 31). Subgroup analysis of immunotherapy in patients with EGFR/ALK-negative NSCLC with SMS-low showed significantly better PFS than those with SMS-high. We found that the SMS model showed better prediction of OS in subgroups of smoking status (never), TMB-high, and PD-L1 low expression than in those of smoking status (ever), TMB-low, and PD-L1 high expression. Of note, the number of patients with smoking status "never" was relatively small. Additionally, all patients with SMS-low showed longer medium-OS time than those with SMS-high did. This indicated that our SMS model could serve as a well-stratified tool and improve the value of TMB or PD-L1 expression for predicting the prognosis in patients with EGFR/ALK-negative NSCLC receiving anti-PD-1/PD-L1 therapy.

Our study has three limitations. First, the sample size was relatively small, and the three cohorts were from American Medical Centers. Although the result of predicting response to immunotherapy was performed well, a large prospective study based on this SMS model should be tested across an international multicenter population in a clinical trial. Second, our study focused on mutational genes, and tumor heterogeneity might affect the results of genomic variants in WES or NGS. Thus, a multi-omics model, including tumor genomics, radiology, and pathology, should be considered to predict the response. Third, the sequencing tumor tissues were obtained from biopsy or surgery, and this was an invasive procedure. An SMS model based on the ctDNA of peripheral blood would be a non-invasive model that should be further investigated in the future.

Overall, our research supports the 15-gene SMS classification as a reliable prediction tool for identifying patients who may benefit from anti-PD-1 treatment in patients with EGFR/ALK-negative NSCLC. The new findings described in this study may help us develop a sequence database to explore new strategies for cancer immunotherapy. In the future, comprehensive pan-cancer research is needed to make better use of multigene SMS panels as predictive biomarkers for immunotherapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

[2]https://pengjie.shinyapps.io/Somatic/

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 Second Affiliated Hospital of Guizhou Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JP: conception, design, data analysis and interpretation, and administrative support. JP and DZ: provision of study materials or patients. JP and LX: collection and assembly of data. JP, LX, DZ, and LH: manuscript writing and final approval of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Qian Dong Nan Science and Technology Program (grant no. qdnkhJz2020-013) and the Science and Technology Foundation of Guizhou Province (grant no. Qian Ke He Ji Chu-ZK 2021, Yi Ban 454).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.808378/full#supplementary-material

REFERENCES

1. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med.* (2018) 378:2288–301. doi: 10.1056/NEJMoa1716948
2. Lemery S, Keegan P, Pazdur R. First FDA approval agnostic of cancer site - when a biomarker defines the indication. *N Engl J Med.* (2017) 377:1409–12. doi: 10.1056/NEJMp1709968
3. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet.* (2019) 51:202–6.
4. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther.* (2017) 16:2598–608. doi: 10.1158/1535-7163.MCT-17-0386
5. Cristescu RA-O, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science.* (2018) 362:eaar3593. doi: 10.1126/science.aar3593
6. Brahmer JR, Rodriguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Health-related quality-of-life results for pembrolizumab versus chemotherapy in advanced, PD-L1-positive NSCLC (KEYNOTE-024): a multicentre, international, randomised, open-label phase 3 trial. *Lancet Oncol.* (2017) 18:1600–9. doi: 10.1016/S1470-2045(17)30690-3
7. Diem S, Hasan Ali O, Ackermann CJ, Bomze D, Koelzer VH, Jochum W, et al. Tumor infiltrating lymphocytes in lymph node metastases of stage III melanoma correspond to response and survival in nine patients treated with ipilimumab at the time of stage IV disease. *Cancer Immunol Immunother.* (2018) 67:39–45. doi: 10.1007/s00262-017-2061-4
8. Miao D, Margolis CA, Vokes NI, Liu D, Taylor-Weiner A, Wankowicz SM, et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet.* (2018) 50:1271–81. doi: 10.1038/s41588-018-0200-2
9. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R, et al. Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res.* (2017) 23:4242–50. doi: 10.1158/1078-0432.CCR-16-3133
10. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol.* (2018) 36:633–41. doi: 10.1200/JCO.2017.75.3384
11. Peng J, Zou D, Gong W, Kang S, Han LJ. Deep neural network classification based on somatic mutations potentially predicts clinical benefit of immune checkpoint blockade in lung adenocarcinoma. *Oncoimmunology.* (2020) 9:1734156. doi: 10.1080/2162402X.2020.1734156
12. Skoulidis F, Goldberg ME, Greenawalt DM, Hellmann MD, Awad MM, Gainor JF, et al. STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov.* (2018) 8:822–35. doi: 10.1158/2159-8290.CD-18-0099
13. Biton J, Mansuet-Lupo A, Pecuchet N, Alifano M, Ouakrim H, Arrondeau J, et al. TP53, STK11, and EGFR mutations predict tumor immune profile and the response to anti-PD-1 in lung adenocarcinoma. *Clin Cancer Res.* (2018) 24:5710–23. doi: 10.1158/1078-0432.CCR-18-0163
14. Wang ZJ, Zhao J, Wang G, Zhang F, Zhang ZM, Zhang F, et al. Comutations in DNA damage response pathways serve as potential biomarkers for immune checkpoint blockade. *Cancer Res.* (2018) 78:6486–96. doi: 10.1158/0008-5472.CAN-18-1814
15. Li X, Wang Y, Li X, Feng G, Hu S, Bai YF, et al. The impact of NOTCH pathway alteration on tumor microenvironment and clinical survival of immune checkpoint inhibitors in NSCLC. *Front Immunol.* (2021) 12:638763. doi: 10.3389/fimmu.2021.638763
16. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Vega FS, Ahuja A, et al. Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell.* (2018) 33:843.e–52.e. doi: 10.1016/j.ccell.2018.03.018
17. Peng J, Zhang J, Zhang Q, Xu YK, Zhou J, Liu L, et al. A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma. *Diagnos Intervent Radiol.* (2018) 24:121–7. doi: 10.5152/dir.2018.17467
18. Han L, Zhao K, Li Y, Han HH, Zhou LZ, Ma P, et al. A gut microbiota score predicting acute graft-versus-host disease following myeloablative allogeneic hematopoietic stem cell transplantation. *Am J Transplant.* (2020) 20:1014–27. doi: 10.1111/ajt.15654
19. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* (2017) 23:703–13.
20. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med.* (2017) 376:2109–21.
21. Campbell JA-O, Alexandrov AA-O, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet.* (2016) 48:607–16. doi: 10.1038/ng.3564
22. Dong ZY, Zhong WZ, Zhang XC, Su J, Xie Z, Liu SY, et al. Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. *Clin Cancer Res.* (2017) 23:3012–24. doi: 10.1158/1078-0432.CCR-16-2554
23. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* (2017) 127:2930–40. doi: 10.1172/JCI91190

24. Wang Z, Duan J, Cai S, Han M, Dong H, Zhao J, et al. Assessment of blood tumor mutational burden as a potential biomarker for immunotherapy in patients with non-small cell lung cancer with use of a next-generation sequencing cancer gene panel. *JAMA Oncol.* (2019) 5:696–702. doi: 10.1001/jamaoncol.2018.7098
25. Ricciuti B, Kravets S, Dahlberg SE, Umeton R, Albayrak A, Subegdjo SJ, et al. Use of targeted next generation sequencing to characterize tumor mutational burden and efficacy of immune checkpoint inhibition in small cell lung cancer. *J Immunother Cancer.* (2019) 7:87. doi: 10.1186/s40425-019-0572-6
26. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
27. Johannet PA-O, Coudray NA-O, Donnelly DM, Jour G, Bochaca II, Xia YH, et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. *Clin Cancer Res.* (2021) 27:131–40. doi: 10.1158/1078-0432.CCR-20-2415
28. Harding JJ, Nandakumar S, Armenia J, Khalil DN, Albano M, Ly M, et al. Prospective genotyping of hepatocellular carcinoma: clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin Cancer Res.* (2019) 25:2116–26. doi: 10.1158/1078-0432.CCR-18-2293
29. Arbour KA-O, Luu AT, Luo JA-OX, Rizvi H, Plodkowski AJ, Sakhi M, et al. Deep Learning to Estimate RECIST in Patients with NSCLC Treated with PD-1 Blockade. *Cancer Discov.* (2021) 11:59–67.
30. Zhao W, Jiang W, Wang H, He JB, Su CY, Yu QT, et al. Impact of smoking history on response to immunotherapy in non-small-cell lung cancer: a systematic review and meta-analysis. *Front Oncol.* (2021) 11:703143. doi: 10.3389/fonc.2021.703143
31. Zaleskis G, Pasukoniene V, Characiejus D, Urbonas V. Do the benefits of being a smoker hint at the existence of PD-1/PD-L1 sensitizers for patients on single-agent immunotherapy? *J Immunother Cancer.* (2021) 9:e003191. doi: 10.1136/jitc-2021-003191

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

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

frontiers | Frontiers in Medicine

REVIEW
published: 14 June 2022
doi: 10.3389/fmed.2022.766869

Precision Medicine: An Optimal Approach to Patient Care in Renal Cell Carcinoma

Revati Sharma[1,2], *George Kannourakis*[1,2], *Prashanth Prithviraj*[1,2†] *and Nuzhat Ahmed*[1,2,3,4*†]

[1] Fiona Elsey Cancer Research Institute, Ballarat Central Technology Central Park, Ballarat Central, VIC, Australia, [2] School of Science, Psychology and Sport, Federation University, Mt Helen, VIC, Australia, [3] Centre for Reproductive Health, Hudson Institute of Medical Research and Department of Translational Medicine, Monash University, Clayton, VIC, Australia, [4] Department of Obstetrics and Gynaecology, University of Melbourne, Melbourne, VIC, Australia

Renal cell cancer (RCC) is a heterogeneous tumor that shows both intra- and inter-heterogeneity. Heterogeneity is displayed not only in different patients but also among RCC cells in the same tumor, which makes treatment difficult because of varying degrees of responses generated in RCC heterogeneous tumor cells even with targeted treatment. In that context, precision medicine (PM), in terms of individualized treatment catered for a specific patient or groups of patients, can shift the paradigm of treatment in the clinical management of RCC. Recent progress in the biochemical, molecular, and histological characteristics of RCC has thrown light on many deregulated pathways involved in the pathogenesis of RCC. As PM-based therapies are rapidly evolving and few are already in current clinical practice in oncology, one can expect that PM will expand its way toward the robust treatment of patients with RCC. This article provides a comprehensive background on recent strategies and breakthroughs of PM in oncology and provides an overview of the potential applicability of PM in RCC. The article also highlights the drawbacks of PM and provides a holistic approach that goes beyond the involvement of clinicians and encompasses appropriate legislative and administrative care imparted by the healthcare system and insurance providers. It is anticipated that combined efforts from all sectors involved will make PM accessible to RCC and other patients with cancer, making a tremendous positive leap on individualized treatment strategies. This will subsequently enhance the quality of life of patients.

Keywords: renal cell carcinoma, gut microbiome, artificial intelligence, precision medicine, nanomedicine

OPEN ACCESS

Edited by:
Alice Chen,
National Cancer Institute (NCI),
United States

Reviewed by:
Sean R. Williamson,
Cleveland Clinic, United States
Rahul Mannan,
University of Michigan, United States

***Correspondence:**
Nuzhat Ahmed
nuzhata@unimelb.edu.au;
nuzhat@fecri.org.au

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *30 August 2021*
Accepted: *11 April 2022*
Published: *14 June 2022*

Citation:
Sharma R, Kannourakis G, Prithviraj P and Ahmed N (2022) Precision Medicine: An Optimal Approach to Patient Care in Renal Cell Carcinoma. Front. Med. 9:766869. doi: 10.3389/fmed.2022.766869

INTRODUCTION

Renal cell carcinoma (RCC), an inherently heterogeneous group of cancers, is one of the 10 most common cancers worldwide and accounts for 2% of global cancer cases (1, 2). The incidence of RCC is more common in developed countries, with Belarus having the highest incidence in the world, which has doubled in the last 50 years (2). The incidence of RCC varies considerably among various population and geographical regions. The geographic distribution of RCC shows mainly higher age standardized rate of incidences in Eastern Europe and North America followed by Africa (3). RCC is also known to be influenced by genetic factors, for instance, the risk of RCC is elevated by two-fold for people who have first-degree relative with a history of RCC (4).

Genome wide association studies of RCC have shown multiple susceptible loci on the chromosomal regions of 2p21, 2q22.3, 8q24.21,11q13.3, 12p11.23 and 12q24.31 predisposing those harboring these SNPs to RCC (5). RCC accounts for more than 100,000 deaths worldwide each year (6). Most RCC arises from accumulated cell mutation, leading to uncontrolled cell growth in the proximal convoluted tubule located in the cortex of the kidney. It is an insidious and highly heterogeneous cancer, which in most cases is dominated by mutation in the VHL gene function (2, 7). However, RCC is not considered a single entity as it comprises multiple subtypes, each having characteristic histopathological features along with unique genetic traits and clinical outcomes.

RCC, which arises from the nephron tubules is a heterogeneous group of neoplasm having diverse histological subtypes with varying clinical courses and responses to therapy. The histopathological classification of RCC recently has gone through some major changes owing to the key advances in morphological, genetical and epidemiological understanding of RCC subtypes. The 2016 World health organization (WHO) classification of renal tumors is based on the combination of these RCC features and consists of 16 different subtypes and 4 provisional or emerging entities of RCC (8). The 2016 WHO classification includes subtypes that have been classified largely based on their (i) cytoplasmic features; clear cell RCC that arises from proximal tubules and chromophobe RCCs which arises from intercalated cells of distal tubules; (ii) architectural features; papillary RCC arising from distal tubular and mucinous tubular epithelium; (iii) spindle cell carcinoma arising from the cells of the loop of Henle or the collecting duct epithelium; and (iv) anatomical position of the tumors; collecting duct carcinoma arising from the collecting ducts of the kidney and renal medullary carcinoma that develop in the medullar region of the kidneys. The most common subtypes are clear cell RCC (ccRCC), accounting for 70–75% of the RCC cases followed by papillary RCC (pRCC), which is found in 10–15% of cases and chromophobe RCC accounting for 5% of the RCC cases. Oncocytomas comprising of benign tumors accounts for 4–7% of all kidney tumors. Around 4% of the RCCs are unclassified based on the currently available histopathological or molecular parameters. Amongst the new 7 subtypes that were added were based on molecular alterations such as MiT family translocation renal carcinomas, TRCC (originating from the translocation of transcription factor genes *TFE3* and *TFEB*) and succinate dehydrogenase or SDH deficient RCC caused by a biallelic mutation of one of the four subunits of SDH complex; familial predisposition syndrome associated hereditary leiomyomatosis and RCC syndrome associated specific renal disease acquired cystic disease associated RCC. The others that were included as new RCC entities were tubulocystic RCC; small-intermediate size tubules and cystically dilated larger tubules, multilocular cystic renal neoplasm of low malignant potential composed of cysts without expansive growth and clear cell papillary RCC, which shares morphological similarity with both ccRCC and pRCC. The provisional entities of RCC in the 2016 WHO classification include Oncocytic RCC occurring after neuroblastoma (increased risk of RCC after prior blastoma appearance similar to MiT family TRCC), thyroid-like follicular RCC (morphologically similar to the follicular carcinoma of the thyroid), Anaplastic Lymphoma Kinase (ALK) rearrangements-associated RCC (resembles medullary carcinomas and associated with sickle cell trait) and RCC with angioleiomyomatous stroma (sporadic or associated with tuberous sclerosis) (9). Around 4% of the RCCs are unclassified based on the currently available histopathological or molecular parameters (10).

Historically, RCCs are resistant to traditional cancer treatments like chemotherapy and radiotherapy (11). Treatment of RCC depends on the location of the tumor, stage of the disease, and overall health of the patient. Partial or radical nephrectomy (curative surgery) has been the gold standard for treating localized RCC from stages T1b-T4 (12). Ablation and active monitoring by ultrasonography are other popular choices for management of the localized disease. Despite the curative nature of surgery, approximately 30% of patients with localized ccRCC eventually develop recurrence or metastatic disease (13). More importantly, at the time of diagnosis, approximately 30% of patients present with locally advanced or metastatic disease (14). In the late 1980s, cytokines like interferon α (IFN-α) and interleukin-2 (IL-2) were the mainstays of treatment for metastatic RCCs (mRCCs) (15). However, these treatments were ineffective in terms of overall survival (OS), significantly toxic, and linked with high morbidity in patients (16). The subsequent advent of targeted therapies, mainly tyrosine kinase inhibitors (TKIs) such as sunitinib and pazopanib, has revolutionized the management of inoperable and mRCC. Sunitinib, pazopanib, and cabozantinib are TKIs approved as first-line treatments, while axitinib and sorafenib are second-line treatments available for patients with RCC. However, 30% of patients are innately resistant to these treatments, and 70% of the initial responders acquire resistance in 2 years (17, 18).

In addition, a better understanding of the immune system and a recent finding that tumors can exploit immune checkpoints to favor their survival and growth have led to the development of immunotherapeutic agents (ICIs). Immunotherapies in RCC oriented toward programmed cell death protein (PD-1), programmed cell death protein-ligand (PD-L1), and cytotoxic T lymphocyte antigen-4 (CTLA-4) target tumor and immune cells. Recent data from clinics have shown that a combination of targeted therapeutics and immunotherapy agents has increased the median survival for patients with RCC from 15 to 30 months (19).

Despite the progress made in the treatment of mRCC with TKIs and ICIs, most patients (∼85%) do not benefit from the current therapeutics and eventually succumb to the disease affecting the overall survival and quality of life (20). The effectiveness of the treatment depends on several factors related to patients, such as geography, socioeconomic condition, type and stage of cancer, patient's age, immune status, and overall general health. The mechanism of action of the drugs used and its interaction with the patient's immune system also contribute to the effectiveness of the drugs used (21). Hence, a more effective way of treating patients with mRCC relies on the development of specific individualized treatments tailored to the patient's explicit needs. The approach commonly termed precision medicine (PM)

is gaining momentum in the current arena of cancer treatment. For example, somatic inactivation of the VHL gene is genetically associated with patients with RCC and is one of the major factors regulating RCC pathophysiology. The major effect of an inactivated VHL gene is overexpression of vascular endothelial growth factor A (VEGFA), a key molecule accountable for pathological and physiological angiogenesis-related progression in patients with RCC (22). Agents approved to treat mRCC like sunitinib, which is a standard of treatment in the clinic, have been proven significantly effective as a first-line therapy in patients with RCC harboring VHL gene mutation, as they target VHL gene-associated hypoxia and related angiogenesis regulated mainly by VEGF and its receptors (23). Sunitinib specifically targets the downstream consequences of genetic mutation and, thereby, maximizes the efficacy and minimizes side effects in patients. Despite that, sunitinib resistance occurs in the majority of patients through complementary angiogenic pathways (24). Hence, developing approaches that will not only harness specific genotypes but also other molecular and lifestyle traits of patients are essential to integrating PM into the standard care of patients. This review describes some of the recent progress made in the field of precision medicine (PM)-based approaches in RCC.

In addition, extensive crosstalk between VHL/HIF and PI3K/AKT pathways results in the aberrant activation of PI3/AKT pathway, which contributes to the pathogenesis of RCC (25). In that context, genetic alteration of the PI3K pathway was identified in RCC using a largescale integrated analysis (26). Loss of VHL activation/function with concomitant upregulation of HIF activation facilitates the expression of several growth factors, including VEGF, epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) which activate the downstream PI3K/AKT pathway through their respective membrane bound growth factor receptor. Consistent with that, constitutive activation of PI3K and its downstream components, have been observed in ccRCC (27, 28). This also aligns with the observed activation of downstream AKT, as specified by high phosphorylation levels of AKT and AKT substrates in ccRCC (27, 28). In addition, high frequencies of gene mutations or deletions of *PBRM1* (36%), *SETD2* (15%), *BAP1* (13%), and *KDM5C* (7%) have also been identified in the ccRCC (29). These genes are essential for chromatin remodeling leading to genomic chaos commonly known as 'chromosomal instability' (CIN), a hall mark of RCC and other cancers (30, 31). In that context, PI3K inhibitor TGX221 has been shown to inhibit the growth of RCC cell lines containing VHL and SETD2 mutations suggesting that VHL/HIF and PI3K/AKT pathways may have a role in the deregulation of chromatin remodeling and CIN, involved with the pathogenesis of ccRCC (32). Considering that small-molecule inhibitors such as sorafenib, sunitinib, pazopanib, and axitinib that target VEGFR, and a new generation of inhibitors (such as brivanib, cabozantinib, cediranib, dovitinib, foretinib, lenvatinib, linifanib, nintedanib, regorafenib, tivozanib, vandetanib, and aflibercept) that not only target the signaling through activated VEGFR but additional targets such as platelet derived growth receptor (PDGFR), fibroblast growth factor receptor (FGFR), hepatocyte growth factor (HGF), and its receptor cMET (33), which are currently being tested in clinical trials for ccRCC patients, which can be contemplated to target the PI3K/AKT pathway in RCC, with the goal to inhibit RCC progression. Hence, developing approaches that will target multiple genotypes in a specific group of patients is essential to integrate PM into standard care of patients. This review describes some of the recent progress made in the field of precision medicine (PM)-based approaches in RCC.

Diagrammatic representation of loss of VHL activation/function resulting in constitutive activation of the downstream PI3K/AKT and other pathways is shown in **Figure 1**.

WHAT IS PRECISION MEDICINE (PM)?

Precision medicine is driven by patient data and refers to the identification of unique patient characteristics, whether genetic, molecular, pathological, or lifestyle, that are recorded and can be selected and used to tailor a targeted treatment protocol for a patient or a group of patients (34). The concept of PM has been used for nearly a century where transfusion patients and their donors were matched for blood or tissue type for transfusion or transplant based on their health records. With the advent of the Human Genome Sequencing Project (HGSP) in 2001, genome sequencing was first made available to clinical practice for the treatment of rare diseases (35). This led to the approval of several gene therapies commonly designed for a group of patients susceptible to a specific genetic disposition. A notable example of this category of patients undergoing treatment currently is the one identified with BRCA mutations for breast and ovarian cancer, human epidermal growth factor receptor 2 (HER2), and progesterone receptor (PR) for breast cancer. These discoveries single-handedly have led to increase in the overall survival and reduced the risk of death by 20% in these patient cohorts (36). In the area of chronic myeloid leukemia (CML), the knowledge that BCR-ABL is the genetic mutation that drives the disease in most patients with CML, which led to the development of a targeted agent that improved survival outcomes in patients (37). In addition, the identification of somatic mutation in the gene encoding the serine–threonine protein kinase B-RAF(B-RAF) in the majority of melanoma cases has provided an opportunity to treat these patients successfully with B-RAF inhibitors (38). Similarly, identification of inherited familial colorectal cancer (CRC) syndromes, such as familial adenomatous polyposis (FAP) and Lynch syndrome [hereditary non-polyposis colorectal cancer (HNPCC)], has led to significant understanding of the molecular pathogenesis underlying sporadic CRC spread and designing of appropriate treatment protocols (39). These discoveries indicate that the PM-oriented development of targeted treatment strategies can realistically benefit many patient groups.

Recent advancements in multi-omics technologies have further accelerated PM-based treatment protocols. Several HGS-like projects are currently in progress across the world by various laboratories and consortiums that may help to strengthen the field of PM (35). A new set of diagnostic assays known as *in vitro* diagnostic companion and/or complementary diagnostics

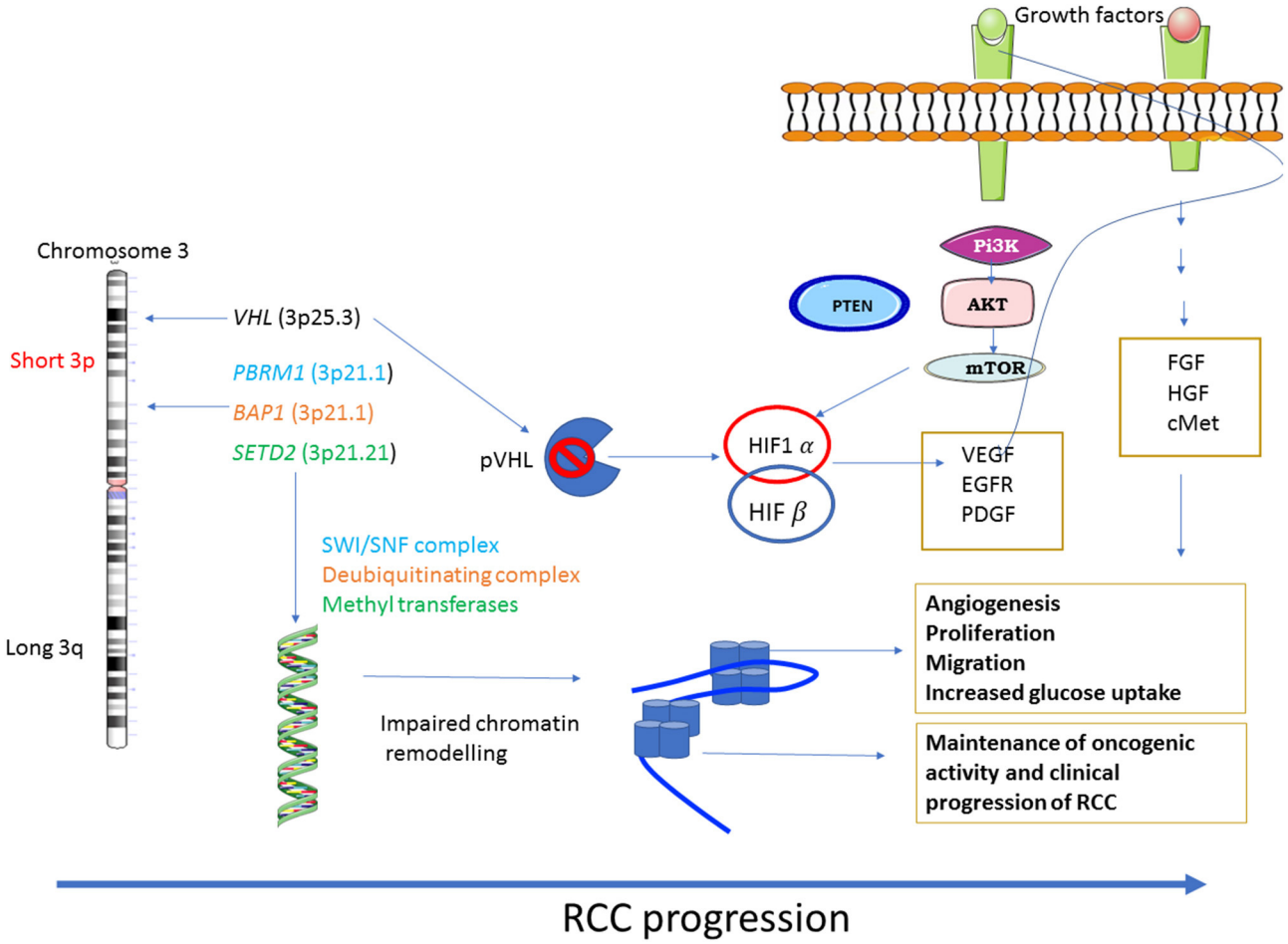

FIGURE 1 | Major dysregulated pathways in RCC: RCC shows a diverse range of genetic mutations. Loss of chromosome 3p tumor suppressor genes play a major role in the pathogenesis of RCC. Genes mostly affected are *VHL*, the gene responsible for sensing oxygen levels within a cell, chromatin remodeling genes such as *PBRM1*, *BAP1* and *SETD2*. The other signaling pathways that are associated with RCC progression are PI3K-AKT-mTOR and the pathways regulated by FGF, HGF and its receptor c-MET.

(CDx) is gaining more popularity in recent times. As per the United States FDA, European Regulation (EU) 2017/2016, and Australian Department of Therapeutic Goods Administration (TGA), PM is defined as a test that measures the level of genes, proteins, or mutations that aids in the benefit-risk decision-making about the use of a therapeutic drug, where the difference in benefit-risk is clinically meaningful (40, 41). There are already FDA-approved CDx assays available in the United States, especially for breast cancer, non-small cell lung cancer, melanoma, and CRC. Few examples of CDx assays are the THXID BRAF Kit which qualitatively detects the presence of BRAF mutations in patients with metastatic melanoma by polymerase chain reaction (PCR) (https://www.biomerieux-diagnostics.com/thxidr-braf), Herceptin, HER2 PharmDx Kit based on immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH) platforms, which determine the overexpression of the HER2 protein and gene in patients with breast cancer (42).

Over the past few years, the paradigm of mRCC treatment and care has changed drastically. Recently, a lot of effort was spent in integrating molecular targets with histopathology and cancer biology into RCC classification. In 2016, MiT family translocation/Xp11 translocation, fumarate hydratase deficiency, and succinate dehydrogenase deficiency were updated by the WHO after molecular and histopathological reclassification of RCC (9, 43), suggesting that the traditional morphological tumor typing is being replaced by evolving molecular tumor subtyping in RCC. To date, no molecular biomarker with prognostic or predictive value has been approved or is in practice in clinics for RCC; the prognostic stratification of patients with mRCC is still based on clinical factors like low levels of hemoglobin, high serum calcium, increased neutrophil and platelet count, and time from diagnosis to treatment (44). The development of multiple therapeutic approaches focused on different molecular targets in the tumor caters to patients with different subtypes of mRCC and requires the identification of robust predictive biomarkers that

will help to classify patients in a way that clinicians can stratify for the right treatment.

IDENTIFICATION OF PREDICTIVE BIOMARKERS IN THE ERA OF PM

The partial success of VEGF-TKI-based targeted therapies in RCC has proved the importance of understanding tumor biology at the molecular level for targeted treatment (45). The introduction of immune checkpoint inhibitor (ICI) has resulted in a paradigm shift in the treatment of RCC patients. In 2015, two different large phase III trials Checkmate 025 (nivolumab vs. everolimus) and Checkmate 214 (ipilimumab plus nivolumab vs. sunitinib) have pushed mRCC into the ICI era (46, 47). Thereafter, a substantial number of trials combining ICIs and angiogenic inhibitors have revolutionized the clinical practice in mRCC (48, 49). As encouraging as it sounds, a significant subset of patients has shown to remain non-responsive (inherent resistance) or stop responding (acquired resistance) to these life-changing treatment options either as single or in combination therapy. As RCC is a heterogeneous cancer both histologically and clinically, where the tumor ranges from a benign to clinically indolent to a most aggressive phenotype with a vast potential to metastasize, identification of predictive and prognostic biomarkers may provide a platform to stratify patients.

Histopathological Evaluation

Tumor morphology in RCC can be easily underestimated because of the presence of intratumor heterogeneity. This can hinder PM approaches and lead to therapy failure. RCC has been known to show histological heterogeneity across its breadth and the diversity that can be seen in architectural patterns and cytological features. The diversity of clinical behaviors in RCC may occur partly because of pathological and histological variations. A cohort study evaluating the association of ccRCC and nccRCC like papillary and chromophobe RCC with survival showed that, depending on histological subtypes in mRCC, the sites of metastasis differed. The sites of metastasis were also associated with the survival of patients in all histological subtypes (50). Similarly, a significant variation exists within the subtypes of RCC. The phase III trial of sunitinib compared with interferon which pushed sunitinib on to the horizon of systemic therapies available for RCC patients showed that approximately 25% of patients did not respond to sunitinib despite the presence of the clear cell tumor histology. A European clinical trial concluded that the phenotypic heterogeneity seen during the treatment resembles the genotypic and transcriptomic diversity in RCC. Another group identified variation in the treatment responses between subsets of metastases within same patients (51, 52). RCC also exhibits distinct cytological variations, like sarcomatoid and rhabdoid features that are associated with high grade RCCs. Rhabdoid and sarcomatoid features in ccRCC are associated with worse prognosis and poor survival (53, 54). A recent study identified 9 distinct RCC tumor patterns including compact small nests, large nests, bleeding follicles, alveolar, papillary and pseudopapillary, thick trabecular/insular, solid sheet, microcystic, tubular/acinar (55). The RCC subtypes differ in terms of their immune microenvironment for e.g., ccRCC are immune cell rich tumors that respond well to immunotherapy; whereas chRCC and pRCC are immune cold (poor immune cell infiltration), tumors that respond poorly to immunotherapy.

Many studies have shown that the histological pattern in a tumor closely resonates with the molecular features within the tumor (56). For example, ccRCC tumor cells are filled with lipids and glycogen that represents the faulty metabolism associated with fatty acid and glucose breakdown in these tumors. The pathways are altered due to the uncontrolled function of HIF gene that results in the mitochondrial dysfunction that subsequently redirects glucose and glutamine towards glycogen and lipid metabolism (57). For a successful administration of PM to patients an integration of genetical, morphological and molecular data is needed. The role of pathologists has evolved in this era of PM from just identifying and classifying tumors to playing an increasingly involved role in the clinical management of patients. Due to rapid development of technologies and range of different tests undertaken by the pathologists they are able to provide personalized and integrated information to the clinicians who can then provide tailored therapies to their patients (58).

Histological Biomarkers

Several histological biomarkers (pathological stage, nuclear grade, histology variant, etc.) have been studied across different pathological spectra of RCC. Morphologically, ccRCC is the most common subtype of RCC. However, not all cases of ccRCC have conventional clear cell features with a nest of large uniform cells having clear cytoplasm. High-grade ccRCC tumors do not retain the conventional ccRCC morphology and may contain eosinophilic cytoplasm and papillary or pseudopapillary formation (59). Non-clear cell RCC like papillary RCC contains tumor cells forming finger-like projections called papillae and tubules. Chromophobe RCC consists of cells with atypical nuclei along with granular cells in a solid growth pattern. The remaining uncommon types of RCC show aggressive clinical behavior and a poor prognosis. They are classified according to their unique features; for example, medullary RCC is associated with sickle cell trait, but low-grade oncocytic RCC affects pediatric neuroblastoma survivors (60). In addition, Xp11 translocation renal cell carcinoma, in which the transcription factor gene (TFE3) located on chromosome Xp11.2 is fused by translocation to proline-rich mitotic checkpoint control factor (PRCC) or disheveled segment polarity protein 2 (DVL2), is common in pediatric patients with RCC (61). Most importantly, rhabdoid and sarcomatoid differentiations are features more commonly associated with ccRCC and are associated with worst prognosis (62–64); the higher the clear cell component of a patient's tumor, the greater the chances of the patient benefitting from anti-VEGF therapy (65). Hypoxia-related HIF1α is implicated in the development of RCC (66). In an immunohistochemistry study, complete or partial response to sunitinib was correlated with the expression of HIF1α. Longer progression-free survival (PFS) was associated with lower levels of HIF1α levels (67).

Enhanced expression of carbonic anhydrase IX (CA9) and C-X-C chemokine receptor type 4 (CXCR4) has shown promise as biomarkers that can predict the response in patients treated with angiogenic inhibitors. While higher CA9 expression predicted longer PFS in patients with sorafenib treatment, higher CXCR4 predicted poor outcomes in patients treated with sunitinib (68, 69).

Genomic Biomarkers

The most studied genetic event in RCC is the loss of chromosome 3p that leads to mutations in VHL, polybromo1 (PBRM1), BRCA1-associated protein 1 (BAP1), and set domain containing 2 (SETD2) affecting 90% of ccRCC cases (70). Most biomarker studies have been conducted on VHL, making it the most studied biomarker (71). Although it is the most widely studied biomarker, there is no stable relationship derived between aberrant VHL gene expression and patient outcomes (72). Studies on the predictive and prognostic relationships of VHL mutations in patients treated with anti-VEGF therapy have shown that there is no correlation between abnormal VHL gene expression and the patient's PFS or OS (73). Similar attempts to find a positive correlation between VHL gene expression and patient responses to anti-VEGF treatments showed that, for mutations of VHL, loss of function to be an independent prognostic marker linked to improved response rate, no positive correlation with PFS and OS were reported (74, 75). The presence of PBRM1 or BAF180 tumor suppressor genes encoded at a gene locus near VHL was associated with patients who could be treated for a longer duration with anti-VEGF therapy (76). PBRM1 truncating mutation has been studied much more extensively as a potential biomarker in ICI treatment and has been positively associated with aggressive clinical behavior (77). A pan-cancer study on PBRM1 mutations revealed an association of PBRM1 mutation with OS (HR: 1.24, $p = 0.47$) in 189 patients with mRCC treated with ICI (78). In another independent cohort of the randomized CheckMate 025 trial, the investigators found PBRM1 mutation to be associated with clinical benefits in patients treated with nivolumab (79). Multiple studies found that patients harboring a functional somatic mutation in the BAP1 gene that binds to the BRCA1 and acts as a tumor suppressor gene did not respond well to everolimus and sunitinib treatments as compared to patients with wild-type BAP1 (80, 81). SETD2, which codes for a methyl transferase and is a tumor suppressor protein, has not been associated with significant differences in PFS in patients treated with anti-VEGF therapy (82). However, mutations in the telomerase reverse transcriptase (TERT) promoter region were associated with no clinical benefit in patients (83).

As discussed above, a tremendous amount of effort was spent in identifying putative RCC-specific biomarkers, which can be used as a podium to monitor drug responses. However, there is a serious lack of validated biomarkers for response to mRCC treatments in clinical trials as well as in clinical practice. Proper identification and validation of a robust biomarker panel in addition to the currently used IMDC (International Metastatic RCC Database Consortium) clinical scores are needed to further the goal of PM in RCC (84).

INTEGRATED OMICS EMPOWERS PRECISION HEALTHCARE

RCC is a heterogeneous group of diseases, with each subtype having unique and complex biology. To study, analyze, and infer useful insights from tumor biology, it is important to combine and integrate powerful high-throughput tools and available techniques. One such tool is "OMICS," a multidisciplinary platform that analyzes the interaction and function of biological information obtained from various sets of molecules in an organism such as DNA, protein, lipid, and metabolite. OMICS approaches that integrate and combine data across multiple platforms in a sequential manner to study the interplay of biomolecules in a cancer, in combination with clinical information, can endow clinicians with valuable data to stratify a treatment strategy for an individual patient at a personalized level. This concept is particularly important for non-responsive patients where standard treatments are ineffective, and some extra help in the form of combination therapies are needed to design "patient-specific" protocols that may enable success in better management of those patients.

Genome Sequencing Technology

The first step in understanding any cancer is to investigate the genetic code and underlying DNA sequence. Hence, as described earlier, differentially and/or unique gene analysis of the patient's samples have been used as an established platform as a genome-based PM in designing suitable therapy for these patients (85). The well-established landmark achievement in genome sequencing technology is the Human Genome Sequencing Project (HGSP). The discovery of 20,500 genes in the normal human genome has paved the way for the development of PM in diseased patients (86). Sanger sequencing and bacterial artificial chromosome techniques were one of the earliest available sequencing techniques (87, 88). However, the more user-friendly next-generation sequencing (NGS) technique that provided a cheaper and effective platform of high-throughput sequencing was later introduced (89). The earliest sequencing data that came out for RCC in 2009 established a higher frequency of mutations in chromatin remodeling genes like lysine demethylase 6A (KDM6A) and SETD2, KDM5C or lysine demethylase 5C and KMT2D or MLL2, and lysine methyltransferase 2D (90). In the following years, PBRM1 and BAP1, genes involved in chromatin remodeling, were discovered to be important drivers of ccRCC by whole genome exome sequencing (WGES) (91, 92). Similarly, in nccRCC, NGS allowed the discovery of unique mutations and somatic copy number alterations (SCNA) in MET proto-oncogene (MET), SETD2, and Neurofibromin 2 (NF2) in pRCC and tumor suppressor genes TP53 and PTEN in chRCC (93, 94).

Intermediate OMIC Levels: Transcriptomics, Proteomics, and Metabolomics

Although studying the genomics of a disease is an important step in understanding the pathogenesis and progression of the disease, the instruction contained in the gene is transcribed

as a functional protein. Gene readouts are commonly known as gene transcripts (messenger RNA/mRNA), and the study of total mRNA is known as transcriptomics. RNA sequencing and microarray are two important tools that allow researchers to know which gene is active and to determine the amount of gene expression in cancer cells. Knowledge of the presence of a gene and its activity aids in understanding the active and inactive pathways in cancer types. Both tools have high-throughput capabilities; however, microarray has proven to be a cost-effective method (95).

Transcriptomic tools have been able to profile RCC subtypes in the past 3 decades. A microarray is a powerful tool that is used to distinguish between clinical subtypes using a very small amount of sample, like in the case of core biopsies (96). Very recently, microarray profiling carried out on liquid biopsies (circulating molecules) is fast becoming an attractive and non-invasive approach for RCC diagnosis and progression of the disease (97). RNA sequencing is conducted to understand signature patterns that provide important data on response to treatments and survival of tumor cells in the heterogeneous RCC tumor microenvironment (TME). Recently, transcriptomic signatures of peritumoral adipose tissues showed that the RCC TME varies depending on the body mass index. This study indicates that the survival impact of obese patients with ccRCC may differ compared to normal-weight patients (98).

Proteomics facilitates the global study of protein profiles across organisms, tissues, or cellular organizations (99). Genomic and transcriptomic profiling provide an understanding of an altered gene sequence or mutation that may be present in tumor cells, but integrating proteomics is crucial as it provides vital information about the functional effect of a particular mutated gene sequence. More importantly, proteins undergo post-translational modification (PTM) after they are translated by ribosomes from the messenger RNA (mRNA) to form mature proteins. However, transcriptomics do not offer comprehensive insights into PTMs that affect the functions and interactions with the cell surrounding (100). The field of proteomics can help in unraveling the alterations in the proteome that are more likely to mirror tumorigenesis and the TME, which is important in PM in order to diagnose and predict diseases. The majority of targets in mRCC therapy are proteins, and conducting protein analysis will result in finding therapeutic targets that have the potential of direct clinical translational capabilities (101).

Two-dimensional electrophoresis was one of the earliest tools used for protein profiling. Multiple studies in the past have leveraged this technique to identify several dysregulated proteins in RCC (102, 103). However, there are several disadvantages of this technique, including the issue of reproducibility. It is also important to mention the low-throughput nature of this technique, which made it less favorable for use as a tool for protein profiling (104). Techniques like flow cytometry or immunohistochemistry have also been used to detect the expression of dysregulated proteins, but over the past two decades, mass spectrometry (MS) has been employed to assess the proteome in RCC (105). The glycolytic enzyme phosphoenolpyruvate carboxykinase 1 (PCK1) and small nuclear ribonucleoprotein polypeptide F (SNRPF) were shown to be significantly dysregulated in ccRCC by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The reliability of LC-MS/MS was validated in ccRCC samples by western blotting (106). The world of proteomics uses two crucial strategies to produce proteomic data, bottom-up proteomics, also called shotgun proteomics (useful for analyzing a mixture of protein) and top-down proteomics (total protein as a start sample) (107, 108). The shotgun method enables to generate a protein fingerprint of individual patients and is an approach suitable for PM, enabling the identification of key biomarkers in patients with RCC (107). A large-scale study identified 596 proteins that were variably expressed in ccRCC in comparison with a normal adjacent kidney tissue. They were also able to validate two proteins, Coronin-1A (CORO1A) and Perilipin (ADFP), that were found to be differentially expressed in ccRCC tissues using immunohistochemistry. Interestingly, while validating, they found that CORO1A was overexpressed in infiltrating lymphocytes and not in tumor cells (109). Middle-down proteomics, which takes advantage of both the shotgun and top-down techniques and uses partial protein digestion to characterize coexisting PTMs, is slowly starting to gain popularity in the field of proteomics. It is emerging as a promising tool for PM as it can explore potential biomarkers by quantifying the expression of a larger number of proteins along with the analysis of individual protein modifications (110). To date, to the best of the authors' knowledge, no study has been conducted on RCC using this new approach.

Studying genetic makeup is an entry point in omics, and the phenotype is the final physical makeup of an organism. Metabolites, along with genetic constitution and cellular microenvironment, constitute the closest reflection of a phenotype of a tumor (111). The metabolome represents all low molecular intermediates or compounds that are products of a metabolic reaction in a cell and, hence, the closest representation of the microenvironment. Studying the metabolome will help in acquiring a better understanding of the cellular process. Metabolomics is a new omics platform when compared to the other omics disciplines (112). It can strengthen PM by being able to predict drug response and safety in a patient (113). Metabolomics is one of the fastest growing platforms and has an upper hand over all the other omics because metabolites are small low molecular-weight substances that can easily be secreted in bio fluids like blood or urine and, hence, can be measured non-invasively (114). Not long ago, nuclear magnetic resonance (NMR) spectroscopy was conducted to analyze metabolites (115). A single RCC study investigated the urinary metabolome profile of patients with ccRCC before and after nephrectomy by NMR. It showed that the levels of creatine, lactate, alanine, and pyruvate were increased, and that the levels of citrate, hippurate, and betaine were decreased in patients with ccRCC in comparison to healthy subjects (116). Another study used the NMR platform and showed that the serum from patients with RCC had elevated levels of very low-density lipids, isoleucine, leucine alanine, N-acetyl glycoproteins, pyruvate, glycerol, and unsaturated lipids along with lower levels of glucose, glutamine, and acetoacetate before nephrectomy and that, interestingly, the pattern was reversed after nephrectomy (117). Recently, a study used a

biomarker-based cluster using NMR-based serum metabolomics and self-organizing maps to create an artificial neural network to predict early RCC diagnosis. The study proposed a cluster of 7 metabolites namely alanine, creatinine, choline, isoleucine, leucine, lactate, and valine to validate the metabolomics changes in patients with RCC before and after nephrectomy (118). MS is another data acquisition platform widely utilized to generate metabolomics patterns from biological samples (119). Although an increasingly large number of studies that have used metabolomics to find potential biomarkers are available, there is a high number of studies that have shown an excessive number of false-positive rates as being the greatest challenge in the world of metabolomics (120). However, metabolomics can become a powerful tool in PM when used in conjunction with another available omics platform.

A multi-omics approach integrating various platforms is essential to bridge the wide gap in understanding RCC tumorigenesis that is driven by oncogenes, rewired metabolic pathways, and altered signaling cascades, all leading to a dynamic and, at the same time, defective phenotype. Many studies have highlighted the importance of taking an integrative approach to explore the intertwined association of various biomolecules and its effect on cancer biology, which is the key link in bringing PM approaches to the clinic (121, 122). There are many computational tools now available to integrate different branches of omics, e.g., Metabox (an R-based web application), O-miner, and Galaxy (123–125). From a PM perspective, integrative approaches will help to understand the biology of RCC holistically, thereby improving the ability to predict an early diagnosis as well as drug response in patients.

LIFESTYLE-BASED DATA

Hereditary RCC constitutes approximately 2–5% of all RCCs (126). The remaining RCCs are influenced mainly by lifestyle and environmental factors (127). Cigarette smoking, obesity, and hypertension are established risk factors that are associated with RCC. On the other hand, physical activity and alcohol consumption are known protective factors against RCC (128). Combining lifestyle data with integrated omics data along with clinical and diagnostic information will help scientists generate patterns identifying the risk of developing RCC and predict the disease earlier and will help in determining the most effective intervention against RCC.

MANAGEMENT AND APPLICATION OF COMPLEX OMICS DATA

Acquisition of omics data using high-throughput technologies has allowed the generation of huge amounts of data often in terabytes and petabytes. Notably, The Cancer Genome Atlas (TCGA) contains petabytes of genomics, epigenomics, transcriptomics, and proteomics data (129). To put that into perspective, 1 petabyte is equivalent 223,000 DVDs, each storing 4.7 Gb together. It is estimated that, by 2025, 60 million genomes will be sequenced (130). With huge amounts of data collection come the massive challenge of sorting and storing appropriately. The gigantic amount of data must be stored appropriately so that it is easily accessible by scientists and clinicians for it to be utilized efficiently for clinical diagnosis and treatment. Like the TCGA, there are other publicly available oncology data sets pertaining to patient-based multi-omics data sets (**Table 1**). The TCGA contains information on 941 renal cancer patient samples (126). The TCGA and other public domain databases aid investigators in combing the vast and diversified omics information into a well-annotated structured data set that may contribute toward stratifying PM for patient groups. There is a range of publications available on RCC that have used these databases to analyze the clinical and biological parameters of patients with the aim of prioritizing structured treatment (131–133).

The key challenge in using these multi-omics approaches is to understand, interpret, apply, and translate the knowledge generated from the huge omics data. According to the International Medical Informatics Association (IMIA), bioinformatics techniques and algorithms are used to analyze and extract the hidden knowledge in the diverse and complex "Big Data" (134). Although harnessing Big Data related to patients is essential in making the PM approach in oncology a success, it can come with its share of adversities in patients. Collection of Big Data about patients from different sources and networks may, in certain cases, result in data theft and misuse, leading to uncertainty, social discrimination, and biases for patients, which again can result in undesirable health effects for patients.

NANOTECHNOLOGY-BASED APPROACHES IN PM FOR RCC

One of the challenges in cancer therapy and more so in RCC is the broad non-specific target-based treatment approach, which not only causes severe side effects in patients but also in most cases enhances drug resistance by enhancing the survival of chemotherapy-treated residual tumor cells into an aggressive phenotype. The advantage in the application of nanotechnology is the use of "nanocarriers" as drug delivery vehicles that mediate targeted delivery of drugs to tumor sites without causing much harm to normal tissues (135). However, the technology is still in infancy and, currently, there are a few clinical applications of this technology in cancer treatment [127, 130–133]. Nonetheless, several *in vitro* and *in vivo* mouse model studies have demonstrated that the technology has the potential to have a significant impact on clinics.

Nanocarriers, usually having a size of less than 100 nm in one dimension, are organic/inorganic or hybrid particles shaped in the form of micelles, dendrimers, liposomes, or virus-like particles and are used to encapsulate/covalently conjugate or absorb cancer drugs. These nanocarriers have shown significantly greater efficacy in different cancer models compared to drugs on its own (136). Nanoparticle formulated (nanocarriers) drugs allow for specific targeting of tumor cells by providing superior solubility and stability of drugs in the tumor microenvironment, resulting in improved internalization of the drugs in the tumor

TABLE 1 | Description of multi-omics data repositories related to cancer.

Data repository	Short form	Data links	Description
International cancer genomics ***Whole genome analysis***	ICGC	ICGC DATA PORTAL	The repository is a global initiative that provides user- friendly platform for visualizing, querying, and downloading cancer data
The cancer genome atlas ***RNA sequencing, DNA sequencing, single nucleotide variant, copy number variation***	TCGA	TCGA DATA PORTAL	A cancer genomics program spanning 33 cancer types and >20,000 primary cancer and matched normal samples
Clinical proteomic tumor analysis consortium ***Mass spectrometry-based proteomics data***	CPTAC	CPTAC DATA PORTAL	A proteogenomic Cancer Atlas of comprehensive sequence of proteomic datasets
Cancer cell line encyclopedia ***Gene expression, mRNA expression SNP genotyping, pharmacological profiles of 24 anticancer drugs***	CCLE	CCLE DATA PORTAL	The project validates >1000 human cancer cell line models by detailed genetic and pharmacological characterization
Therapeutically applicable research to generate effective treatments ***Gene expressions, copy number, miRNA expressions***	TARGET	TARGET DATA LINK TARGET DATA MATRIX	TARGET applies a comprehensive genomic approach to determine molecular changes that drive childhood cancer
Cancer genome characterization initiative ***Gene expressions, copy number, sequencing data***	CGCI	CGC1DATA PORTAL CGC1DATA MATRIX	CGCI uses molecular characterization to uncover distinct features of rare cancer
Omics discovery index ***Genomics, transcriptomics, proteomics, and metabolomics***	OmicsDI	OMICSDI	The tool provides a framework across heterogenous omics datasets
Molecular taxonomy of breast cancer international consortium ***Single nucleotide polymorphisms, Gene expressions, computational biology***	METABRIC	MOBCCRC	Breast cancer PM and Computational Cancer Biology programs incorporates multidisciplinary techniques to develop statistical models to understand genomic abnormalities

without much loss in the circulation, circumventing harmful side effects in patients. One great advantage of nanoparticle-formulated drugs is that they cannot pass the tight junctions of the normal vascular lining but can easily pass through the leaky vascular lining of tumors to enhance their concentration at the tumor site (137). This phenomenon known as "enhanced permeability and retention effect (EPR)" is the fundamental principle of nanoparticle-conjugated drug treatment of tumors (138). However, in that scenario, the surface area and size of nanocarriers play an important role in active tumor targeting and are adjusted for EPR effects to occur without unwanted uptake of nanodrugs by the normal endothelial system (138). Common examples of these formulations are Abraxane (albumin conjugated paclitaxel), PEGylated (polyethylene glycol formulated), doxorubicin (DOX), and DOXIL recently approved by FDA for cancer treatments (139, 140). These nanodrugs have shown enhanced efficacy in patients with less cardiotoxicity compared to conventional drugs. In addition, active targeting of tumors by nanoparticle-conjugated drugs also occurs by physically attaching the surface of nanocarriers to certain overexpressed antigens on the surface of tumor cells. Tumor-specific cell surface overexpressed antigens with nanoparticle-formulated antibody conjugation have recently been shown to target folate receptor overexpressing prostate, breast, and lung cancer cells *in vitro* (141). Moreover, the application of nanoparticle-conjugated epidermal growth factor receptor (EGFR) and HER2 receptors on different cancers leading to increased efficacy of different cancer drugs have recently been demonstrated (142, 143). Besides these, the cluster of differentiation (CD), estrogen, integrin, and other growth factor receptor-based targeting using different nanoparticles have successfully been shown to enhance the efficacy of current cancer drugs *in vitro* and *in vivo* mouse models (144–146). In cases where there is poor penetration of antibodies inside cells, antigen fragments (Fab) or single fragments (scFv) are also used to overcome these deficiencies.

As such, nanomedicine has the potential to empower PM in oncology, as the key step is to select the right drug delivery platform for a patient cohort that can target specifically overexpressed tumor-specific protein sets aberrantly regulated by cancer. Human kidneys are an ultrafiltration unit and are responsible for filtering the circulating blood. In that case, the shape, size, and charge of nanoparticles are important factors when designing nanocarriers for RCC. In that scenario, cationic spherical nanoparticles with diameters between 6 and 8 nm have been shown to have better renal clearance (147, 148). Pre-clinical studies focusing on RCC have concentrated on selectively targeting kidney tumors (149, 150). Tumor hypoxia is a leading cause of drug resistance in RCC (151). Recently, it has been shown that certain nanoparticles can be activated under hypoxia-induced oxygen stress. Among those, hypoxia receptive electron acceptor nitroimidazole conjugated with carboxymethyl dextran and loaded with DOX showed accelerated release of DOX to hypoxic tissues *via* the EPR

effect (151, 152). In addition, iron oxide conjugated DOX linked with azobenzene-4, 4-dicarboxylic acid released 80% of DOX in hypoxia compared to only 10% release in normoxic tissues (153). There are nanocarriers that have demonstrated the application for targeting hypoxia in tumors, which may have clinical application for RCC (154–156). In addition, few gold, silver, silica, and iron-based nanoparticles including liposomes have shown applicability in targeting angiogenesis, which is a major decisive factor in RCC progression and resistance (157). In that context, tumor necrosis factor (TNF)-mediated secretion of vascular endothelial growth factor (VEGF) was decreased by curcumin-loaded PLGA nanoparticles (158). In a hepatocellular cancer model, chitosan nanoparticles suppressed the expression of VEGFR2 with subsequent suppression of VEGF synthesis, leading to an anti-tumor effect in the mouse model (159). These and other nanoparticle-based studies targeting hypoxia and angiogenesis hold a great promise in increasing the efficacy of current drugs used for RCC treatment. However, significant challenges exist in modulating these anti-hypoxia and anti-angiogenesis nanomedicine-based approaches successfully *in vitro* and in mouse model studies with consequent application to patients with RCC.

Recently, nanoparticles have been used in RCC with an intention to reverse sunitinib resistance. Cuprous oxide nanoparticles (CONPs) can downregulate the expression of AXL, MET, AKT, and ERK to improve responsiveness to sunitinib in resistant RCC cells and may be a less toxic way to treat patients with acquired sunitinib resistance (160). Although few nanomedicines such as liposome-based Onivyde and Vyxeos have been approved to treat certain solid cancers (pancreatic, esophageal, and colorectal) (161), there are no FDA-approved nanoparticle-based drugs for RCC. Zinostatin stimalamer, a lipophilic analog of the antitumor antibiotic zinostatin, is the only approved nanomedicine-based drug available in Japan for RCC treatment (162, 163). CRLX101, a novel nanoparticle-drug conjugate containing camptothecin and inhibitor of topoisomerase I, HIF1, and HIF 2α was tested recently in a randomized phase II trial along with the anti-angiogenic drug bevacizumab vs. the standard of care for mRCC like bevacizumab, axitinib, everolimus, pazopanib, sorafenib, and sunitinib. Unfortunately, the combination failed to demonstrate any improvements in the PFS of patients with mRCC when compared to standard treatments (164).

ARTIFICIAL INTELLIGENCE IN PM

Artificial intelligence (AI) is an arm of computer science that makes use of computer-generated data to mimic human intelligence. RCC is a multi-faceted disease with many genetic and epigenetic variations, and AI algorithms can push forward personalized RCC detection and diagnosis in leaps and bounds toward the autonomous disease diagnosis field with the help of big data sets. Machine learning (ML) and deep learning (DL) are driving today's AI advancements. ML is an important type of AI that can learn from a whole heap of data and make predictions. DL is a subset of ML that uses an artificial neural network that mimics the human brain's information processing approach. A well-planned DL can diagnose and classify diseases and make predictions with high accuracy.

AI together with ML and DL can improve RCC diagnosis and treatment in digital healthcare. Most patients with RCC are diagnosed incidentally when scanning for other diseases. However, there is no sure way of predicting that renal masses are cancer by imaging alone. Tissue biopsy is a gold standard in many cancers; however, renal mass biopsy has significantly higher non-diagnostic rates and fails to aid in diagnosis (165, 166). Approximately 20% of small renal masses (<4 cm) are non-malignant and do not need surgeries but still end up undergoing partial or full nephrectomy (167). Differentiation between small renal masses and RCC is an important aspect of patient management, and this is where AI can play an important part in PM. AI algorithms can be used to accurately predict whether the renal mass shown in patient scans is cancer or not (168, 169). Many research studies have developed complex neural networking programs that can process digitized renal histopathology slides and learn patterns to identify tumors (170). AI has touched every aspect of the prognosis of patients with RCC right from diagnosing RCC to predicting prognosis and recurrence in patients (171–173). To accurately tailor-make treatments for patients with RCC, an important step is to accurately predict drug response in individual patients. Therapeutic resistance, both innate and acquired, is a financial burden in any disease. Thirty percent of patients with RCC are known to be innately resistant to the targeted therapy, and another 30% respond initially, develop resistance later, and show up with increased tumor burden (174). RCC researchers have very recently designed deep neural networks to predict tumor drug response (175). ML algorithms have perfectly predicted the chemoresistance of cancer cell lines (176). AI is already gaining momentum in clinical medicine, specifically in medical imaging. In early 2021, the FDA released the agency's first action plan named the "AI/ML-based software," which outlined the FDA's next steps toward the oversight of AI/ML-based medical action (177). Very recently, the FDA approved GI Genius, the first device that uses AI/ML to assist clinicians in real-time detection of polyps or tumors in the colon (178). Few AI algorithms like WRDensity, HealthMammo, Profound AI, and Transpara have already been approved in breast cancer that help radiologists and clinicians in identifying suspicious mammograms (179, 180).

The advantage of AI/ML-based programs is that they are continuously evolving with new data sets. Harnessing its true potential is important to achieve the goal of PM where it can help healthcare systems to automate tasks that are time-consuming and overwhelmingly difficult for physicians.

THE ROLE OF THE GUT MICROBIOME IN RCC PM

Over the past decade, considerable advances have occurred in understanding the implication of the gut microbiome in normal biology and how it changes with tumor initiation and progression through manipulation of the immune system (181). In the human body, extensive interaction exists between host cells and millions of symbiotically blossoming microbes.

These symbiotic microbes colonize in different organs of human bodies and undergo constant changes triggered by endogenous and exogenous stimuli. However, substantial portions of these "healthy microbes" remain unidentified by current microbiological techniques. Nonetheless, distinct "dysbiotic" microbiome signatures have been associated with cancer and other diseases (182). Changes in the dysbiotic gut microbiome occur with initiation of cancer in patients in response to surgery and chemotherapy treatments and are associated with cancer recurrence and contribute substantially to the efficacy of cancer therapies (183, 184). A plethora of integrated omics studies have emphasized the role of gut commensals in tumorigenesis and cancer treatment (185–187). Recent studies also suggest that host-microbiome signature can even correlate with survival parameters in patients with cancer, suggesting that the modulation of gut microbiome signature can potentially dictate the success of cancer therapy (188, 189). Hence, the gut microbiome signature is rapidly becoming a target for new treatments and diagnostic models.

In RCC, gut microflora are definitely an important non-genetic contributing factor to the progression of the disease, as microbial populations in the gut of patients with RCC have been found to be distinct from their adjacent normal tissues (190, 191). Gut microbiota can also modulate the tumor microenvironment (TME) by influencing the levels of various metabolites such as dietary amino acids and short-chain fatty acids, for example, butyrate, acetate, and propionate, which have an anti-inflammatory property (192). Immunotherapy treatment founded on ICI-based PD-L1 and PD-1 antibodies has changed the landscape of RCC treatment (193). However, primary resistance occurs in the majority of patients with RCC. Recently, studies on mice xenografts and on patients with cancer have shown that abnormal gut microbiome profiles can influence primary resistance to ICI treatment (194). Particularly, an antibiotic therapy prior to ICI treatment in patients with RCC significantly reduced the overall and progression-free survival compared to a cohort of patients who were not treated with the antibiotic (195). In univariate and multivariate analyses, antibiotic treatment was a single prognosticator of ICI treatment failure in these patients. Using quantitative metagenomics, the composition of the gut microbiome of patients participating in the study was analyzed. Positive clinical outcome in responding patients with RCC was associated with the enrichment of the *Akkermansia muciniphila* (*A. muciniphila*) microbe in the gut microbiome (177). Further studies are in progress in which fecal bacteria from patients responding to ICI treatment will be transferred to non-responding patients. Such treatments have been performed with success in other groups of patients with cancer (196). In addition, the microbiota of non-responding patients can be controlled by diet and use of prebiotics for a favorable outcome in patients in response to therapies. Hypothetically, if the prevalence of microbial population in non-responders can be changed toward the microbiome profile in responders, that would enable the non-responders to respond to ICI therapy (197).

In the same context, a shotgun DNA sequencing technique was used to screen the gut microbial composition of patients with advanced RCC (198). It was found that TKIs given to patients prior to ICI treatment shifted the gut microbiome profile of patients positively and enhanced the growth of immunostimulatory microbes like *Alistipes senegalensis* and *Akkermansia muciniphila*. Hence, harnessing the positive effect of gut microbiota in patients with RCC can enhance the clinical efficacy of ICI therapy (198). Few other studies have also emphasized the importance of modulating gut microflora to achieve successful outcomes while using ICIs on patients with mRCC (199, 200).

With respect to PM, microbiome-based treatment approaches are gaining momentum where individual host-microbiome patterns can be integrated with other personal health-related information of a patient for evaluation if the microbiome-based PM approach will be a suitable treatment option. As analyzing and storing of individual patient's microbiome profile will be challenging, AI in that scenario may play an important role in finding essential clinical evaluations, thereby promoting the field of PM to evolve. In that context, the application of AI in predicting chemotherapy resistance in patients with ovarian cancer based on gut microbiota profiles has recently been reported (201).

Although evaluation of the microbiome-cancer axis has started, a lot of heavy lifting is needed for translational possibilities. There is an extensive diversity of gut microbiomes in individuals depending on their genetic makeup, diet, and diverse lifestyle. Adding to that complexity, microbes interact with the host and the tumor in a diverse and context-specific manner. Hence, elucidation of microbiome composition and unique mechanisms by which the gut microbiome interact with the host and tumors may prove challenging for PM. However, a significant breakthrough in microbiome-based diagnostics and therapeutics is underway based on the identification of unique tumor-specific gut microbiome and/or metabolic signatures combined with existing immune cell types. This could potentially stratify patients into different risk groups and may facilitate the prediction of models for early-stage screening (202–204). In these scenarios, big data analysis using AI/ML algorithms may play significant roles in developing precise and reliable prediction, diagnostic, and therapeutic tools for cancer treatment.

CANCER VACCINES AND CELLULAR THERAPIES: PROMISING PM STRATEGIES IN RCC

Vaccines have been, for a very long time, thought to be a viable treatment for cancer. The first published literature on oncology vaccines dates to 1893 when an inoperable soft tissue sarcoma was targeted by injecting streptococcal toxins that enhanced a non-specific immune response in a patient (205). Many cancer vaccines are in clinical trials with the aim to enhance patient's immune response against tumor cells. Vaccine therapy can be implemented if a tumor is immunogenic, and RCC is a proven immunogenic tumor, as there is significant infiltration of lymphocytes in the tumors (206). A recent phase 1 clinical trial studied the safety and efficacy of autologous dendritic

cells transduced with AdGMCA9 (GM-granulocyte macrophage colony-stimulating factor + CAIX delivered by an adenoviral vector) and was found to be well tolerated without any significant safety concerns in patients with mRCC (207). Neoantigens or mutation-generated novel epitopes of self-antigens in tumor cells have recently been proven to have greater potential than tumor-associated antigens like CAIX. PM tools like NGS have made it possible to identify several neoepitopes in individual patient tumors that are potential targets to develop treatments. GEN-009 is a personalized neoantigen vaccine that is in the phase 1/2a trial (NCT0363310) for RCC and other solid tumors. Patients who received PD-1 based immunotherapies were selected for the NCT0363310 vaccine trial. GEN-009 was able to enhance T cell responses in 100% of evaluated patients. The trial emphasized that the approach of combining vaccines with ICI has a good therapeutic potential. GEN-009 consists of neoantigens that were identified based on a cell-based bioassay platform that uses patient's own monocyte-derived dendritic cells to identify neoepitopes (208). Although GEN-009 is still far from clinical use, the excellent results of phase 1/2a is a perfect example that vaccine-based PM strategies can be tailored toward specific patient groups or individuals.

Chimer antigen receptor (CAR)-T cells are genetically engineered autologous T cells to express CAR that can specifically recognize tumor cells. CAR-T cells, after their success with hematologic malignancies, have been proven to be an illustrative example highlighting the use of the PM platform to improve patient outcome. CAR-T cells are tailor-made for patients using their own white blood cells and, so far, are approved for treating leukemia and lymphoma. KymriahTM (relapsed or refractory acute lymphoblastic leukemia and large B-cell lymphoma), YescartaTM (lymphoma), and TecartusTM (relapsed or refractory mantle cell lymphoma) are the only three autologous CAR-T cell treatments that are currently approved by the FDA and have reached the market (209, 210). CAR-T cell therapy in the case of hematologic cancers has been particularly successful because it harnesses CD19, a unique protein only found in hematological malignancies and is not expressed by solid tumors. As RCC tumors express proteins that are also expressed by healthy kidneys, the potential application of CART-cell therapy may not be an option for patients with RCC until RCC-specific proteins are discovered that can be targeted by CART-cells.

Recently, human endogenous retrovirus E (HERV-E) derived antigen was found to be expressed in majority of ccRCC cells with no expression in normal healthy tissues. A phase I clinical trial based on this finding is currently actively recruiting patients to study the safety and efficacy of HLA-A11:01 restricted HERV-E specific CART-cells in patients with mRCC (NCT03354390). Another clinical trial currently recruiting patients is a phase 1/1b open-label multi-center study that will analyze the safety and efficacy of TRQ-1501 in patients with relapsed or refractory solid tumors like RCC (NCT03815682). TRQ-1501 is an immunotherapy developed from patient's own T-cells capable of targeting heterogeneous tumor antigens in addition to overcoming immunosuppression in TME. TRQ-1501 is also loaded with 1L-15, 1L-12, and TLR agonists whose prime function will be to activate the immune system. Another clinical trial that is worth mentioning is NCT03393936, an umbrella trial that is a phase 1 and 2 trial studying the safety and efficacy of two CART-cells, CCT301-38 and CCT301-59, in relapsed and refractory patients with stage IV mRCC. CCT301-38 will target AXL on tumor cells while CCT301-59 will target receptor tyrosine kinase-like orphan receptor 2 (ROR2) antigens in tumors. Patients in these trials are selected based on the expression of tumor antigens they express. Although all the above trials are in their infantry, they sure are a step ahead toward the PM era in RCC.

CIRCULATING TUMOR CELLS AND PATIENT-DERIVED TUMOR ORGANOIDS: THEIR RELEVANCE TO PM IN RCC

The trans-circulatory pathway is one of the most important pathways in RCC metastasis (211). A metastatic disease is the result of circulating tumor cells (CTCs), a rare subset of tumor-disseminated cells that are shed in the patient's blood. CTCs maintain the inherent primary tumor properties and can be detected very early as cancer progresses, making them a powerful clinical biomarker. RCC, being a highly invasive cancer, can benefit from the early detection of RCC-specific CTCs. The CellSearch circulating tumor cell test is the only FDA-approved CTC detection test used in clinics. Baseline CTC detection has been proven to be an important prognostic marker of PFS and is a significant predictor of poor response to tyrosine kinase inhibitors in patients with mRCC (212). CTCs can be a good candidate for a preclinical model because they retain the heterogeneity and properties of primary tumors. These models can then be tested for drug screening and disease modeling (213).

The potential PM application for RCC drug and biomarker screening could be generating patient-specific pre-clinical 3D-organoid models. Patient-derived 3D organoids have been developed from primary tumors and CTCs in various cancer types (214, 215). Primary tumor and CTC-derived explants have recently been studied in *in vivo* drug-resistant models. Analysis of the genetic makeup from resistant models and relapsed patients can provide a clearer picture about the development of resistance in cancer (216). A clinical study on treatment guided by patient-derived xenografts (PDXs) could identify an effective treatment regimen for 11 out of 12 patients with advanced cancer types (1 patient died before receiving treatment) (217). Thus, 3D organoids and PDXs have immense potential for a guided treatment regimen. Although the studies require further development, CTC, patient-derived 3D organoids and PDX are in use for the development of PM-based cancer therapies.

CURRENT PM-BASED TARGET THERAPIES AND IMMUNE CHECKPOINT INHIBITOR STRATEGIES IN RCC

Renal cell carcinoma, in terms of available treatment, has come a long way from being a therapeutic orphan to one with multiple treatment options. RCCs are innately resistant

to chemotherapy and radiotherapy. The exponential growth of research on RCC has led to the clinical knowledge that RCC has mutated pathways like the VHL pathway that sustains RCC cell growth by supporting angiogenesis, and the PI3K-Akt-mTOR pathway that supports the progression of the disease (24). Not surprisingly, targeting angiogenesis as well as the mTOR pathway with angiogenic inhibitors and mTOR inhibitors have shown tremendous improvement in the overall response of patients with RCC. However, as mentioned before, only selected patients can reap the benefits of these treatments. Another PM-based approach in RCC has been the use of monoclonal antibodies that block the immune checkpoints. RCC is an immunogenic tumor; tumor cells evade the immune system by overexpressing the immune regulators that keep autoimmunity and self-tolerance in check (218). PD-1 and PD-L1 are transmembrane immune and tumor cell regulatory proteins that modulate activation or inhibition of immune cells (219). PD-L1 tumor expression is a poor prognostic factor but a good response predictor for the use of both PD-1 and PD-L1 inhibitors in RCC (220). A recent meta-analysis study on the expression of PD-L1 has shown that patients with RCC harboring high expression of PD-L1 in tumors responded significantly better to both PD-1 and PD-L1 antibody therapies compared to patients with low or negative PD-L1 expression in tumor cells (221). Each PD-1/PD-L1 drug approved by the FDA takes into consideration PD-L1 expression based on an immunohistochemistry (IHC)-based tissue assay. Only a small fraction of patients with a negative PD-L1 expression by IHC assay showed any response to PD-1/PD-L1 antibody therapy. This suggests that the identification and utilization of PD-L1 expression in tumors is of great significance for a better selection of patients who may respond to PD-1/PD-L1 therapy.

More than dozens of immune checkpoints have been identified over the past decade, nivolumab, anti PD-1, was the first immune checkpoint inhibitor (ICI) that was approved by the FDA as a monotherapy for treating advanced RCC in 2015. Thereafter, in April 2018, a CTLA-4 inhibitor, ipilimumab, in combination with nivolumab, was approved by the FDA for intermediate and poor risk in previously untreated patients. Soon, anti-VEGF/ICI combination treatments appeared to produce favorable outcomes in patients with RCC. As a result, in 2019, two combinations, pembrolizumab (anti-PD-1) plus axitinib (tyrosine kinase inhibitor) and avelumab (anti-PD-L1) plus axitinib, made their way as FDA-approved drugs in mRCC treatment plans (193). As with any other treatment, the success of ICI is also very vague, so effective use of ICIs with respect to PM needs robust predictive biomarkers that can predict response to tumors. Biomarkers that have claimed to predict ICI responses include PD-L1 expression, tumor mutational/neoantigen burden, microenvironment signatures, and occurrence of immune-related adverse events (222). Recently, a study showed that the frequency of PD-$1^{+}CD8^{+}$ T cells relative to PD-1^{+} regulatory T cells or Tregs in the TME is a far superior biomarker that can predict the efficacy of anti-PD-1 treatments than PD-L1 expression and tumor mutation burden (223). Other studies that are gaining increased momentum in recent years have highlighted the success of re-challenging patients treated with ICI to explore its safety and efficacy (224). If ICI re-challenging studies translate to clinical settings, research on exploring the identification of potential biomarkers will be in high demand for the selection of the right group of patients.

REGULATIONS FOR THE IMPLEMENTATION OF PM

PM is an opportunity to treat individual patients with cancer at a greater targeted therapeutic resolution. With the rapid shift of cancer therapeutics toward the use of PM in patient care, there is a greater need of refined regulatory guidelines to ensure the best and safe PM-based patient care. Many countries like Australia have very recently introduced regulatory laws revolving around the use of PM, while countries like the United States and European countries modified their regulatory landscape to include regulatory laws pertaining to PM. Most regulatory approvals are recognized by the FDA and Centers for Medicare and Medicaid Services (CMS) in the United States and the European Medicines Agency in the European Union (225). Different centers under the FDA and CMS regulate the approval of PM as per their jurisdiction and regulations. For example, laboratory diagnostic tests are regulated by CMS. Similarly, in the European Union, the EU regulatory framework for pharmaceuticals offers several different legislations that regulate the development of PM. *In vitro* diagnostics and medical device legislations aim at adapting the EU legislation to the technological and scientific progress in the PM sector and provide a better consultation process for companion diagnostics, similar to clinical trial regulations that aim to simplify the conduct of clinical trials and research in therapies using PM (226, 227).

The regulatory scenario for PM is changing worldwide to accommodate its fast-changing landscape. In 2018, 42% of all new drug approvals by the FDA were personalized therapies (228). However, PM has multiple moving parts, and that not only involves PM-based drugs but different services, devices, and associated technologies that currently are regulated by multiple agencies, with no clear regulatory framework associated with PM treatment in patients. With the PM era pacing rapidly in the direction of modern medicine, there is an urgent need for associated regulatory boards to continually update and adapt to maintain the high quality and safety of PM-based treatments and associated products and technologies.

THE DOWNSIDE OF PM

PM is a young and rapidly growing field and, like any new scientific/therapeutic field, personnel involved are in uncharted territories. Although PM has the potential of touching a patient's life and be able to make a positive change to their health, it still must make a huge amount of progress in order to fill the gaps between the research, clinical trials, and clinics for successful therapeutic application in patients.

Privacy Concerns and Ethical Issues

One of the foremost issues is the amount of patient data required, generated, and stored for the proper implementation of PM. This can lead to privacy issues related to handling and storing the massive data sets. The use of technologically big data-generating tools like the NGS and OMICS platforms may worsen this issue further. To apply PM successfully, integration of NGS- or OMICS-generated data along with patient's other health records including diagnosis, laboratory works, and demographic information is needed. Patient data may differ depending on how they are handled, collected, and stored (229). Generation of error-free clinical data and their recording and interpretation are of prime importance. Digital health records and networks are used increasingly to ease the daunting task of storage of large volumes of patient information (230). Generation of large amounts of data like whole-genome sequencing (WES) of a patient may raise concerns related to privacy acts and discrimination. In addition, the big data sets generated by the NGS, WES, and OMICS platforms need treating physicians to have special interpretation skills. In that context, treating physicians will need adequate genomic knowledge and training for data interpretation, as these data sets will formulate prescribed treatments for patients. The issue may be exacerbated further by genomic findings related to "variants of unknown significance" or "false positives/negatives" that may make the interpretation difficult and flawed. In addition, redundancy in major biological pathways adds a further layer of complication in the interpretation of the bioinformatics of big data sets. These exercises may involve increased time management and, at times, may require multidisciplinary clinical/bioinformatics efforts to make a decisive treatment plan. According to a recent survey, the collection of huge data sets from patients potentially caused stress in clinicians, as it required a substantial amount of time for data interpretation and explanation to patients (231). In addition, patient information is sensitive and private and requires confidentiality and secured recording and storage. There are additional risks for patients to suffer from stigma and discrimination by insurance providers or employers if health records/data are made accessible or inappropriately disclosed. These can lead to serious ethical issues surrounding patients. For example, incidental detection of the presence of a life-threatening disease in an individual while genetic screening for one disease can have damaging impacts on the patient's physical and mental health, especially if there is no cure for the disease. To further complicate the issue, the rights of the patient's immediate family members or related members must also be considered in such situations. A genetic disease can be managed better with early intervention, and family members have the right to know if they are at any risk. However, it will mean encroaching patient's privacy and his/her right to disclose information to family members. Such an ethical dilemma is a huge concern for clinicians and patients practicing PM. In that context, the American Medical Association has formulated guidelines for physicians who require counseling patients about sharing the results of genetic tests with family members (232). The American Society of Human Genetics provides legislature to physicians to share the results of genetic tests that can pose health risks to family members if patients refuse to do so (233). However, this declared right given to treating physicians might result in unintentional aftereffects on families. For example, in cases of testing to determine potential birth effects in an unborn child, both parents undergo genetic testing. This, at times, can lead to paternity issues, which can have serious consequences for concerned families. All the above issues highlight the importance of legislative and confidentiality issues related to patients' health information and require due consideration while developing PM-oriented laws.

Economic Impact of PM

The principal barrier to implementing PM in clinics is meeting the cost required for its execution. Economic evaluations of PM interventions for making policy decisions pertaining to investment in research and development (R and D) and reimbursement in healthcare systems have been suggested (234). PM has the potential to reduce the cost of healthcare as it can predict the right treatment for the right patient, thereby reducing any unnecessary and multiple trial-and-error attempts to achieve healthy patient outcomes. However, to reach this stage, huge investment in multiple directions is required, right from the secured and massive infrastructure to hold and collect patient data to the development of a precision drug. The R and D of PM can be costlier than traditional medicine because it requires the use of expensive techniques to test, like sequencing a large amount of DNA or RNA and other genetic testing. Hence, despite its potential to decrease the treatment time and unnecessary side effects due to the use of broad-spectrum drugs currently used in cancer therapy, the overall cost of healthcare may prove to be a burden for not-so-affluent patients and insurance payers. To put that in perspective, the US invested $ 215 million in funding toward PM out of $35 billion in health research and development in 2016 (235). The development of a drug is the costliest aspect of any healthcare system. It currently exceeds $2.7 billion for cancer drugs, which is regularly used as a justification for higher cancer drug prices (236). Reimbursement from health insurance companies will get tougher with the rise in patients' out-of-pocket costs. All these factors will have an indirect and a direct effect on patients and can mean higher costs for patients to avail PM. This will only lead to economic discrimination, with only the affluent sector of the public being able to afford the PM treatments. Strategic approaches need to be in place like changing regulations governing health insurance as well as pharmaceutical companies to overcome the cost challenges involved in bringing PM to deliver improved health outcomes to a broader sector of patients. A recent study showed that stratifying patients according to PM-based approaches could save approximately $7 million in overall healthcare costs per 1,000 patients (237). In that context, detection of HER2, a validated biomarker in a specific cohort of patients with breast cancer overexpressing HER2 in tumors, reduced the clinical trial risk by 50 and 27% reduction in cost (238). There are studies that convincingly prove that using precise-protocoled PM therapeutics with well-designed analytic strategies can reduce clinical trial risks in patients with cancer (239–241).

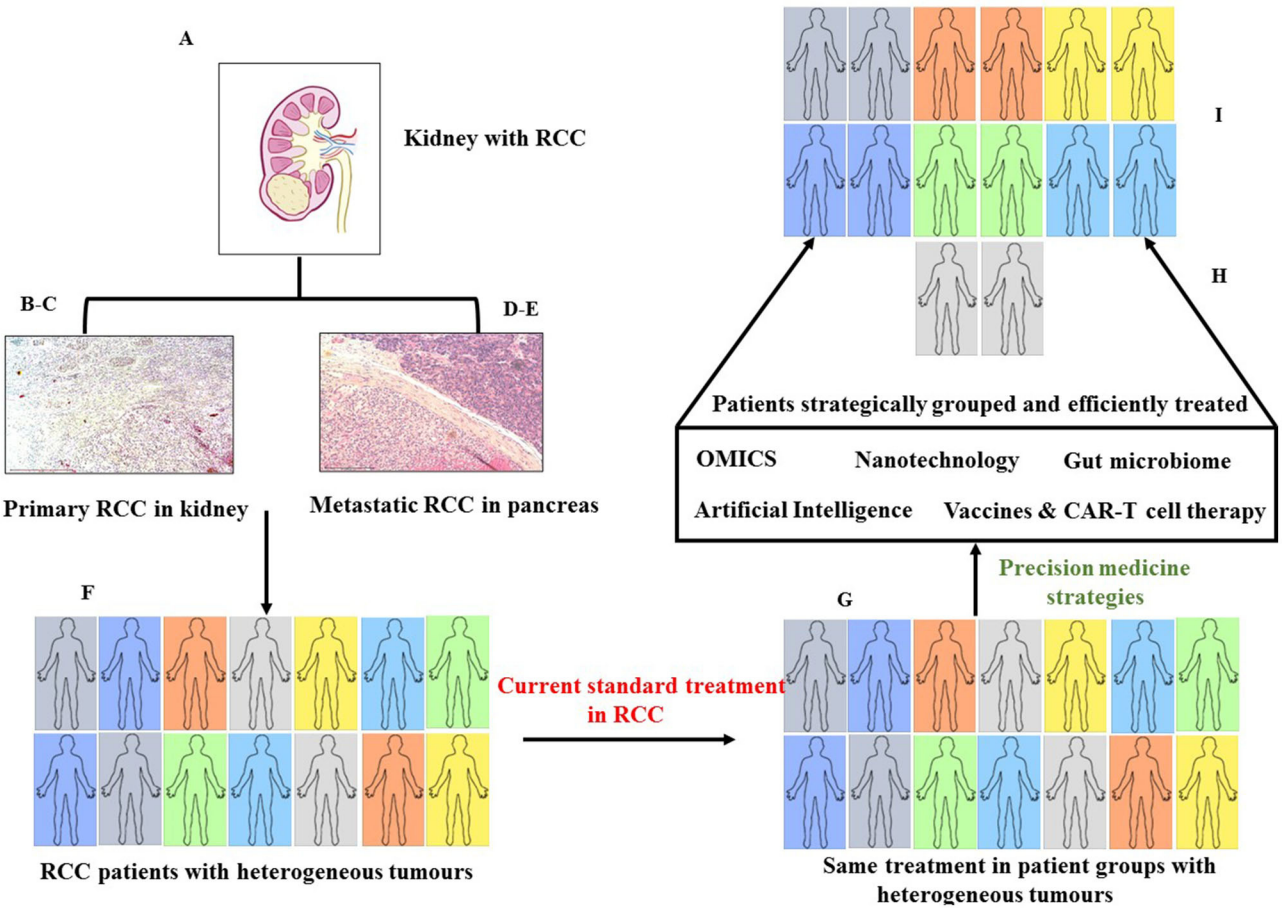

FIGURE 2 | Role of precision medicine (PM) in RCC. **(A)** Hematoxylin and Eosin (H & E) staining of human kidney with RCC, **(B)** H & E of primary RCC showing typical epithelial nests of RCC cells with clear cytoplasm (Magnification: 100X), **(C)** adjacent normal kidney tissue, **(D)** metastatic RCC invading the pancreas of a patient tissue, **(E)** adjacent normal spleen in a patient tissue (Magnification: 100X). **(F)** heterogeneity in RCC patients due to diverse clinical and environmental factors, **(G)** current practice of treating patients – every patient treated with same standard drugs resulting in treatment failure and secondary resistance, **(H)** precision medicine approach utilizing data obtained from various platforms and stratifying patients with personalized treatments, **(I)** individualized treatment (PM) resulting in better clinical outcomes for patients.

TABLE 2 | Summary of applications of precision medicine (PM) to facilitate treatment of patients with renal cell cancer (RCC).

Tools for personalized data collection				
Lifestyle data	Genomics	Proteomics	Transcriptomics	Metabolomics
New approaches to precision introduced in precision medicine				
Artificial Intelligence		Gut Microbiology	Nanotechnology	
Strategies of delivering PM in RCC				
Vaccines		Cellular therapies & organoids	Monoclonal antibodies	

CONCLUSIONS

The number of PM-based pharmaceutical drugs used in clinics has increased more than double from 2016 [132] to 2020 [286] (242). In that context, 31 genome-targeted anti-cancer drugs were used in 2018 [159]. The conventional cancer therapeutics reduce tumor burden and treas cancer-related symptoms, but relapse is common in most cancer cases. However, with the aid of genetic testing and the identification of specific protein biomarkers, more predictable patient health outcomes are possible.

RCC is a heterogeneous tumor at both the clinical and molecular levels. At the same time, RCC is a vascular, immunogenic, and metabolic tumor. The extensive variability in RCC tumor subtypes makes it a perfect candidate for PM, as the requirement for a patient-centered approach is high. Although much is known about RCC, the current treatment regimen

relies highly on prognostic stratification of patients based on the assessment of clinical factors (243). Although several attempts were made to find accurate biomarkers to evaluate response to target therapies and immunotherapy, none of them have been successfully adopted for mass patients with RCC. Hence, finding validated genetic or molecular biomarkers to assist in clinical decision-making is of clinical priority and an important step in taking RCC toward the PM arena. The use of patient-derived organoids and PDX and the development of cancer vaccines and CAR-T cells are slowly but steadily changing the arena of current PM. This is pushing the boundaries of current treatments more toward individually focused treatment in RCC. With skyrocketing costs of healthcare, PM has a scope of providing personalized healthcare at affordable costs by cutting repeated trial-and-error treatment strategies, which not only take a heavy toll on patients' health but also affect the immediate families. While celebrating the accomplishments of PM is necessary, its failures should be addressed as well. Tight regulations should govern PM products and services to provide essential and cost-effective positive patient-oriented health outcomes. While conventional medicine is considered as a shotgun that shoots pellets in a wide range with the hope to hit the target in the line of fire, PM may mitigate its way as a laser with focus only on the precise targets. It has the potential to reduce the cancer-associated economic and social burden. While there is still a lot to be achieved in the struggle for PM in RCC treatment, which is far from reaching clinics, PM is the way that offers more specific and individualized treatments for patients. **Figure 2** and **Table 2** depicts the importance of PM in the treatment of RCC patients.

AUTHOR CONTRIBUTIONS

RS undertook the literature review, was involved with the conception of the idea, and wrote the article. PP and GK edited and read the final version of the manuscript. NA was involved in conceiving the idea and writing and editing the manuscript. All the authors have read and approved the manuscript.

FUNDING

This study was supported by John Turner Cancer Research Funds to Fiona Elsey Cancer Research Institute, Ballarat, Australia. RS is a recipient of a John Turner Cancer Research PhD scholarship, Fiona Elsey Cancer Research Institute, and Federation University Australia, Ballarat, Australia. This study was made possible through funds from John Turner Cancer Research Funds and the Victorian State Government Operational Infrastructure Support to Hudson Institute of Medical Research.

REFERENCES

1. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. *Nat Rev Dis primers.* (2017) 3:17009. doi: 10.1038/nrdp.2017.9
2. Padala SA, Barsouk A, Thandra KC, Saginala K, Mohammed A, Vakiti A, et al. Epidemiology of renal cell carcinoma. *World J Oncol.* (2020) 11:79–87. doi: 10.14740/wjon1279
3. Lindblad P. Epidemiology of renal cell carcinoma. *Scand J Surg.* (2004) 93:88–96. doi: 10.1177/145749690409300202
4. Hung RJ, Moore L, Boffetta P, Feng BJ, Toro JR, Rothman N, et al. Family history and the risk of kidney cancer: a multicenter case-control study in Central Europe. *Cancer Epidemiol Biomarkers Prev.* (2007) 16:1287–90. doi: 10.1158/1055-9965.EPI-06-0963
5. Bensouilah FZ, Chellat-Rezgoune D, Garcia-Gonzalez MA, Carrera N, Abadi N, Dahdouh A, et al. Association of single nucleotide polymorphisms with renal cell carcinoma in Algerian population. *African J Urol.* (2020) 26:48. doi: 10.1186/s12301-020-00055-4
6. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol.* (2015) 67:519–30. doi: 10.1016/j.eururo.2014.10.002
7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
8. Athanazio DA, Amorim LS, da Cunha IW, Moreira KR, da Paz AR, Gomes RX, et al. Classification of renal cell tumors – current concepts and use of ancillary tests: recommendations of the Brazilian Society of Pathology. *Surg Exp Pathol.* (2021) 4:1–21. doi: 10.1186/s42047-020-00084-x
9. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO classification of tumours of the urinary system and male genital organs-part a: renal, penile, and testicular tumours. *Eur Urol.* (2016) 70:93–105. doi: 10.1016/j.eururo.2016.02.029
10. Chen YB, Xu J, Skanderup AJ, Dong Y, Brannon AR, Wang L, et al. Molecular analysis of aggressive renal cell carcinoma with unclassified histology reveals distinct subsets. *Nat Commun.* (2016) 7:13131. doi: 10.1038/ncomms13131
11. Riazalhosseini Y, Lathrop M. Precision medicine from the renal cancer genome. *Nat Rev Nephrol.* (2016) 12:655–66. doi: 10.1038/nrneph.2016.133
12. Krabbe LM, Bagrodia A, Margulis V, Wood CG. Surgical management of renal cell carcinoma. *Semin Intervent Radiol.* (2014) 31:27–32. doi: 10.1055/s-0033-1363840
13. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. *J Urol.* (2002) 168:2395–400. doi: 10.1016/S0022-5347(05)64153-5
14. Martel CL, Lara PN. Renal cell carcinoma: current status and future directions. *Crit Rev Oncol Hematol.* (2003) 45:177–90. doi: 10.1016/S1040-8428(02)00076-8
15. Ramsey S, Aitchison M. Treatment for renal cancer: are we beyond the cytokine era? *Nat Clin Pract Urol.* (2006) 3:478–84. doi: 10.1038/ncpuro0581
16. Koneru R, Hotte SJ. Role of cytokine therapy for renal cell carcinoma in the era of targeted agents. *Curr Oncol.* (2009) 16:S40–44. doi: 10.3747/co.v16i0.417
17. Morais C. Sunitinib resistance in renal cell carcinoma. *J Kidney Cancer VHL.* (2014) 1:1–11. doi: 10.15586/jkcvhl.2014.7
18. Li Q, Zhang Z, Fan Y, Zhang Q. Epigenetic alterations in renal cell cancer with TKIs resistance: from mechanisms to clinical applications. *Front Genet.* (2021) 11:562868. doi: 10.3389/fgene.2020.562868
19. Hsieh JJ, Le V, Cao D, Cheng EH, Creighton CJ. Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. *J Pathol.* (2018) 244:525–37. doi: 10.1002/path.5022
20. Burney IA, Lakhtakia R. Precision Medicine: Where have we reached and where are we headed? *Sultan Qaboos Univ Med J.* (2017) 17:e255–8. doi: 10.18295/squmj.2017.17.03.001
21. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized

medicine for cancer treatment. *Technology (Singap World Sci).* (2018) 6:79–100. doi: 10.1142/S2339547818300020
22. Carmeliet P, VEGF. as a key mediator of angiogenesis in cancer. *Oncology.* (2005) 3:4–10. doi: 10.1159/000088478
23. Jonasch E, McCutcheon IE, Waguespack SG, Wen S, Davis DW, Smith LA, et al. Pilot trial of sunitinib therapy in patients with von Hippel-Lindau disease. *Ann Oncol.* (2011) 22:2661–6. doi: 10.1093/annonc/mdr011
24. Sharma R, Kadife E, Myers M, Kannourakis G, Prithviraj P, Ahmed N. Determinants of resistance to VEGF-TKI and immune checkpoint inhibitors in metastatic renal cell carcinoma. *J Exp Clin Cancer Res.* (2021) 40:186. doi: 10.1186/s13046-021-01961-3
25. Guo H, German P, Bai S, Barnes S, Guo W, Qi X, et al. The PI3K/AKT pathway and renal cell carcinoma. *J Genet Genomics.* (2015) 42:343–53. doi: 10.1016/j.jgg.2015.03.003
26. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet.* (2013) 45:860–7. doi: 10.1038/ng.2699
27. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods. (2013) 10:1046–7. doi: 10.1038/nmeth.2650
28. Akbani R, Ng PK, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun.* (2014) 5:3887. doi: 10.1038/ncomms4887
29. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet.* (2013) 45:1113–20. doi: 10.1038/ng.2764
30. Correa AF, Ruth KJ, Al-Saleem T, Pei J, Dulaimi E, Kister D, et al. Overall tumor genomic instability: an important predictor of recurrence-free survival in patients with localized clear cell renal cell carcinoma. *Cancer Biol Ther.* (2020) 21:424–31. doi: 10.1080/15384047.2020.1721251
31. Ma Q, Wang J, Qi J, Peng D, Guan B, Zhang J, et al. Increased chromosomal instability characterizes metastatic renal cell carcinoma. *Transl Oncol.* (2021) 14:100929. doi: 10.1016/j.tranon.2020.100929
32. Feng C, Sun Y, Ding G, Wu Z, Jiang H, Wang L et al. PI3Kβ Inhibitor TGX221 Selectively Inhibits Renal Cell Carcinoma Cells with Both VHL and SETD2 mutations and Links Multiple Pathways. *Sci Rep.* (2015) 5:9465. doi: 10.1038/srep09465
33. Banumathy G, Cairns P. Signaling pathways in renal cell carcinoma. *Cancer Biol Ther.* (2010) 10:658–64. doi: 10.4161/cbt.10.7.13247
34. Shin SH, Bode AM, Dong Z. Precision medicine: the foundation of future cancer therapeutics. *NPJ Precis Oncol.* (2017) 1:12. doi: 10.1038/s41698-017-0016-z
35. Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. Human genomics projects and precision medicine. *Gene Ther.* (2017) 24:551–61. doi: 10.1038/gt.2017.77
36. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* (2001) 344:783–92. doi: 10.1056/NEJM200103153441101
37. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med.* (2010) 363:809–19. doi: 10.1056/NEJMoa1002011
38. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res.* (2012) 5:19–27.
39. Scheerens H, Malong A, Bassett K, Boyd Z, Gupta V, Harris J, et al. Current Status of Companion and Complementary Diagnostics: Strategic Considerations for Development and Launch. *Clin Transl Sci.* (2017) 10:84–92. doi: 10.1111/cts.12455
40. Administration, A.G.D.o.H.T.G. IVD companion diagnostics-Guidance on regulatory requirements. (2020). Version 1.1: Available online at: https://www.tga.gov.au/publication/ivd-companion-diagnostics (accessed July 23, 2021)
41. Arteaga CL, Baselga J. Impact of genomics on personalized cancer medicine. *Clin Cancer Res.* (2012) 18:612–8. doi: 10.1158/1078-0432.CCR-11-2019
42. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and male genital organs-part B: prostate and bladder tumours. *Eur Urol.* (2016) 70:106–19. doi: 10.1016/j.eururo.2016.02.028
43. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol.* (1999) 17:2530–40. doi: 10.1200/JCO.1999.17.8.2530
44. Ellis LM, Hicklin DJ. VEGF-targeted therapy: mechanisms of anti-tumour activity. *Nat Rev Cancer.* (2008) 8:579–91. doi: 10.1038/nrc2403
45. Haibe Y, Kreidieh M, El Hajj H, Khalifeh I, Mukherji D, Temraz S, et al. Resistance mechanisms to anti-angiogenic therapies in cancer. *Front Oncol.* (2020) 10:221. doi: 10.3389/fonc.2020.00221
46. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med.* (2018) 378:1277–90. doi: 10.1056/NEJMoa1712126
47. Rini BI, Plimack ER, Stus V, Gafanov R, Hawkins R, Nosov D, et al. Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med.* (2019) 380:1116–27. doi: 10.1056/NEJMoa1816714
48. Motzer RJ, Penkov K, Haanen J, Rini B, Albiges L, Campbell MT, et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med.* (2019) 380:1103–15. doi: 10.1056/NEJMoa1816047
49. Rini BI, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, et al. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet.* (2019) 393:2404–15. doi: 10.1016/S0140-6736(19)30723-8
50. Dudani S, de Velasco G, Wells JC, Gan CL, Donskov F, Porta C, et al. Evaluation of clear cell, papillary, and chromophobe renal cell carcinoma metastasis sites and association with survival. *JAMA Netw Open.* (2021) 4:e2021869. doi: 10.1001/jamanetworkopen.2020.21869
51. Crusz SM, Tang YZ, Sarker SJ, Prevoo W, Kiyani I, Beltran L, et al. Heterogeneous response and progression patterns reveal phenotypic heterogeneity of tyrosine kinase inhibitor response in metastatic renal cell carcinoma. *BMC Med.* (2016) 14:185. doi: 10.1186/s12916-016-0729-9
52. Ferronika P, Hof J, Kats-Ugurlu G, Terpstra M, De lange K, Leliveld-Kors A et al. DNA and RNA analysis of intratumour heterogeneity in metastatic clear cell renal cell carcinoma. *Ann Oncol.* (2017) 28:vii31. doi: 10.1093/annonc/mdx510.001
53. Zhi H, Feng M, Liu S, Na T, Zhang N, BiLiGe W. Prognostic significance of sarcomatoid differentiation in patients with metastatic renal cell carcinoma: a systematic review and meta-analysis. *Front Oncol.* (2020) 10:591001. doi: 10.3389/fonc.2020.591001
54. Singh RR, Murugan P, Patel LR, Voicu H, Yoo SY, Majewski T, et al. Intratumoral morphologic and molecular heterogeneity of rhabdoid renal cell carcinoma: challenges for personalized therapy. *Mod Pathol.* (2015) 28:1225–35. doi: 10.1038/modpathol.2015.68
55. Cai Q, Christie A, Rajaram S, Zhou Q, Araj E, Chintalapati S, et al. Ontological analyses reveal clinically-significant clear cell renal cell carcinoma subtypes with convergent evolutionary trajectories into an aggressive type. *EBioMedicine.* (2020) 51:102526. doi: 10.1016/j.ebiom.2019.10.052
56. Stanta G, Bonin S. Overview on clinical relevance of intra-tumor heterogeneity. *Front Med.* (2018) 5:85. doi: 10.3389/fmed.2018.00085
57. Hakimi AA, Reznik E, Lee CH, Creighton CJ, Brannon AR, Luna A, et al. An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell.* (2016) 29:104–16. doi: 10.1016/j.ccell.2015.12.004
58. D'Abbronzo G, Franco R. The changing role of the pathologist in the era of targeted therapy in personalized medicine. *Expert Rev Precis Med Drug Develop.* (2021) 6:295–297. doi: 10.1080/23808993.2021.1923400
59. Srigley JR, Delahunt B, Eble JN, Egevad L, Epstein JI, Grignon D, et al. The international society of urological pathology (ISUP) vancouver classification of renal neoplasia. *Am J Surg Pathol.* (2013) 37:1469–89. doi: 10.1097/PAS.0b013e318299f2d1
60. Perlman EJ. Pediatric Renal Cell Carcinoma. *Surg Pathol Clin.* (2010) 3:641–651. doi: 10.1016/j.path.2010.06.011
61. Argani P, Zhong M, Reuter VE, Fallon JT, Epstein JI, Netto GJ, et al. TFE3-Fusion variant analysis defines specific clinicopathologic associations among Xp11 translocation cancers. *Am J Surg Pathol.* (2016) 40:723–37. doi: 10.1097/PAS.0000000000000631

62. Signoretti S, Flaifel A, Chen YB, Reuter VE. Renal cell carcinoma in the era of precision medicine: from molecular pathology to tissue-based biomarkers. *J Clin Oncol*. (2018) 36:JCO2018792259. doi: 10.1200/JCO.2018.79.2259
63. Cheville JC, Lohse CM, Zincke H, Weaver AL, Leibovich BC, Frank I, et al. Sarcomatoid renal cell carcinoma: an examination of underlying histologic subtype and an analysis of associations with patient outcome. *Am J Surg Pathol*. (2004) 28:435–41. doi: 10.1097/00000478-200404000-00002
64. Zhang BY, Cheville JC, Thompson RH, Lohse CM, Boorjian SA, Leibovich BC, et al. Impact of rhabdoid differentiation on prognosis for patients with grade 4 renal cell carcinoma. *Eur Urol*. (2015) 68:5–7. doi: 10.1016/j.eururo.2015.01.002
65. Maitland ML, Xu CF, Cheng YC, Kistner-Griffin E, Ryan KA, Karrison TG et al. Identification of a variant in KDR associated with serum VEGFR2 and pharmacodynamics of Pazopanib. *Clin Cancer Res*. (2015) 21:365–72. doi: 10.1158/1078-0432.CCR-14-1683
66. Hoefflin R, Harlander S, Schäfer S, Metzger P, Kuo F, Schönenberger D, et al. HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nat Commun*. (2020) 11:4111. doi: 10.1038/s41467-020-17873-3
67. Motzer RJ, Hutson TE, Hudes GR, Figlin RA, Martini JF, English PA, et al. Investigation of novel circulating proteins, germ line single-nucleotide polymorphisms, and molecular tumor markers as potential efficacy biomarkers of first-line sunitinib therapy for advanced renal cell carcinoma. *Cancer Chemother Pharmacol*. (2014) 74:739–50. doi: 10.1007/s00280-014-2539-0
68. Choueiri TK, Regan MM, Rosenberg JE, Oh WK, Clement J, Amato AM, et al. Carbonic anhydrase IX and pathological features as predictors of outcome in patients with metastatic clear-cell renal cell carcinoma receiving vascular endothelial growth factor-targeted therapy. *BJU Int*. (2010) 106:772–8. doi: 10.1111/j.1464-410X.2010.09218.x
69. D' Alterio C, Portella L, Ottaiano A, Rizzo M, Carteni G, Pignata S, et al. High CXCR4 expression correlates with sunitinib poor response in metastatic renal cancer. *Curr Cancer Drug Targets*. (2012) 12:693–702. doi: 10.2174/156800912801784820
70. Errarte P, Larrinaga G, López JI. The role of cancer-associated fibroblasts in renal cell carcinoma. An example of tumor modulation through tumor/non-tumor cell interactions. *J Adv Res*. (2019) 21:103–8. doi: 10.1016/j.jare.2019.09.004
71. Gossage L, Eisen T. Alterations in VHL as potential biomarkers in renal-cell carcinoma. *Nat Rev Clin Oncol*. (2010) 7:277–88. doi: 10.1038/nrclinonc.2010.42
72. Kondo K, Yao M, Yoshida M, Kishida T, Shuin T, Miura T, et al. Comprehensive mutational analysis of the VHL gene in sporadic renal cell carcinoma: relationship to clinicopathological parameters. *Genes Chromosomes Cancer*. (2002) 34:58–68. doi: 10.1002/gcc.10054
73. Kim BJ, Kim JH, Kim HS, Zang DY. Prognostic and predictive value of VHL gene alteration in renal cell carcinoma: a meta-analysis and review. *Oncotarget*. (2017) 8:13979–85. doi: 10.18632/oncotarget.14704
74. Choueiri TK, Fay AP, Gagnon R, Lin Y, Bahamon B, Brown V, et al. The role of aberrant VHL/HIF pathway elements in predicting clinical outcome to pazopanib therapy in patients with metastatic clear-cell renal cell carcinoma. *Clin Cancer Res*. (2013) 19:5218–26. doi: 10.1158/1078-0432.CCR-13-0491
75. Choueiri TK, Vaziri SA, Jaeger E, Elson P, Wood L, Bhalla IP, et al. von Hippel-Lindau gene status and response to vascular endothelial growth factor targeted therapy for metastatic clear cell renal cell carcinoma. *J Urol*. (2008) 180:860–5. doi: 10.1016/j.juro.2008.05.015
76. Ho TH, Choueiri TK, Wang K, Karam JA, Chalmers Z, Frampton G, et al. Correlation between molecular subclassifications of clear cell renal cell carcinoma and targeted therapy response. *Eur Urol Focus*. (2016) 2:204–9. doi: 10.1016/j.euf.2015.11.007
77. D'Aniello C, Berretta M, Cavaliere C, Rossetti S, Facchini BA, Iovane G, et al. Biomarkers of prognosis and efficacy of anti-angiogenic therapy in metastatic clear cell renal cancer. *Front Oncol*. (2019) 9:1400. doi: 10.3389/fonc.2019.01400
78. Hakimi AA, Attalla K, DiNatale RG, Ostrovnaya I, Flynn J, Kyle A, et al. A pan-cancer analysis of PBAF complex mutations and their association with immunotherapy response. *Nat Commun*. (2020) 11:4168. doi: 10.1038/s41467-020-17965-0
79. Braun DA, Ishii Y, Walsh AM, Van Allen EM, Wu CJ, Shukla SA, et al. Clinical validation of PBRM1 alterations as a marker of immune checkpoint inhibitor response in renal cell carcinoma. *JAMA Oncol*. (2019) 5:1631–3. doi: 10.1001/jamaoncol.2019.3158
80. Hsieh JJ, Chen D, Wang PI, Marker M, Redzematovic A, Chen YB, et al. Genomic biomarkers of a randomized trial comparing first-line everolimus and sunitinib in patients with metastatic renal cell carcinoma. *Eur Urol*. (2017) 71:405–14. doi: 10.1016/j.eururo.2017.01.013
81. Jin S, Wu J, Zhu Y, Gu W, Wan F, Xiao W, et al. Comprehensive analysis of *BAP1* somatic mutation in clear cell renal cell carcinoma to explore potential mechanisms in silico. *J Cancer*. (2018) 9:4108–16. doi: 10.7150/jca.27281
82. Stenehjem DD, Hahn AW, Gill DM, Albertson D, Gowrishankar B, Merriman J, et al. Predictive genomic markers of response to VEGF targeted therapy in metastatic renal cell carcinoma. *PLoS ONE*. (2019) 14:e0210415. doi: 10.1371/journal.pone.0210415
83. Salgia N, Dizman N, Lyou Y, Bergerot PG, Hsu J, Byron SA, et al. Genomic and transcriptomic correlates of clinical benefit from immunotherapy and targeted therapy among patients with metastatic renal cell carcinoma (mRCC). *J Clin Oncol*. (2020) 38:5076–5076. doi: 10.1200/JCO.2020.38.15_suppl.5076
84. Ko JJ, Xie W, Kroeger N, Lee JL, Rini BI, Knox JJ, et al. The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. *Lancet Oncol*. (2015) 16:293–300. doi: 10.1016/S1470-2045(14)71222-7
85. Williams MS, Taylor CO, Walton NA, Goehringer SR, Aronson S, Freimuth RR, et al. Genomic information for clinicians in the electronic health record: lessons learned from the clinical genome resource project and the electronic medical records and genomics network. *Front Genet*. (2019) 10:1059. doi: 10.3389/fgene.2019.01059
86. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet*. (2011) 52:413–35. doi: 10.1007/s13353-011-0057-x
87. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. (2016) 107:1–8. doi: 10.1016/j.ygeno.2015.11.003
88. Sterky F, Holmberg A, Alexandersson G, Lundeberg J, Uhlén M. Direct sequencing of bacterial artificial chromosomes (BACs) and prokaryotic genomes by biotin-capture PCR. *J Biotechnol*. (1998) 60:119–29. doi: 10.1016/S0168-1656(97)00196-X
89. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. (2016) 17:333–51. doi: 10.1038/nrg.2016.49
90. van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet*. (2009) 41:521–3.
91. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*. (2011) 469:539–42. doi: 10.1038/nature09639
92. Peña-Llopis S. Vega-Rubín-de-Celis S, Liao A, Leng N, Pavía-Jiménez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet*. (2012) 44:751–9. doi: 10.1038/ng.2323
93. Cancer Genome Atlas Research Network, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N Engl J Med*. (2016) 374:135–45. doi: 10.1056/NEJMoa1505917
94. Durinck S, Stawiski EW, Pavía-Jiménez A, Modrusan Z, Kapur P, Jaiswal BS, et al. Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. *Nat Genet*. (2015) 47:13–21. doi: 10.1038/ng.3146
95. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. (2017) 13:e1005457. doi: 10.1371/journal.pcbi.1005457
96. Tan MH, Rogers CG, Cooper JT, Ditlev JA, Maatman TJ, Yang X, et al. Gene expression profiling of renal cell carcinoma. *Clin Cancer Res*. (2004) 10:6315S−21S. doi: 10.1158/1078-0432.CCR-050002
97. Lakshminarayanan H, Rutishauser D, Schraml P, Moch H, Bolck HA. Liquid biopsies in renal cell carcinoma-recent advances and promising new technologies for the early detection of metastatic disease. *Front Oncol*. (2020) 10:582843. doi: 10.3389/fonc.2020.582843

98. Sanchez A, Furberg H, Kuo F, Vuong L, Ged Y, Patil S, et al. Transcriptomic signatures related to the obesity paradox in patients with clear cell renal cell carcinoma: a cohort study. *Lancet Oncol.* (2020) 2:283–93. doi: 10.1016/S1470-2045(19)30797-1
99. Roehrl MH, Roehrl VB, Wang JY. Proteome-based pathology: the next frontier in precision medicine. *Expert Rev Precis Med Drug Dev.* (2021) 6:1–4. doi: 10.1080/23808993.2021.1854611
100. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. (2012) 13:227–32. doi: 10.1038/nrg3185
101. Heng DY, Kollmannsberger C, Chi KN. Targeted therapy for metastatic renal cell carcinoma: current treatment and future directions. *Ther Adv Med Oncol.* (2010) 2:39–49. doi: 10.1177/1758834009352498
102. Raimondo F, Salemi C, Chinello C, Fumagalli D, Morosi L, Rocco F, et al. Proteomic analysis in clear cell renal cell carcinoma: identification of differentially expressed protein by 2-D DIGE. *Mol Biosyst.* (2012) 8:1040–51. doi: 10.1039/c2mb05390j
103. Starita-Geribaldi M, Thaon-Scarzello S, Le Blanc M, Van Obberghen E, Poustis-Delpont C. Two-dimensional polyacrylamide gel electrophoresis of the protease SP220K, a renal cell carcinoma marker. *Bioseparation.* (2000) 9:133–44. doi: 10.1023/A:1008198521231
104. Magdeldin S, Enany S, Yoshida Y, Xu B, Zhang Y, Zureena Z, et al. Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clin Proteomics.* (2014) 11:16. doi: 10.1186/1559-0275-11-16
105. Clark DJ, Zhang H. Proteomic approaches for characterizing renal cell carcinoma. *Clin Proteomics*. (2020) 17:28. doi: 10.1186/s12014-020-09291-w
106. Sun X, Zhang H, Luo L, Zhong K, Ma Y, Fan L, et al. Comparative proteomic profiling identifies potential prognostic factors for human clear cell renal cell carcinoma. *Oncol Rep.* (2016) 36:3131–8. doi: 10.3892/or.2016.5159
107. Duarte TT, Spencer CT. Personalized proteomics: the future of precision medicine. *Proteomes.* (2016) 4:29. doi: 10.3390/proteomes4040029
108. Gregorich ZR, Ge Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics.* (2014) 14:1195–210. doi: 10.1002/pmic.201300432
109. Atrih A, Mudaliar MA, Zakikhani P, Lamont DJ, Huang JT, Bray SE, et al. Quantitative proteomics in resected renal cancer tissue for biomarker discovery and profiling. *Br J Cancer.* (2014) 110:1622–33. doi: 10.1038/bjc.2014.24
110. Sidoli S, Lu C, Coradin M, Wang X, Karch KR, Ruminowicz C, et al. Metabolic labeling in middle-down proteomics allows for investigation of the dynamics of the histone code. *Epigenet Chromatin*. (2017) 10:34. doi: 10.1186/s13072-017-0139-z
111. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. *Curr Med Chem.* (2013) 20:257–71. doi: 10.2174/092986713804806621
112. Dettmer K, Hammock BD. Metabolomics–a new exciting field within the "omics" sciences. *Environ Health Perspect*. (2004) 112:A396–7. doi: 10.1289/ehp.112-1241997
113. Everett JR, Loo RL, Pullen FS. Pharmacometabonomics and personalized medicine. *Ann Clin Biochem.* (2013) 6:523–45. doi: 10.1177/0004563213497929
114. Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancerr. *Anal Biochem.* (2007) 363:185–95. doi: 10.1016/j.ab.2007.01.028
115. Zira AN, Theocharis SE, Mitropoulos D, Migdalis V, Mikros E. (1)H NMR metabonomic analysis in renal cell carcinoma: a possible diagnostic tool. *J Proteome Res.* (2010) 9:4038–44. doi: 10.1021/pr100226m
116. Ragone R, Sallustio F, Piccinonna S, Rutigliano M, Vanessa G, Palazzo S, et al. Renal cell carcinoma: a study through NMR-based metabolomics combined with transcriptomics. *Diseases.* (2016) 4:7. doi: 10.3390/diseases4010007
117. Gao H, Dong B, Liu X, Xuan H, Huang Y, Lin D. Metabonomic profiling of renal cell carcinoma: high-resolution proton nuclear magnetic resonance spectroscopy of human serum with multivariate data analysis. *Anal Chim Acta.* (2008) 624:269–77. doi: 10.1016/j.aca.2008.06.051
118. Zheng H, Ji J, Zhao L, Chen M, Shi A, Pan L, et al. Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps. *Oncotarget.* (2016) 7:59189–98. doi: 10.18632/oncotarget.10830
119. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev.* (2007) 26:51–78. doi: 10.1002/mas.20108
120. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom.* (2016) 27:1897–905. doi: 10.1007/s13361-016-1469-y
121. Lucarelli G, Rutigliano M, Sallustio F, Ribatti D, Giglio A, Lepore, et al. Integrated multi-omics characterization reveals a distinctive metabolic signature and the role of NDUFA4L2 in promoting angiogenesis, chemoresistance, and mitochondrial dysfunction in clear cell renal cell carcinoma. *Aging.* (2018) 10:3957–85. doi: 10.18632/aging.101685
122. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights.* (2020) 14:1177932219899051. doi: 10.1177/1177932219899051
123. Wanichthanarak K, Fan S, Grapov D, Barupal DK, Fiehn O. Metabox: A toolbox for metabolomic data analysis, interpretation and integrative exploration. *PLoS ONE.* (2017) 12:e0171046. doi: 10.1371/journal.pone.0171046
124. Sangaralingam A, Dayem Ullah AZ, Marzec J, Gadaleta E, Nagano A, Ross-Adams H, et al. 'Multi-omic' data analysis using O-miner. *Brief Bioinform.* (2019) 20:130–43. doi: 10.1093/bib/bbx080
125. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* (2018) 46:W537–44. doi: 10.1093/nar/gky379
126. Gupta S, Kang HC, Ganeshan DM, Bathala TK, Kundra V. Diagnostic approach to hereditary renal cell carcinoma. *AJR Am J Roentgenol.* (2015) 204:1031–41. doi: 10.2214/AJR.14.13514
127. Kabaria R, Klaassen Z, Terris MK. Renal cell carcinoma: links and risks. *Int J Nephrol Renovasc Dis.* (2016) 9:45–52. doi: 10.2147/IJNRD.S75916
128. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. *Nat Rev Urol.* (2010) 7:245–57. doi: 10.1038/nrurol.2010.46
129. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn).* (2015) 19:A68–77. doi: 10.5114/wo.2014.47136
130. Birney E, Vamathevan J, Goodhand P. Genomics in healthcare: GA4GH looks to 2022. *bioRxiv [Preprint]*. (2017). doi: 10.1101/203554
131. Chen B, Chen W, Jin J, Wang X, Cao Y, He Y. Data mining of prognostic microenvironment-related genes in clear cell renal cell carcinoma: a study with TCGA database. *Dis Markers.* (2019) 8901649. doi: 10.1155/2019/8901649
132. Liu B, Chen X, Zhan Y, Wu B, Pan S. Identification of a Gene Signature for Renal Cell Carcinoma-Associated Fibroblasts Mediating Cancer Progression and Affecting Prognosis. *Front Cell Dev Biol.* (2021) 8:604627. doi: 10.3389/fcell.2020.604627
133. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep*. (2018) 23:313–326.e5.
134. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. *Yearb Med Inform.* (2014) 9:8–13. doi: 10.15265/IY-2014-0024
135. van der Meel R, Sulheim E, Shi Y, Kiessling F, Mulder WJM, Lammers T. Smart cancer nanomedicine. *Nat Nanotechnol.* (2019) 11:1007–17. doi: 10.1038/s41565-019-0567-y
136. Peer D, Karp JM, Hong S, Farokhzad OC, Margalit R, Langer R. Nanocarriers as an emerging platform for cancer therapy. *Nat Nanotechnol.* (2007) 2:751–60. doi: 10.1038/nnano.2007.387
137. Perrault SD, Chan WC. In vivo assembly of nanoparticle components to improve targeted cancer imaging. *Proc Natl Acad Sci U S A.* (2010) 107:11194–9. doi: 10.1073/pnas.1001367107
138. Maeda H, Nakamura H, Fang J. The EPR effect for macromolecular drug delivery to solid tumors: Improvement of tumor uptake, lowering of systemic toxicity, and distinct tumor imaging in vivo. *Adv Drug Deliv Rev.* (2013) 65:71–9. doi: 10.1016/j.addr.2012.10.002
139. Ma P, Mumper RJ. Paclitaxel nano-delivery systems: a comprehensive review. *J Nanomed Nanotechnol.* (2013) 4:1000164.
140. Danhier F, Lecouturier N, Vroman B, Jérôme C, Marchand-Brynaert J, Feron O, et al. Paclitaxel-loaded PEGylated PLGA-based nanoparticles:

in vitro and in vivo evaluation. *J Control Release.* (2009) 133:11–7. doi: 10.1016/j.jconrel.2008.09.086
141. Angelopoulou A, Kolokithas-Ntoukas A, Fytas C, Avgoustakis K. Folic acid-functionalized, condensed magnetic nanoparticles for targeted delivery of doxorubicin to tumor cancer cells overexpressing the folate receptor. *ACS Omega.* (2019) 4:22214–27. doi: 10.1021/acsomega.9b03594
142. Pizetta B, Raggi LG, Rocha KSS, Cerqueira-Santos S, de Lyra-Jr DP, Dos Santos Júnior GA. Does drug dispensing improve the health outcomes of patients attending community pharmacies? A systematic review. *BMC Health Serv Res*. (2021) 21:764. doi: 10.1186/s12913-021-06770-0
143. Mazzucchelli S, Truffi M, Fiandra L, Sorrentino L, Corsi F. Targeted approaches for HER2 breast cancer therapy: News from nanomedicine? *World J Pharmacol.* (2014) 3:72–85. doi: 10.5497/wjp.v3.i4.72
144. Liu C, Zhang P, Zhai X, Tian F, Li W, Yang J, et al. Nano-carrier for gene delivery and bioimaging based on carbon dots with PEI-passivation enhanced fluorescence. *Biomaterials.* (2012) 33:3604–13. doi: 10.1016/j.biomaterials.2012.01.052
145. Sofias AM, Toner YC, Meerwaldt AE, van Leent MMT, Soultanidis G, Elschot M, et al. Tumor Targeting by $\alpha_v\beta_3$-Integrin-Specific Lipid Nanoparticles Occurs via Phagocyte Hitchhiking. *ACS Nano.* (2020) 14:7832–46. doi: 10.1021/acsnano.9b08693
146. Madhankumar AB, Slagle-Webb B, Mintz A, Sheehan JM, Connor JR. Interleukin-13 receptor-targeted nanovesicles are a potential therapy for glioblastoma multiforme. *Mol Cancer Ther.* (2006) 5:3162–9. doi: 10.1158/1535-7163.MCT-06-0480
147. Choi HS, Liu W, Misra P, Tanaka E, Zimmer JP, Itty Ipe B, et al. Renal clearance of quantum dots. *Nat Biotechnol.* (2007) 25:1165–70. doi: 10.1038/nbt1340
148. Longmire M, Choyke PL, Kobayashi H. Clearance properties of nano-sized particles and molecules as imaging agents: considerations and caveats. *Nanomedicine.* (2008) 3:703–17. doi: 10.2217/17435889.3.5.703
149. Sancey L, Kotb S, Truillet C, Appaix F, Marais A, Thomas E, et al. Long-term in vivo clearance of gadolinium-based AGuIX nanoparticles and their biocompatibility after systemic injection. *ACS Nano.* (2015) 9:2477–88. doi: 10.1021/acsnano.5b00552
150. Bennett KM, Zhou H, Sumner JP, Dodd SJ, Bouraoud N, Doi K, et al. MRI of the basement membrane using charged nanoparticles as contrast agents. *Magn Reson Med.* (2008) 60:564–74. doi: 10.1002/mrm.21684
151. Alsaab HO, Sau S, Alzhrani RM, Cheriyan VT, Polin LA, Vaishampayan U, et al. Tumor hypoxia directed multimodal nanotherapy for overcoming drug resistance in renal cell carcinoma and reprogramming macrophages. *Biomaterials*. (2018) 183:280–294. doi: 10.1016/j.biomaterials.2018.08.053
152. Thambi T, Deepagan VG, Yoon HY, Han HS, Kim SH, Son S, et al. Hypoxia-responsive polymeric nanoparticles for tumor-targeted drug delivery. *Biomaterials.* (2014) 35:1735–43. doi: 10.1016/j.biomaterials.2013.11.022
153. Chen Z, Ailing L, Zhao Y, Zhu Li. Preparation of HR magnetic IONPs for drug delivery. *Micro Nano Lett.* (2019) 14:38–41. doi: 10.1049/mnl.2018.5263
154. Li Y, Lu A, Long M, Cui L, Chen Z, Zhu L. Nitroimidazole derivative incorporated liposomes for hypoxia-triggered drug delivery and enhanced therapeutic efficacy in patient-derived tumor xenografts. *Acta Biomater.* (2019) 83:334–48. doi: 10.1016/j.actbio.2018.10.029
155. Zhang P, Yang H, Shen W, Liu W, Chen L, Xiao C. Hypoxia-Responsive Polypeptide Nanoparticles Loaded with Doxorubicin for Breast Cancer Therapy. *ACS Biomater Sci Eng.* (2020) 6:2167–74. doi: 10.1021/acsbiomaterials.0c00125
156. Xie Z, Paras CB, Weng H, Punnakitikashem P, Su LC, Vu K, et al. Dual growth factor releasing multi-functional nanofibers for wound healing. *Acta Biomater.* (2013) 9:9351–9. doi: 10.1016/j.actbio.2013.07.030
157. Mukherjee A, Madamsetty VS, Paul MK, Mukherjee S. Recent advancements of nanomedicine towards antiangiogenic therapy in cancer. *Int J Mol Sci.* (2020) 21:455. doi: 10.3390/ijms21020455
158. Anand P, Thomas SG, Kunnumakkara AB, Sundaram C, Harikumar KB, Sung B, et al. Biological activities of curcumin and its analogues (Congeners) made by man and Mother Nature. *Biochem Pharmacol.* (2008) 76:1590–611. doi: 10.1016/j.bcp.2008.08.008
159. Andrews LP, Marciscano AE, Drake CG, Vignali DA. LAG3 (CD223) as a cancer immunotherapy target. *Immunol Rev.* (2017) 276:80–96. doi: 10.1111/imr.12519
160. Yang Q, Wang Y, Yang Q, Gao Y, Duan X, Fu Q, et al. Cuprous oxide nanoparticles trigger ER stress-induced apoptosis by regulating copper trafficking and overcoming resistance to sunitinib therapy in renal cancer. *Biomaterials.* (2017) 146:72–85. doi: 10.1016/j.biomaterials.2017.09.008
161. Zhang H. Osimertinib making a breakthrough in lung cancer targeted therapy. *Onco Targets Ther.* (2016) 9:5489–93. doi: 10.2147/OTT.S114722
162. Tran S, DeGiovanni PJ, Piel B, Rai P. Cancer nanomedicine: a review of recent success in drug delivery. *Clin Transl Med.* (2017) 6:44. doi: 10.1186/s40169-017-0175-0
163. von Roemeling C, Jiang W, Chan CK, Weissman IL, Kim BYS. Breaking down the barriers to precision cancer nanomedicine. *Trends Biotechnol.* (2017) 35:159–71. doi: 10.1016/j.tibtech.2016.07.006
164. Voss MH, Hussain A, Vogelzang N, Lee JL, Keam B, Rha SY, et al. A randomized phase II trial of CRLX101 in combination with bevacizumab versus standard of care in patients with advanced renal cell carcinoma. *Ann Oncol.* (2017) 28:2754–2760. doi: 10.1093/annonc/mdx493
165. Ma LX, Craig KM, Mosquera JM, Robinson BD, Scherr DS, Pizzo JD, et al. Contemporary results and clinical utility of renal mass biopsies in the setting of ablative therapy: a single center experience. *Cancer Treat Res Commun.* (2020) 25:100209. doi: 10.1016/j.ctarc.2020.100209
166. Scheckner B, Peyser A, Rube J, Tarapore F, Frank R, Vento S, et al. Diagnostic yield of renal biopsies: a retrospective single center review. *BMC Nephrol.* (2009) 10:11. doi: 10.1186/1471-2369-10-11
167. Woo S, Cho JY. Imaging findings of common benign renal tumors in the era of small renal masses: differential diagnosis from small renal cell carcinoma: current status and future perspectives. *Korean J Radiol.* (2015) 16:99–113. doi: 10.3348/kjr.2015.16.1.99
168. Kocak B, Kaya OK, Erdim C, Kus EA, Kilickesmez O. Artificial intelligence in renal mass characterization: a systematic review of methodologic items related to modeling, performance evaluation, clinical utility, and transparency. *AJR Am J Roentgenol.* (2020) 215:1113–22. doi: 10.2214/AJR.20.22847
169. Lubner MG. Radiomics and artificial intelligence for renal mass characterization. *Radiol Clin North Am.* (2020) 58:995–1008. doi: 10.1016/j.rcl.2020.06.001
170. Zhu M, Ren B, Richards R, Suriawinata M, Tomita N, Hassanpour S. Development and evaluation of a deep neural network for histologic classification of renal cell carcinoma on biopsy and surgical resection slides. *Sci Rep.* (2021) 11:7080. doi: 10.1038/s41598-021-86540-4
171. Kim H, Lee SJ, Park SJ, Choi IY, Hong SH. Machine learning approach to predict the probability of recurrence of renal cell carcinoma after surgery: prediction model development study. *JMIR Med Inform.* (2021) 9:e25635. doi: 10.2196/25635
172. Byun SS, Heo TS, Choi JM, Jeong YS, Kim YS, Lee WK, et al. Deep learning based prediction of prognosis in nonmetastatic clear cell 1331 renal cell carcinoma. *Sci. Rep.* (2021) 11:1242. doi: 10.1038/S41598-020-80262-9
173. Xi IL, Zhao Y, Wang R, Chang M, Purkayastha S, Chang K, et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. *Clin Cancer Res.* (2020) 26:1944–52. doi: 10.1158/1078-0432.CCR-19-0374
174. Greenberg JW, Higyoung K, Moustafa AA, Pedro C, Asim AM. Krane Spencer: a dual drug therapy for sunitinib resistant RCC: An in vitro analysis. *J Clin Oncol.* (2021) 39:340–340. doi: 10.1200/JCO.2021.39.6_suppl.340
175. Chiu YC, Chen HH, Zhang T, Zhang S, Gorthi A, Wang LJ, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics.* (2019) 12:18. doi: 10.1186/s12920-018-0460-9
176. Ataei A, Majidi NS, Zahiri J, Rostami R, Arab SS, Rizvanov AA. Prediction of chemoresistance trait of cancer cell lines using machine learning algorithms and systems biology analysis. *J Big Data.* (2021) 8:97. doi: 10.1186/s40537-021-00477-z
177. U.S Food and Drug Administration. *FDA Releases Artificial Intelligence/Machine Learning Action Plan [Press release]*. (2021). Available online at: https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan
178. U.S Food and Drug Administration. *FDA Authorizes Marketing of First Device that Uses Artificial Intelligence to Help Detect Potential Signs of Colon Cancer, in medical device aids clinicians in detecting potential*

irregularities during colon cancer screening and surveillance [Press release]/ (2021). Available online at: https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-device-uses-artificial-intelligence-help-detect-potential-signs-colon
179. Matthews TP, Singh S, Mombourquette B, Su J, Shah MP, Pedemonte S, et al. A multisite study of a breast density deep learning model for full-field digital mammography and synthetic mammography. *Radiol Artif Intell*. (2020) 3:e200015. doi: 10.1148/ryai.2020200015
180. Fujioka T, Mori M, Kubota K, Oyama J, Yamaga E, Yashima Y, et al. The utility of deep learning in breast ultrasonic imaging: a review. *Diagnostics Basel Switzerland*. (2020) 10:1055. doi: 10.3390/diagnostics10121055
181. Jain T, Sharma P, Are AC, Vickers SM, Dudeja V. New insights into the cancer-microbiome-immune axis: decrypting a decade of discoveries. *Front Immunol*. (2021) 12:622064. doi: 10.3389/fimmu.2021.622064
182. Ciernikova S, Mego M, Chovanec M. Exploring the potential role of the gut microbiome in chemotherapy-induced neurocognitive disorders and cardiovascular toxicity. *Cancers*. (2021) 13:782. doi: 10.3390/cancers13040782
183. Chamber LM, Esakov EL, Braley C, Tretan L, Alali Z, Bayik D, et al. Disruption of the gut microbiota attenuates epithelial ovarian cancer sensitivity to cisplatin therapy. *bioRxiv*. (2020) 13:782. doi: 10.21203/rs.3.rs-80626/v1
184. Jacobson D, Moore K, Gunderson C, Rowland M, *Austin* R, Honap TP, et al. Shifts in gut and vaginal microbiomes are associated with cancer recurrence time in women with ovarian cancer. *PeerJ*. (2021) 9:e11574. doi: 10.7717/peerj.11574
185. Khan S, Ignacio G, Marta C, Giuseppe B, Francesco P, Anna PC, Cancer and the microbiome: potential applications as new tumor biomarker. *Expert Rev Anticancer Ther*. (2015) 3:317–330. doi: 10.1586/14737140.2015.992785
186. Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol*. (2017) 15:109–28. doi: 10.1038/nrmicro.2016.171
187. Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, et al. The microbiome and human biology. *Annu Rev Genomics Hum Genet*. (2017) 18:65–86. doi: 10.1146/annurev-genom-083115-022438
188. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell*. (2019) 178:795–806.e12. doi: 10.1016/j.cell.2019.07.008
189. Lauka L, Reitano E, Carra MC, Gaiani F, Gavriilidis P, Brunetti F, et al. Role of the intestinal microbiome in colorectal cancer surgery outcomes. *World J Surg Oncol*. (2019) 17:204. doi: 10.1186/s12957-019-1754-x
190. Wang J, Li X, Wu X, Wang Z, Zhang C, Cao G, et al. Uncovering the microbiota in renal cell carcinoma tissue using 16S rRNA gene sequencing. *J Cancer Res Clin Oncol*. (2021) 147:481–91. doi: 10.1007/s00432-020-03462-w
191. Oliva M, Mulet-Margalef N, Ochoa-De-Olza M, Napoli S, Mas J, Laquente B, et al. Tumor-associated microbiome: where do we stand? *Int J Mol Sci*. (2021) 22:1446. doi: 10.3390/ijms22031446
192. Cimadamore A, Cheng M, Santoni M, Lopez-Beltran A, Battelli N, Massari F, et al. New prostate cancer targets for diagnosis, imaging, and therapy: focus on prostate-specific membrane antigen. *Front Oncol*. (2018) 8:653. doi: 10.3389/fonc.2018.00653
193. Rini BI, Battle D, Figlin RA, George DJ, Hammers Hans, Hutson T, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of advanced renal cell carcinoma (RCC). *J Immunotherapy Cancer*. (2019) 7:354. doi: 10.1186/s40425-019-0813-8
194. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. (2018) 359:91–97. doi: 10.1126/science.aan3706
195. Derosa L, Hellmann MD, Spaziano M, Halpenny D, Fidelle M, Rizvi H, et al. Negative association of antibiotics on clinical activity of immune checkpoint inhibitors in patients with advanced renal cell and non-small-cell lung cancer. *Ann Oncol*. (2018) 29:1437–44. doi: 10.1093/annonc/mdy103
196. Cimadamore A, Santoni M, Massari F, Gasparrini S, Cheng L, Lopez-Beltran A, et al. Microbiome and cancers, with focus on genitourinary tumors. *Front Oncol*. (2019) 9:178. doi: 10.3389/fonc.2019.00178
197. Reid G. Microbes in food to treat and prevent disease. *Expert Rev Precis Med Drug Develop*. (2018) 2:79–81. doi: 10.1080/23808993.2018.1429217
198. Derosa L, Routy B, Fidelle M, Iebba V, Alla L, Pasolli E, et al. Gut bacteria composition drives primary resistance to cancer immunotherapy in renal cell carcinoma patients. *Eur Urol*. (2020) 78:195–206. doi: 10.1016/j.eururo.2020.04.044
199. Salgia NJ, Bergerot PG, Maia MC, Dizman N, Hsu J, Gillece JD, et al. Stool microbiome profiling of patients with metastatic renal cell carcinoma receiving anti-PD-1 immune checkpoint inhibitors. *Eur Urol*. (2020) 78:498–502. doi: 10.1016/j.eururo.2020.07.011
200. Liss MA, Chen Y, Rodriguez R, Pruthi D, Johnson-Pais T, Wang H, et al. Microbiome within primary tumor tissue from renal cell carcinoma may be associated with PD-L1 expression of the venous tumor thrombus. *Adv Urol*. (2020) 2020:9068068. doi: 10.1155/2020/9068068
201. Gong TT, He XH, Gao S, Wu QJ. Application of machine learning in prediction of chemotherapy resistant of ovarian cancer based on gut microbiota. *J Cancer*. (2021) 12:2877–85. doi: 10.7150/jca.46621
202. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. (2014) 10:766. doi: 10.15252/msb.20145645
203. Ren Z, Wang H, Cui G, Lu H, Wang L, Luo H, et al. Alterations in the human oral and gut microbiomes and lipidomics in COVID-19. *Gut*. (2021) 70:1253–65. doi: 10.1136/gutjnl-2020-323826
204. Lapidot Y, Amir A, Nosenko R, Uzan-Yulzari A, Veitsman E, Cohen-Ezra O, et al. Alterations in the gut microbiome in the progression of cirrhosis to hepatocellular carcinoma. *mSystems*. (2020) 5:e00153–20. doi: 10.1128/mSystems.00153-20
205. Coley WB. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin Orthop Relat Res*. (1991) 2623–11.
206. Xu W, Jiang X, Guan C, Gu M. The prognostic and predictive value of tumor infiltrating Macrophage and Neutrophil in patient with clear cell renal cell carcinoma: Tumor infiltrating lymphocytes in renal cell carcinoma. *Medicine (Baltimore)*. (2020) 99:e23181. doi: 10.1097/MD.0000000000023181
207. Faiena I, Comin-Anduix B, Berent-Maoz B, Bot A, Zomorodian N, Sachdeva A, et al. A phase I, open-label, dose-escalation, and cohort expansion study to evaluate the safety and immune response to autologous dendritic cells transduced with AdGMCA9 (DC-AdGMCAIX) in patients with metastatic renal cell carcinoma. *J Immunother*. (2020) 43:273–282. doi: 10.1097/CJI.0000000000000336
208. Zhang Z, Lu M, Qin Y, Gao W, Tao L, Su W, et al. Neoantigen: A New Breakthrough in Tumor Immunotherapy. *Front Immunol*. (2021) 12:672356. doi: 10.3389/fimmu.2021.672356
209. Dushenkov A, Jungsuwadee P. Chimeric antigen receptor T-cell therapy: Foundational science and clinical knowledge for pharmacy practice. *J Oncol Pharm Pract*. (2019) 25:1217–25. doi: 10.1177/1078155219836480
210. Richards, M. FDA Approves First Cell-Based Gene Therapy For Adult Patients with Relapsed or Refractory MCL. (2020). Available online at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-cell-based-gene-therapy-adult-patients-relapsed-or-refractory-mcl (accessed June 3, 2021)
211. Song J, Yu Z, Dong B, Zhu M, Guo X, Ma Y, et al. Clinical significance of circulating tumour cells and Ki-67 in renal cell carcinoma. *World J Surg Oncol*. (2021) 19:156. doi: 10.1186/s12957-021-02268-5
212. Nayak B, Panaiyadiyan S, Singh P, Karmakar S, Kaushal S, Seth A. Role of circulating tumor cells in patients with metastatic clear-cell renal cell carcinoma. *Urologic Oncol*. (2021) 39:135.e9–135.e15. doi: 10.1016/j.urolonc.2020.10.077
213. Lallo A, Schenk MW, Frese KK, Blackhall F, Dive C. Circulating tumor cells and CDX models as a tool for preclinical drug development. *Transl Lung Cancer Res*. (2017) 6:397–408. doi: 10.21037/tlcr.2017.08.01
214. Gao D, Vela I, Sboner A, Iaquinta PJ, Karthaus WR, Gopalan A, et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell*. (2014) 159:176–87. doi: 10.1016/j.cell.2014.08.016
215. Weeber F, van de Wetering M, Hoogstraat M, Dijkstra KK, Krijgsman O, Kuilman T, et al. Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *Proc Natl Acad Sci U S A*. (2015) 112:13308–11. doi: 10.1073/pnas.1516689112
216. Tellez-Gabriel M, Cochonneau D, Cadé M, Jubellin C, Heymann MF, Heymann D. Circulating tumor cell-derived pre-clinical models for

personalized medicine. *Cancers.* (2018) 11:19. doi: 10.3390/cancers11010019

217. Hidalgo M, Bruckheimer E, Rajeshkumar NV, Garrido-Laguna I, De Oliveira E, Rubio-Viqueira B, et al. pilot clinical study of treatment guided by personalized tumorgrafts in patients with advanced cancer. *Mol Cancer Ther.* (2011) 10:1311–6. doi: 10.1158/1535-7163.MCT-11-0233
218. Wang X, Lopez R, Luchtel RA, Hafizi S, Gartrell B, Shenoy N. Immune evasion in renal cell carcinoma: biology, clinical translation, future directions. *Kidney Int.* (2021) 99:75–85. doi: 10.1016/j.kint.2020.08.028
219. Francisco LM, Salinas VH, Brown KE, Vanguri VK, Freeman GJ, Kuchroo VK, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med.* (2009) 206:3015–29. doi: 10.1084/jem.20090847
220. Mann SA, Lopez-Beltran A, Massari F, Pili R, Fiorentino M, Koch MO, et al. Targeting the programmed cell death-1 pathway in genitourinary tumors: current progress and future perspectives. *Curr Drug Metab.* (2017) 18:700–11. doi: 10.2174/1389200218666170518162500
221. Liu J, Zhang C, Hu J, Tian Q, Wang X, Gu H, et al. Effectiveness of anti-PD-1/PD-L1 antibodies in urothelial carcinoma patients with different PD-L1 expression levels: a meta-analysis. *Oncotarget.* (2018) 9:12400–7. doi: 10.18632/oncotarget.24249
222. Raimondi A, Sepe P, Zattarin E, Mennitto A, Stellato M, Claps M, et al. Predictive biomarkers of response to immunotherapy in metastatic renal cell cancer. *Front Oncol.* (2020) 10:1644. doi: 10.3389/fonc.2020.01644
223. Kumagai S, Togashi Y, Kamada T, Sugiyama E, Nishinakamura H, Takeuchi Y, et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nat Immunol.* (2020) 21:1346–58. doi: 10.1038/s41590-020-0769-3
224. Ravi P, Mantia C, Su C, Sorenson K, Elhag D, Rathi N, et al. Evaluation of the safety and efficacy of immunotherapy rechallenge in patients with renal cell carcinoma. *JAMA Oncol.* (2020) 6:1606–10. doi: 10.1001/jamaoncol.2020.2169
225. Kulshreshtha C, Venkatesh MP, Kumar PTM. The path of personalized medicine: regulatory perspective. *Int J Drug Regul Affair.* (2018) 3:14–29. doi: 10.22270/ijdra.v3i1.155
226. Personalised Medicine; medicinal products. Available online at: https://ec.europa.eu/health/human-use/personalised-medicine_en (accessed June 3, 2021)
227. Knowles L, Luth W, Bubela T. Paving the road to personalized medicine: recommendations on regulatory, intellectual property and reimbursement challenges. *J Law Biosci.* (2017) 4:453–506. doi: 10.1093/jlb/lsx030
228. Konski AF. The Era of Personalized Medicine Has Arrived - PMC's Annual Progress and Outlook Report. (2019). Available online at: https://www.foley.com/en/insights/publications/2019/08/the-era-of-personalized-medicine-has-arrived (accessed June 4, 2021)
229. Cruz-Correia R, Ferreira D, Bacelar G, Marques P, Maranhao P. Personalised medicine challenges: quality of data. *Int J Data Sci Anal.* (2018) 6:251–9. doi: 10.1007/s41060-018-0127-9
230. Brothers KB, Rothstein MA. Ethical, legal and social implications of incorporating personalized medicine into healthcare. *Per Med.* (2015) 12:43–51. doi: 10.2217/pme.14.65
231. Armstrong S. Data, data everywhere: the challenges of personalised medicine. *BMJ.* (2017) 359:j4546. doi: 10.1136/bmj.j4546
232. Council on Ethical and Judicial Affairs, American Medical Association. AMA Code of Medical Ethics' opinions on genetic testing: opinion 2.131 - disclosure of familial risk in genetic testing. *AMA J Ethics.* (2009) 11:683–685. doi: 10.1001/virtualmentor.2009.11.9.code1-0909
233. ASHG statement. Professional disclosure of familial genetic information. The American Society of Human Genetics Social Issues Subcommittee on Familial Disclosure. *Am J Hum Genet.* (1998) 62:474–83. doi: 10.1086/301707
234. Patrinos GP, Mitropoulou C. Measuring the value of pharmacogenomics evidence. *Clin Pharmacol Ther.* (2017) 102:739–41. doi: 10.1002/cpt.743
235. Kichko K, Marschall P, Flessa S. Personalized medicine in the U.S. and Germany: awareness, acceptance, use and preconditions for the wide implementation into the medical standard. *J Pers Med.* (2016) 6:15. doi: 10.3390/jpm6020015
236. Prasad V, Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Intern Med.* (2017) 177:1569–75. doi: 10.1001/jamainternmed.2017.3601
237. Bergman MJ, Kivitz AJ, Pappas DA, Kremer JM, Zhang L, Jeter A, et al. Clinical utility and cost savings in predicting inadequate response to anti-TNF therapies in rheumatoid arthritis. *Rheumatol Ther.* (2020) 7:775–92. doi: 10.1007/s40744-020-00226-3
238. Parker JL, Lushina N, Bal PS, Petrella T, Dent R, Lopes G. Impact of biomarkers on clinical trial risk in breast cancer. *Breast Cancer Res Treat.* (2012) 136:179–85. doi: 10.1007/s10549-012-2247-6
239. Falconi A, Lopes G, Parker JL. Biomarkers and receptor targeted therapies reduce clinical trial risk in non-small-cell lung cancer. *J Thorac Oncol.* (2014) 9:163–9. doi: 10.1097/JTO.0000000000000075
240. Parker JL, Zhang ZY, Buckstein R. Clinical trial risk in Non-Hodgkin's lymphoma: endpoint and target selection. *J Pharm Pharm Sci.* (2011) 14:227–35. doi: 10.18433/J39P45
241. Antoniou M, Kolamunnage-Dona R, Wason J, Bathia R, Billingham C, Bliss JM, et al. Biomarker-guided trials: challenges in practice. *Contemp Clin Trials Commun.* (2019) 16:100493. doi: 10.1016/j.conctc.2019.100493
242. PMC The Personalised Medicine Report 2020 Opportunity, Challenegs and the Future. (2020).
243. Ciccarese C, Brunelli M, Montironi R, Fiorentino M, Iacovelli R, Heng D, et al. The prospect of precision therapy for renal cell carcinoma. *Cancer Treat Rev.* (2016) 49:37–44. doi: 10.1016/j.ctrv.2016.07.003

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

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

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 27 July 2022
DOI 10.3389/fmed.2022.830621

OPEN ACCESS

EDITED BY
Alice Chen,
National Cancer Institute (NCI),
United States

REVIEWED BY
Yi Chiung Hsu,
National Central University, Taiwan
Chia-Hung Yen,
Kaohsiung Medical University, Taiwan

*CORRESPONDENCE
Shih-Yin Chen
chenshihy@gmail.com
Fuu-Jen Tsai
d0704@mail.cmuh.org.tw

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

RECEIVED 07 December 2021
ACCEPTED 27 June 2022
PUBLISHED 27 July 2022

CITATION
Liu T-Y, Liao W-L, Wang T-Y, Chan C-J, Chang J-G, Chen Y-C, Lu H-F, Yang H-H, Chen S-Y and Tsai F-J (2022) Genome-wide association study of hyperthyroidism based on electronic medical record from Taiwan. *Front. Med.* 9:830621. doi: 10.3389/fmed.2022.830621

Genome-wide association study of hyperthyroidism based on electronic medical record from Taiwan

Ting-Yuan Liu[1,2†], Wen-Ling Liao[3,4,5†], Tzu-Yuan Wang[6,7†], Chia-Jung Chan[5], Jan-Gowth Chang[1], Yu-Chia Chen[1,2], Hsing-Fang Lu[2], Hsien-Hui Yang[8], Shih-Yin Chen[5,9*] and Fuu-Jen Tsai[5,9,10*]

[1]Center for Precision Medicine, China Medical University Hospital, Taichung, Taiwan, [2]Million-Person Precision Medicine Initiative, Department of Medical Research, China Medical University Hospital, Taichung, Taiwan, [3]College of Chinese Medicine, Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan, [4]Center for Personalized Medicine, China Medical University Hospital, Taichung, Taiwan, [5]Genetics Center, Medical Research, China Medical University Hospital, Taichung, Taiwan, [6]Department of Internal Medicine, College of Medicine, China Medical University, Taichung, Taiwan, [7]Division of Endocrinology, China Medical University Hospital, Taichung, Taiwan, [8]China Medical University Beigang Campus, Yunlin, Taiwan, [9]School of Chinese Medicine, China Medical University, Taichung, Taiwan, [10]Department of Medical Genetics, China Medical University Hospital, Taichung, Taiwan

Excess thyroid hormones have complex metabolic effects, particularly hyperthyroidism, and are associated with various cardiovascular risk factors. Previous candidate gene studies have indicated that genetic variants may contribute to this variable response. Electronic medical record (EMR) biobanks containing clinical and genomic data on large numbers of individuals have great potential to inform the disease comorbidity development. In this study, we combined electronic medical record (EMR) -derived phenotypes and genotype information to conduct a genome-wide analysis of hyperthyroidism in a 35,009-patient cohort in Taiwan. Diagnostic codes were used to identify 2,767 patients with hyperthyroidism. Our genome-wide association study (GWAS) identified 44 novel genomic risk markers in 10 loci on chromosomes 2, 6, and 14 ($P < 5 \times 10{-}14$), including CTLA4, HCP5, HLA-B, POU5F1, CCHCR1, HLA-DRA, HLA-DRB9, TSHR, RPL17P3, and CEP128. We further conducted a comorbidity analysis of our results, and the data revealed a strong correlation between hyperthyroidism patients with thyroid storm and stroke. In this study, we demonstrated application of the PheWAS using large EMR biobanks to inform the comorbidity development in hyperthyroidism patients. Our data suggest significant common genetic risk factors in patients with hyperthyroidism. Additionally, our results show that sex, body mass index (BMI), and thyroid storm are associated with an increased risk of stroke in subjects with hyperthyroidism.

KEYWORDS

genome-wide association study (GWAS), phenome-wide association studies (PheWAS), hyperthyroidism, electronic medical record (EMR), stroke

Introduction

Hyperthyroidism is a common endocrine disorder with a prevalence of ∼0.3–0.5% in an iodine-replete area (1, 2). Excessive amounts of thyroid hormones have profound effects on the cardiovascular system (3). Hyperthyroidism can cause increased heart rate, contractility, wide pulse pressure, systolic hypertension, changes in peripheral vascular resistance, and predisposition to dysrhythmias (3, 4).

In Taiwan, the prevalence of hyperthyroidism is ∼2% (5). Autoimmune thyroid diseases account for 40–70% of hyperthyroidism sufferers, including Graves' disease and Hashimoto's thyroiditis. The remainder includes hyperfunctioning thyroid adenomas, subacute thyroiditis, thyroid cancer, and pituitary tumors (6, 7).

Although hyperthyroidism may involve both short- and long-term cardiovascular consequences (8), data concerning the association between hyperthyroidism and cardiovascular outcomes are inconsistent (9). Thyroid dysfunction, which leads to effects on the cardiovascular system and increases an individual's risk of death, is currently under debate. In particular, there is few data to demonstrate that hyperthyrodism increase the risk of stroke in young adults.

It is well-known that genetic factors play an important role in disease etiology and pathogenesis (10). Genetic diseases result from the accumulation of genetic alterations. Therefore, genetic alterations could serve as effective biomarkers for the early detection, monitoring, and prognosis of genetic diseases. In the present study, we summarize the accumulation and achievements of big genomic data and show how they could contribute to precision medicine by using hyperthyrodism as a genetic disease model.

Methods

Data sources and informed consent

The China Medical University Hospital Precision Medicine Project was initiated in 2018 and remains operational. This project was approved by the ethical committees of CMUH (CMUH107-REC3-058 and CMUH110-REC3-005). Thus, far, more than 170,000 people have contributed. In this study, all clinical information was collected from the electronic medical records (EMRs) of CMUH and approved by the respective ethical committees of CMUH (CMUH110-REC1-095). The EMR data were collected between 1992 and 2019.

SNP array data quality control

We used the TPMv1 customized SNP array (Thermo Fisher Scientific, Inc., Santa Clara, CA, USA), which was designed from the Academia Sinica and Taiwan Precision Medicine Initiative (TPMI) teams. The SNP array contained approximately 714,457 SNPs. PLINK1.9 (11) was used for the analysis. We excluded subjects and SNPs with missing rates (subjects excluded: missingness per marker –geno 0.1 > 10% for SNPs and missingness per individual –mind 0.1 >10% for subjects). We filtered out variants with a Hardy–Weinberg equilibrium $p < 10^{-6}$ (–hwe 10^{-6}) and minor allele frequency (MAF) of $<10^{-4}$ (–maf 0.0001). Therefore, 508,004 variants and 173,135 subjects passed the filters and the quality control process; then, we used Beagle 5.2 to impute. The imputed data were filtered out using an alternate allele dose of <0.3 and a genotype posterior probability of <0.9 as the criteria (12). After the quality control and imputation process, we analyzed 13,034,044 variants and 173,135 subjects (13) (Figure 1).

Genome-wide association study

We used PLINK 1.9 for the summary statistics. Subjects who had been diagnosed with hyperthyroidism three times in the EMR were defined as cases. These patients also included those taking medications for hyperthyroidism. Data included values from thyroid-related tests (free T4 and TSH). Subjects who had never been diagnosed with thyroid-related disease in the EMR were defined as controls. There were no abnormal values in the thyroid-related tests. We kept only one person from the same family in the control and case groups. We determined the members of the same family based on the results of Identical-by-state (IBS)/Identical-by-descent (IBD) (IBS/IBD > 0.25: is the coefficient used to calculate kinship, which we used to exclude people from the same family to ensure independence between samples) computation using PLINK 1.9 (–logistic, –covar sex and PC1∼PC4). Finally, we randomly divided the subjects into two cohorts (70% for discovery cohort, 30% for replication cohort), divided the subjects into two groups (cases and controls) based on clinical annotation. There were 1,850 cases and 21,499 controls in the discovery cohort. There were 917 cases and 10,743 controls in the replication cohort. Logistic regression with multiple covariates was used to analyze the data. The covariates used in the logistic regression were sex and PC1–PC4. PC1 to PC4 were the results of principal component analysis (PCA) analysis using PLINK 1.9. We also adjusted for statistical significance. We plotted the Manhattan plot and quantile-quantile (QQ) plot with the *p*-value using R studio.

Statistical analysis

Statistical analysis was performed according to our previous study (14). Comparisons between two groups was performed using the Student's *t*-test. Statistical comparisons of more than two groups were performed using one-way analysis of

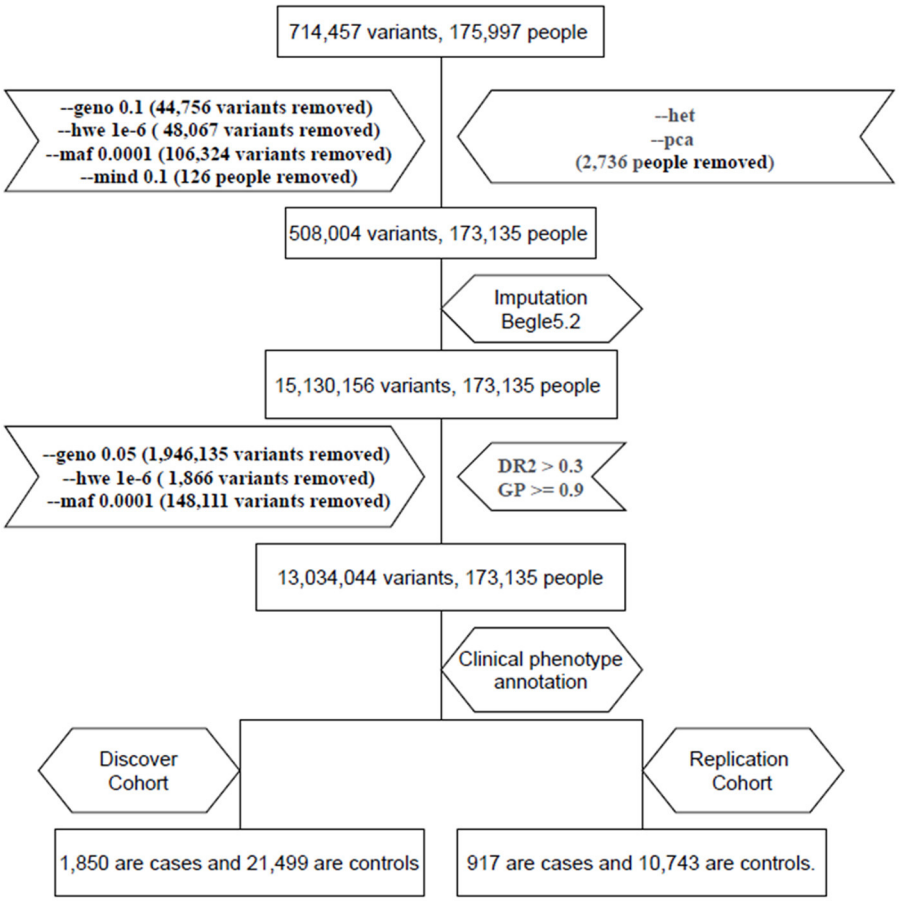

FIGURE 1
Flow chart of electronic medical record (EMR) data mining and the genome-wide association studies (GWAS) (GWAS) pipeline. We enrolled two cohorts: one with 1,850 hyperthyroidism patients and 21,499 age and sex matched individuals and a replication cohort with 917 hyperthyroidism patients and 10,743 age and sex matched individuals.

variance (ANOVA). In all cases, $p < 0.05$, was considered to be statistically significant.

Result

We followed a flowchart for EMR data mining and the GWAS pipeline (Figure 1) and an Abstract Graph representation for research concept is shown in Figure 2. GWAS analyses with hyperthyroidism were performed on the discovery batch of 23,349 individuals included from 1992. The replication batch consisted of 11,660 individuals recruited from 1992 (Table 1). The same exclusion criteria were applied to both batches.

The total genotyping rate of the remaining samples was 0.992366. A total of 3,034,044 variants and 35,009 people passed the filters and quality control among the remaining phenotypes. There were 23,349 people, including 1,850 cases and 21,499 controls in the discovery cohort, and 11,660 people, including 917 cases and 10,743 controls in the replication cohort. Manhattan plot is used to visualize GWAS analysis. The genome-wide significance level was set at $p = 5 \times 10^{-8}$ in the discovery batch (upper red line, Figure 3A) and $p = 1.75 \times 10^{-5}$ in the replication batch (upper red line, Figure 3C). The association of these single nucleotide polymorphisms (SNPs) that passed quality control are plotted on the X-axis according to their chromosomal positions against Y-axis (- $\log_{10}$ p-value). We also used QQ plots for genome-wide association analysis to investigate the correlation between hyperthyroidism patients and controls in the discovery and replication cohorts (Figures 3B,D). As new loci within the hyperthyroidism patients were identified by GWAS, we proved the associations with allelic variants of these new loci in linkage disequilibrium were shown to be stronger than previously observed associations (Supplemental material 1, SP1).

In this present study, we identified more than 1,500 SNPs associate with hyperthyroidism (SP2). For reducing the numbers and focused on the most significantly top 10 genes. Our GWAS data identified 44 novel genomic risk markers in 10 loci on

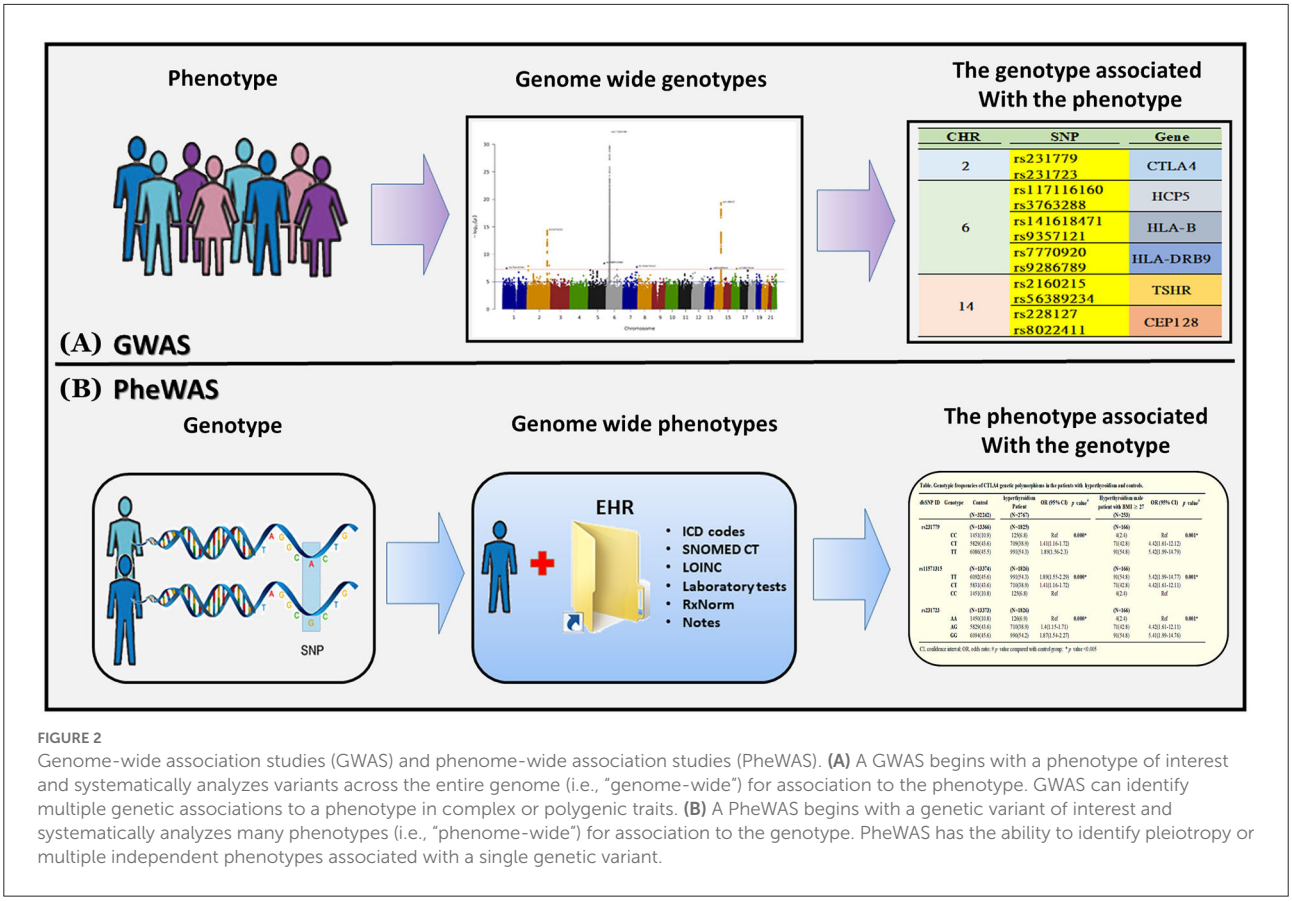

FIGURE 2
Genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS). **(A)** A GWAS begins with a phenotype of interest and systematically analyzes variants across the entire genome (i.e., "genome-wide") for association to the phenotype. GWAS can identify multiple genetic associations to a phenotype in complex or polygenic traits. **(B)** A PheWAS begins with a genetic variant of interest and systematically analyzes many phenotypes (i.e., "phenome-wide") for association to the genotype. PheWAS has the ability to identify pleiotropy or multiple independent phenotypes associated with a single genetic variant.

TABLE 1 Descriptive information on the discovery and replication batches.

Sample characteristics		
Batch name	**Discovery ($n = 23,349$)**	**Replication ($n = 11,660$)**
Sex (M/F) (%)	10,289 (44.1) / 13,060 (55.9)	5,199 (44.6)/6,461 (55.4)
Age (SD)	50.0 (19.497)	49.87 (19.372)
BMI (SD)	25.95 (5.991)	25.91 (5.905)
Hyperthyroidism	1,850 (7.9)	917 (7.9)

Genome-wide association studies (GWAS) analyses with hyperthyroidism were performed on a discovery batch comprising 23,349 individuals included from 1992 to 2019. The replication batch consisted of 11,660 individuals recruited from 1992 to 2019. The same exclusion criteria were applied to both batches.

chromosomes 2, 6, and 14 with the threshold of $p < 5 \times 10^{-14}$ in discovery analysis and $p < 1.75 \times 10^{-5}$ in replication analysis, including genes CTLA4, HCP5, HLA-B, POU5F1, CCHCR1, HLA-DRA, HLA-DRB9, TSHR, RPL17P3, and CEP128. The genes that showed a significant difference in our study have all been considered in the involvement of disease in previous studies (Table 2).

Briefly, ***CTLA4*:** Function of gene: Inhibitory receptor acting as a major negative regulator of T-cell responses (15, 16). Involvement in disease: Systemic lupus erythematosus (SLE) (17), diabetes mellitus, insulin-dependency, 12 (IDDM12) (18), celiac disease 3 (CELIAC3) (19), autoimmune lymphoproliferative syndrome 5 (ALPS5) (20). ***HCP5*:** Function of gene: *HCP5* (HLA Complex P5) is an RNA gene and is affiliated with the lncRNA class. Involvement in disease: acquired immunodeficiency syndrome (21), thyroid glandular carcinoma (22). ***HLA-B*:** Function of gene: *HLA-B* (major histocompatibility complex, class I, B) is a protein-coding gene. Involvement in disease: Stevens-Johnson syndrome (SJS) (23) and Spondyloarthropathy 1 (SPDA1) (24). ***POU5F1*:** Function of gene: Critical for early embryogenesis and embryonic stem cell pluripotency (25). Involvement in disease: Embryonal carcinoma (26), Teratoma (27). ***CCHCR1*:** Function of gene: Critical for early embryogenesis and embryonic stem cell pluripotency. Involvement in disease: Psoriasis (28), Psoriasis 1 (PSORS1) (29). ***HLA-DRA*:** Function of gene: HLA-DRA (major histocompatibility complex, class II, DR alpha) is a protein-coding gene. Involvement in disease: Graham-Little-Piccardi-Lassueur Syndrome (30), Penicillin Allergy (31). ***HLA-DRB9*:** Function of gene: *HLA-DRB9* [major histocompatibility complex, class II, DR beta 9 (pseudogene)]

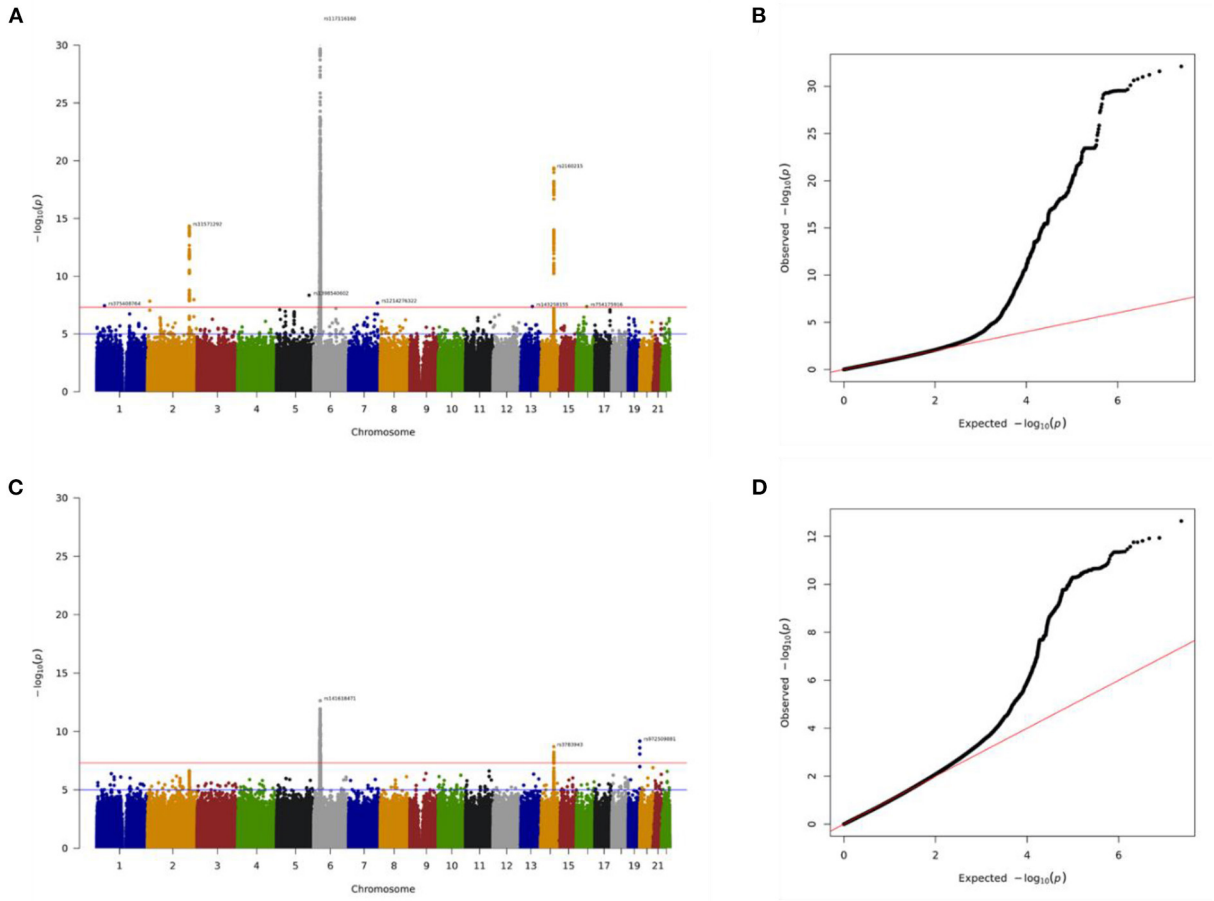

FIGURE 3
Association of genome-wide variants with hyperthyroidism diagnosed in discovery batch **(A,B)** and replication **(C,D)** batch using Manhattan plot **(A,C)** and QQ plot **(B,D)** analysis. In Manhattan plot analysis, single nucleotide polymorphism (SNP) that passed quality control are plotted on the X-axis according to their chromosomal positions against Y-axis (- $\log_{10}$ p-value). The upper and lower dotted lines indicate the genome-wide significance threshold ($p = 5.0\times10^{-8}$) and the cut-off level for selecting SNPs for replication study ($p = 1.75\times10^{-5}$), respectively.

is a pseudogene. Involvement in disease: rheumatoid arthritis (32), Vogt-Koyanagi-Harada disease (33), multiple sclerosis (34). ***TSHR*:** Function of gene: plays a central role in controlling thyroid cell metabolism (by similarity) (35). Involvement in disease: Hypothyroidism, congenital, non-goitrous, 1 (CHNG1) (36), familial gestational hyperthyroidism (HTFG) (37), hyperthyroidism, non-autoimmune (HTNA) (38). ***RPL17P3*:** Function of gene: *RPL17P3* (ribosomal protein L17 pseudogene 3) is a pseudogene. Involvement in disease: thyroid (39). ***CEP128*:** Function of gene: *CEP128* (Centrosomal Protein 128) is a protein-coding gene. Involvement in disease: Hypothyroidism, Congenital, Nongoitrous, 1 (CHNG1) (40), Hyperthyroidism, Non-autoimmune (HTNA) (41).

Based on the prevalence of comorbidities among our study population, we further conducted a comorbidity analysis of our results using EMR data. A total of 2,767 subjects with a hyperthyroidism diagnosis (International Classification of Diseases, 9th Revision, Clinical Modification [ICD-9-CM] 242.90, 242.00, 242.900 or ICD10-code E05.0), with at least one TSH and free T4 or total T4 value, and with genotyping information were identified as subjects with hyperthyroidism (the case group). The gender were grouped by the available data in the study. Obesity in this study was defined as body mass index (BMI) $\geq$ 27 kg/m^2, according to the Ministry of Health and Welfare of Taiwan. As shown as Supplementary material 3 (SP3), the extracted comorbidities were defined by the studies with disease diagnosis (ICD code). We observed that the incidence of thyroid storm in hyperthyroidism individuals was 1.3% (36/2767). We also observed that the risk of stroke in male individuals with hyperthyroidism was significantly higher than that female individuals with hyperthyroidism ($p < 0.05$, Table 3). Similar results were observed for stroke, heart disease, diabetes, and hypertension with statistical significance ($p < 0.05$, Table 4). Compared with normal body weight, individuals

TABLE 2 Lead SNPs from the discovery- and replication-analysis.

CHR	SNP	A1/A2	*p*-Value (GWAS results)	*p*-Value (Replication)	Gene	Function of gene	Involvement in disease
2	rs1427680	G/A	2.07E-14	8.15E-06	CTLA4	Inhibitory receptor acting as a major negative regulator of T-cell responses (15, 16).	Systemic lupus erythematosus (SLE) (17)
	rs736611	C/T	2.19E-14	9.96E-06			Diabetes mellitus, insulin-dependent, 12 (IDDM12) (18)
	rs11571315	T/C	2.19E-14	9.43E-06			Celiac disease 3 (CELIAC3) (19)
	rs231723	A/G	3.24E-14	9.25E-06			Autoimmune lymphoproliferative syndrome 5 (ALPS5) (20)
6	rs117116160	C/T	7.73E-33	1.23E-12	HCP5	HCP5 (HLA Complex P5) is an RNA Gene, and is affiliated with the lncRNA class.	Acquired immunodeficiency syndrome (21)
	rs117884751	T/A	2.56E-32	1.76E-12			Thyroid gland follicular carcinoma (22)
	rs3763287	C/A	5.25E-25	1.41E-11			
	rs114202986	T/A	4.60E-24	3.45E-10			
	rs3763288	G/A	5.04E-24	3.88E-10			
	rs141618471	A/G	1.01E-31	2.31E-13	HLA-B	HLA-B (Major Histocompatibility Complex, Class I, B) is a Protein Coding gene.	Stevens-Johnson syndrome (SJS) (23)
	rs9378228	G/T	1.70E-31	4.48E-12			Spondyloarthropathy 1 (SPDA1) (24)
	rs12524692	T/A	7.74E-29	1.76E-12			
	rs72860306	C/T	5.84E-28	2.16E-11			
	rs9357121	T/G	9.24E-22	4.98E-12			
	rs117588763	C/T	2.88E-30	3.55E-11	POU5F1	Critical for early embryogenesis and for embryonic stem cell pluripotency (25).	Embryonal carcinoma (26)
	rs9357112	A/G	2.88E-30	3.55E-11			Teratoma (27)
	rs9357114	T/G	2.88E-30	3.55E-11			
	rs9348855	A/C	2.88E-30	3.55E-11			
	rs4713439	A/G	3.47E-30	4.10E-11			
	rs28652698	G/A	3.57E-28	2.06E-11	CCHCR1	Critical for early embryogenesis and for embryonic stem cell pluripotency.	Psoriasis (28)
	rs28383832	(-/CGCC)	1.70E-20	6.47E-10			Psoriasis 1 (PSORS1) (29)
	rs1265082	G/A	1.84E-20	4.21E-10			
	rs1265113	C/G	1.84E-20	4.21E-10			

(Continued)

TABLE 2 Continued

CHR	SNP	A1/A2	*p*-Value (GWAS results)	*p*-Value (Replication)	Gene	Function of gene	Involvement in disease
	rs9469112	C/T	3.30E-26	1.71E-10	HLA-DRA	HLA-DRA (Major Histocompatibility Complex, Class II, DR Alpha) is a Protein Coding gene.	Graham-little-piccardi-lassueur syndrome (30)
	rs16822660	T/C	1.36E-22	1.94E-08			Penicillin allergy (31)
	rs9469113	G/A	2.50E-21	1.55E-07			
	rs7770920	T/A	3.51E-24	1.70E-10	HLA-DRB9	HLA-DRB9 [Major Histocompatibility Complex, Class II, DR Beta 9 (Pseudogene)] is a Pseudogene.	Rheumatoid arthritis (32)
	rs6457596	C/T	3.51E-24	1.70E-10			Vogt-koyanagi-harada disease (33)
	rs111573974	(-/G)	3.51E-24	1.70E-10			Multiple sclerosis (34)
	rs6924760	A/G	3.52E-24	1.71E-10			
	rs9286789	T/G	3.52E-24	1.71E-10			
14	rs2160215	T/A	4.45E-20	8.99E-09	TSHR	Plays a central role in controlling thyroid cell metabolism (By similarity) (35).	Hypothyroidism, congenital, non-goitrous, 1 (CHNG1) (36)
	rs1023586	T/C	4.45E-20	8.99E-09			Familial gestational hyperthyroidism (HTFG) (37)
	rs28414437	A/C	1.05E-19	1.25E-08			Hyperthyroidism, non-autoimmune (HTNA) (38)
	rs11159479	C/T	1.68E-18	2.98E-08			
	rs56389234	G/A	1.68E-18	2.98E-08			
	rs4903962	A/G	6.08E-19	1.63E-08	RPL17P3	RPL17P3 (Ribosomal Protein L17 Pseudogene 3) is a Pseudogene.	Thyroid (39)
	rs2268459	A/G	6.61E-19	4.76E-08			
	rs12323356	A/C	1.05E-18	1.72E-08			
	rs228127	G/A	9.84E-15	3.16E-06	CEP128	CEP128 (Centrosomal Protein 128) is a Protein Coding gene.	Hypothyroidism, Congenital, Non-goitrous, 1 (CHNG1) (40)
	rs7154132	C/T	1.05E-14	6.05E-06			Hyperthyroidism, Non-autoimmune (HTNA) (41)
	rs35176982	(-/AA)	1.40E-14	4.79E-06			
	rs1025253	G/A	1.46E-14	4.25E-06			
	rs8022411	G/T	1.46E-14	4.25E-06			

with a body mass index (BMI) of >28 (609/2556, 23.83%) also increased the risk of stroke, heart disease, diabetes, and hypertension in patients with hyperthyroidism, and the data were statistically significant ($p < 0.05$, Table 5). Moreover, a higher incidence of stroke (4/36, 11.1%) was observed in hyperthyroidism individuals with thyroid storm. Our data yield

TABLE 3 Comorbidity analysis in patients with hyperthyroidism using electronic medical record (EMR) data by gender.

	Patients with hyperthyroidism ($n = 2,767$)		*p*-Value
	Male ($n = 637$)	Female ($n = 2, 130$)	
Thyroid storm	5 (0.8)	31 (1.5)	0.234
Cancer	22 (3.5)	78 (3.7)	0.904
Heart disease	33 (5.2)	58 (2.7)	0.003[a]
Osteoporosis	0 (0)	8 (0.4)	0.211
Infertility	3 (0.5)	26 (1.2)	0.122
Stroke	32 (5.0)	61 (2.9)	0.012[a]
Diabetes	93 (14.6)	168 (7.9)	0.000[a]
Hypertension	49 (7.7)	99 (4.6)	0.005[a]
Hyperlipidemia	10(1.6)	27(1.3)	0.557
gallstone	12(1.9)	26(1.2)	0.242

[a]Significant difference at $p < 0.05$.

TABLE 4 Comorbidity analysis in hyperthyroidism patients with thyroid storm using electronic medical record (EMR) data.

	Hyperthyroidism patient with thyroid storm ($n = 36$)	Hyperthyroidism patient without thyroid storm ($n = 2,731$)	*p*-Value
Cancer	2 (5.6)	98 (3.6)	0.376
Heart disease	1 (2.8)	90 (3.3)	1.000
Osteoporosis	1 (2.8)	7 (0.3)	0.100
Infertility	1 (2.8)	28 (1.0)	0.317
Stroke	4 (11.1)	89 (3.3)	0.031[a]
Diabetes	3 (8.3)	258 (9.4)	1.000
Hypertension	3 (8.3)	145 (5.3)	0.439
Hyperlipidemia	0 (0)	37 (1.4)	1.000
Gallstone	1 (2.8)	37 (1.4)	0.394

[a]Significant difference at $p < 0.05$.

a strong correlation between hyperthyroidism patients with thyroid storm and stroke; the data were statistically significant ($p < 0.05$, Table 4).

Discussion

In the present study, we identified 44 novel variants in 10 loci associated with hyperthyroidism, including *CTLA4*, *HCP5*, *HLA-B*, *POU5F1*, *CCHCR1*, *HLA-DRA*, *HLA-DRB9*, *TSHR*, *RPL17P3*, and *CEP128*. To consider differences in racial backgrounds, and proved that these SNPs is really significant associate with the disease. We further compared the Taiwan Biobank data. Such as Supplementary material 4 (SP4), the SNPs data from Taiwan Biobank indicated that

TABLE 5 Comorbidity analysis in patients with hyperthyroidism using electronic health record (EHR) data of BMI.

	Patient with hyperthyroidism ($n = 2,556$)[b]		*p*-Value
	BMI ≥ 27 ($n = 773$)	BMI < 27 ($n = 1,783$)	
Thyroid storm	11 (1.4)	24 (1.3)	0.8546
Cancer	38 (4.9)	57 (3.2)	0.0402
Heart disease	45 (5.8)	40 (2.2)	0.0000[a]
Osteoporosis	3 (0.4)	5 (0.3)	0.7046
Infertility	9 (1.2)	20 (1.1)	1.0000
Stroke	39 (5.0)	51 (2.9)	0.0072[a]
Diabetes	127 (16.4)	116 (6.5)	0.0000[a]
Hypertension	81 (10.5)	54 (3.0)	0.0000[a]
Hyperlipidemia	14 (1.8)	21 (1.2)	0.2006
Gallstone	19 (2.5)	16 (0.9)	0.0028[a]

[a]With significant differences and $P < 0.05$.
[b]211 Patients without HER data of BMI.

these candidate SNPs in our study were indeed significant difference from those in Taiwan without hyperthyroidism. First, Principal components (PC1-10) were added into GWAS to exclude the effect of racial backgrounds. Second, the subjects enrolled from Taiwan Biobank are Han Chinese in Taiwan and used as general controls. Therefore, the significantly difference of genotype distributions for those selected SNPs between Taiwan Biobank population (as general control) and subjects with hyperthyroidism in our study population could provide evidences that these SNPs are associated with hypertension. To our knowledge, five novel genes that have never before been discussed were found to be associated with hyperthyroidism: *POU5F1*, *CCHCR1*, *HLA-DRB9*, *RPL17P3*, and *CEP128*. Here, we show the biological function and involvement of disease in these genes, which have been discussed previously (Table 2). The related pathways of these candidate genes were then analyzed by DAVID with the recently updated Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/pathway.html). Our data showed that these genes were significantly associated in pathways related to autoimmune thyroid disease (SP5).

Comorbidity analysis in patients with hyperthyroidism and thyroid storm is shown in Table 4. There were no differences in cancer, heart disease, osteoporosis, infertility, diabetes, hypertension, hyperlipidemia, or gallstones. We observed that the percentage of stroke in hyperthyroidism patients with thyroid storm was much higher than that in hyperthyroidism patients without thyroid storm ($p < 0.05$). Briefly, our data indicate that hyperthyroidism patients with thyroid storm may have a higher risk of developing stroke (Table 4).

TABLE 6 Genotypic frequencies of CTLA4 genetic polymorphisms in the patients with hyperthyroidism and controls.

dbSNP ID	Genotype	Control ($N = 32,242$)	hyperthyroidism Patient ($N = 2,767$)	OR (95% CI)	p-Value#	Hyperthyroidism male patient with BMI ≧ 27 ($N = 253$)	OR (95% CI)	p-Value#
rs1427680		($N = 13,357$)	($N = 1,822$)			($N = 165$)		
	GG	6,083 (45.5)	989 (54.3)	1.90 (1.56–2.3)	0.000*	90 (54.6)	5.35 (1.96–14.59)	0.001*
	GA	5,828 (43.6)	709 (38.9)	1.42 (1.16–1.73)		71 (43.0)	4.40 (1.61–12.08)	
	AA	1,446 (10.9)	124 (6.8)	Ref		4 (2.4)	Ref	
rs736611		($N = 13,360$)	($N = 1,824$)			($N = 165$)		
	CC	6,086 (45.6)	989 (54.2)	1.90 (1.56–2.30)	0.000*	90 (54.6)	5.35 (1.96–14.58)	0.001*
	CT	5,828 (43.6)	711 (39.0)	1.42 (1.17–1.74)		71(43.0)	4.40 (1.61–12.08)	
	TT	1,446 (10.8)	124 (6.8)	Ref		4 (2.4)	Ref	
rs11571315		($N = 13,374$)	($N = 1,826$)			($N = 166$)		
	TT	6,092 (45.6)	991 (54.3)	1.89 (1.55–2.29)	0.000*	91 (54.8)	5.42 (1.99–14.77)	0.001*
	CT	5,831 (43.6)	710 (38.9)	1.41 (1.16–1.72)		71 (42.8)	4.42 (1.61–12.11)	
	CC	1,451 (10.8)	125 (6.8)	Ref		4 (2.4)	Ref	
rs231723		($N = 13,373$)	($N = 1,826$)			($N = 166$)		
	AA	1,450 (10.8)	126 (6.9)	Ref	0.000*	4 (2.4)	Ref	0.001*
	AG	5,829 (43.6)	710 (38.9)	1.40 (1.15–1.71)		71 (42.8)	4.42 (1.61–12.11)	
	GG	6,094 (45.6)	990 (54.2)	1.87 (1.54–2.27)		91 (54.8)	5.41 (1.99–14.76)	

CI, confidence interval; OR, odds ratio; # p-value compared with control group; * $p < 0.005$.

Hormones and the cardiovascular system are strongly associated, and disorders of hormonal secretion may lead to increased cardiovascular risk (42, 43). In addition to these well-known effects, there is increasing evidence that hyperthyroidism may accelerate atherosclerosis (44, 45). Endothelial dysfunction, hypercoagulability, and thyroid autoimmunity have been suggested as potential contributors (45–48). Thyroid hormones exert important effects on the cardiovascular system, as demonstrated by the adverse clinical effects that can occur in states of hyperthyroidism and hypothyroidism. Thyroid disorders can impair cardiovascular risk factors, such as those included in the definition of metabolic syndrome (49, 50). Indeed, excess as well as lack of thyroid hormone has been linked to alterations in cardiovascular hemodynamics (51), modifications of heart rhythm (52, 53), and arterial wall structure (54–56). While the effects of thyroid hormone excess on cardiovascular risk factors are clear for some of them, others are still debatable (57, 58). In the United States, stroke is the third leading cause of death, and ∼795,000 people suffer from a new or recurrent stroke annually (59). The prevalence of stroke in Taiwan is reported to be 14.27–19.3 per 1,000 person-years; stroke is the most common cause of complex disability in Taiwan (60, 61). In a Taiwan National Health Insurance Research Database (NHIRD) study, Sheu et al. reported an increased risk of ischemic stroke in young patients with hyperthyroidism (1%) compared with a comparable population without thyroid disease (0.7%) after adjusting for AF (62). In this present study, we defined the significant common genetic risk factors in patients with hyperthyroidism. It might be contributed to the disease early diagnosis with precision medicine.

There was a significant difference in HLA gene loci in our results (Table 2). The human leukocyte antigen (HLA) system, located in the major histocompatibility complex (MHC) on chromosome 6, is highly polymorphic. This region has been shown to be important in human diseases, adverse drug reactions, and organ transplantation. For instance, HLA-B*46:01 is associated with Graves' disease in Taiwan (63). However, the HLA subtype cannot be predicted using a single-nucleotide polymorphism (SNP)-based tagging approach. To understand the relationship between HLA subtypes and diseases, machine learning methods such as HIBAG can be used to better predict HLA subtypes (64). The detailed relationship between these HLA subtypes should be studied in the future.

In order to connect the data between GWAS and PheWAS and further demonstrate the SNPs we identified could predict the risk of comorbidities associated with hyperthyroidism. We analysis the genotypic frequencies of CTLA4 genetic polymorphisms in hyperthyroidism patients and controls (Table 6). Compared with controls, the statistically significant difference was observed in the genotype frequency distribution of CTLA4 rs231779, rs1427680, rs736611, rs11571315, and

rs231723 SNPs. We observed the Odds ratio (OR) were from 1.40 to 1.90. We further included the data of gender and BMI ≧ 27 for analysis, the similar results was observed with statistically significant difference in the genotype frequency distribution of CTLA4 SNPs. However, there was a surprising finding in the section of Odds ratio (OR). Our data showed that the Odds ratio all increased in these five SNPs and the OR value was observed from 4.40 to 5.42 (Table 6).

In conclusion, our findings strongly suggest an association between 44 genetic variants in ten loci and hyperthyroidism susceptibility, and that these genes contribute to the genetic background of hyperthyroidism pathogenesis. Moreover, our data indicate that hyperthyroidism patients with thyroid storm may have a higher risk of developing stroke. We also connected the data between GWAS and PheWAS and demonstrated the SNPs we identified could predict the risk of comorbidities associated with hyperthyroidism. These findings should prompt specific considerations for the diagnosis and treatment of patients with hyperthyroidism, especially in preventing stroke.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found at: https://my.locuszoom.org/gwas/239175/?token=ba633ab327054c3f82c34f7d3cf346d5; https://my.locuszoom.org/gwas/292022/?token=51201ae60be34ae6a8859fe84e908c0c.

Ethics statement

All clinical information was collected from the electronic medical records (EMRs) of CMUH and approved by the respective Ethical Committees of CMUH (CMUH110-REC1-095). The patients/participants provided their written informed consent to participate in this study.

Author contributions

S-YC had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. T-YL, W-LL, T-YW, S-YC, and F-JT: concept and design. C-JC, J-GC, Y-CC, H-FL, H-HY, and S-YC: acquisition, analysis, or interpretation of data. T-YL, W-LL, T-YW, S-YC, and F-JT: drafting of the manuscript. C-JC: statistical analysis. All authors: critical revision of the manuscript for important intellectual content, read, and approved the final manuscript.

Funding

This work was supported by China Medical University Hospital (DMR-108-138 and DMR-109-141).

Acknowledgments

We appreciate the data exploration, statistical analysis, and the support of the iHi Clinical Research Platform from the Big Data Center of CMUH. We thank all colleagues at the Genetic Center, Department of Medical Research, and China Medical University Hospital for their feedback and technical support.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.830621/full#supplementary-material

References

1. Kwon H, Jung JH, Han KD, Park YG, Cho JH, Lee DY, et al. Prevalence and annual incidence of thyroid disease in Korea from 2006 to 2015: a nationwide population-based cohort study. *Endocrinol Metab.* (2018) 33:260–7. doi: 10.3803/EnM.2018.33.2.260

2. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): national health and nutrition examination survey (NHANES III). *J Clin Endocrinol Metab.* (2002) 87:489–99. doi: 10.1210/jcem.87.2.8182

3. Jabbar A, Pingitore A, Pearce SH, Zaman A, Iervasi G, Razvi S. Thyroid hormones and cardiovascular disease. *Nat Rev Cardiol.* (2017) 14:39–55. doi: 10.1038/nrcardio.2016.174

4. Frost L, Vestergaard P, Mosekilde L. Hyperthyroidism and risk of atrial fibrillation or flutter: a population-based study. *Arch Intern Med.* (2004) 164:1675–8. doi: 10.1001/archinte.164.15.1675

5. *The Epidemiology of Hyperthyroidism in Taiwan.* Available online at: http://etds.lib.ncku.edu.tw/etdservice/view_metadata?etdun=U0026029201305445400 (accessed December 23, 2014).

6. Chuang CC, Wang ST, Wang PW Yu ML. Prevalence study of thyroid dysfunction in the elderly of Taiwan. *Gerontology.* (1998) 44:162–7. doi: 10.1159/000022002

7. Lin HD, Lo JG, Ching KN. Etiology of adult goiter in Taiwan–a hospital-based study. *Zhonghua Yi Xue Za Zhi.* (1991) 47:154–60.

8. Osman F, Gammage MD, Franklyn JA. Thyroid disease and its treatment: short-term and long-term cardiovascular consequences. *Curr Opin Pharmacol.* (2001) 1:626–31. doi: 10.1016/S1471-4892(01)00107-2

9. Völzke H, Schwahn C, Wallaschofski H, Dörr M. Review: the association of thyroid dysfunction with all-cause and circulatory mortality: is there a causal relationship? *J Clin Endocrinol Metab.* (2007) 92:2421–9. doi: 10.1210/jc.2007-0179

10. McPherson K, Steel CM, Dixon JM, ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ.* (2000) 321:624–8. doi: 10.1136/bmj.321.7261.624

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

12. Browning BL, Zhou Y, Browning SR, A. One-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* (2018) 103:338–48. doi: 10.1016/j.ajhg.2018.07.015

13. Liu TY, Lin CF, Wu HT, Wu YL, Chen YC, Liao CC, et al. Comparison of multiple imputation algorithms and verification using whole-genome sequencing in the CMUH genetic biobank. *Biomedicine.* (2021) 11:57–65. doi: 10.37796/2211-8039.1302

14. Liu SC, Tsai CH, Wu TY, Tsai CH, Tsai FJ, Chung JG, et al. Soya-cerebroside reduces IL-1β-induced MMP-1 production in chondrocytes and inhibits cartilage degradation: implications for the treatment of osteoarthritis. *Food Agric Immunol.* (2019) 30:620–32. doi: 10.1080/09540105.2019.1611745

15. Linsley PS, Brady W, Urnes M, Grosmaire LS, Damle NK, Ledbetter JA. CTLA-4 is a second receptor for the B cell activation antigen B7. *J Exp Med.* (1991) 174:561–9. doi: 10.1084/jem.174.3.561

16. Teft WA, Kirchhof MG, Madrenas J. A molecular perspective of CTLA-4 function. *Annu Rev Immunol.* (2006) 24:65–97. doi: 10.1146/annurev.immunol.24.021605.090535

17. Lee YH, Harley JB, Nath SK. CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis. *Hum Genet.* (2005) 116:361–7. doi: 10.1007/s00439-004-1244-1

18. Marron MP, Raffel LJ, Garchon HJ, Jacob CO, Serrano-Rios M, Martinez Larrad MT, et al. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet.* (1997) 6:1275–82. doi: 10.1093/hmg/6.8.1275

19. Hunt KA, McGovern DP, Kumar PJ, Ghosh S, Travis SP, Walters JR, et al. A common CTLA4 haplotype associated with coeliac disease. *Eur J Hum Genet.* (2005) 13:440–4. doi: 10.1038/sj.ejhg.5201357

20. Schubert D, Bode C, Kenefeck R, Hou TZ, Wing JB, Kennedy A, et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat Med.* (2014) 20:1410–6. doi: 10.1038/nm.3746

21. Ramírez de Arellano E, Díez-Fuertes F, Aguilar F, de la Torre Tarazona HE, Sánchez-Lara S, Lao Y, et al. Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients. *PLoS ONE.* (2019) 14:e0220459. doi: 10.1371/journal.pone.0220459

22. Liang L, Xu J, Wang M, Xu G, Zhang N, Wang G, et al. Correction: LncRNA HCP5A promotes follicular thyroid carcinoma progression via miRNAs sponge. *Cell Death Dis.* (2019) 10:639. doi: 10.1038/s41419-019-1868-7

23. Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, Ho HC, et al. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature.* (2004) 428:486. doi: 10.1038/428486a

24. Varnavidou-Nicolaidou A, Karpasitou K, Georgiou D, Stylianou G, Kokkofitou A, Michalis C, et al. HLA-B27 in the Greek Cypriot population: distribution of subtypes in patients with ankylosing spondylitis and other HLA-B27-related diseases. The possible protective role of B*2707. *Hum Immunol.* (2004) 65:1451–4. doi: 10.1016/j.humimm.2004.08.177

25. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell.* (2007) 131:861–72. doi: 10.1016/j.cell.2007.11.019

26. El-Bahrawy M. Alpha-fetoprotein-producing non-germ cell tumours of the female genital tract. *Eur J Cancer.* (2010) 46:1317–22. doi: 10.1016/j.ejca.2010.01.028

27. Wu Y, Zhang Y, Mishra A, Tardif SD, Hornsby PJ. Generation of induced pluripotent stem cells from newborn marmoset skin fibroblasts. *Stem Cell Res.* (2010) 4:180–8. doi: 10.1016/j.scr.2010.02.003

28. Du J, Tao J, Xu M, Wang R, Lin L, Huang X, et al. The effects of acupuncture for patients with psoriasis: Study protocol for a randomized controlled trial. *Medicine.* (2021) 100:e26042. doi: 10.1097/MD.0000000000026042

29. Jordan CT, Cao L, Roberson ED, Duan S, Helms CA, Nair RP, et al. Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis. *Am J Hum Genet.* (2012) 90:796–808. doi: 10.1016/j.ajhg.2012.03.013

30. Biondi G, Satta R, Lissia A, Fara AM, Montesu MA. Graham-little-piccardi-lassueur syndrome in siblings. *Int J Dermatol.* (2021) 60:e347–9. doi: 10.1111/ijd.15534

31. Huang CZ, Yang J, Qiao HL, Jia LJ. Polymorphisms and haplotype analysis of IL-4Ralpha Q576R and I75V in patients with penicillin allergy. *Eur J Clin Pharmacol.* (2009) 65:895–902. doi: 10.1007/s00228-009-0659-y

32. Jiang X, Källberg H, Chen Z, Ärlestig L, Rantapää-Dahlqvist S, Davila S, et al. An Immunochip-based interaction study of contrasting interaction effects with smoking in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Rheumatology.* (2016) 55:149–55. doi: 10.1093/rheumatology/kev285

33. Hou S, Du L, Lei B, Pang CP, Zhang M, Zhuang W, et al. Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p312 and 10q213. *Nat Genet.* (2014) 46:1007–11. doi: 10.1038/ng.3061

34. Goris A, Pauwels I, Gustavsen MW, van Son B, Hilven K, Bos SD, et al. Genetic variants are major determinants of CSF antibody levels in multiple sclerosis. *Brain.* (2015) 138:632–43. doi: 10.1093/brain/awu405

35. Nakabayashi K, Matsumi H, Bhalla A, Bae J, Mosselman S, Hsu SY, et al. Thyrostimulin, a heterodimer of two new human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. *J Clin Invest.* (2002) 109:1445–52. doi: 10.1172/JCI0214340

36. Lábadi Á, Grassi ES, Gellén B, Kleinau G, Biebermann H, Ruzsa B, et al. Loss-of-function variants in a hungarian cohort reveal structural insights on TSH receptor maturation and signaling. *J Clin Endocrinol Metab.* (2015) 100:E1039–1045. doi: 10.1210/jc.2014-4511

37. Rodien P, Brémont C, Sanson ML, Parma J, Van Sande J, Costagliola S, et al. Familial gestational hyperthyroidism caused by a mutant thyrotropin receptor hypersensitive to human chorionic gonadotropin. *N Engl J Med.* (1998) 339:1823–6. doi: 10.1056/NEJM199812173392505

38. Vaidya B, Campbell V, Tripp JH, Spyer G, Hattersley AT, Ellard S. Premature birth and low birth weight associated with nonautoimmune hyperthyroidism due to an activating thyrotropin receptor gene mutation. *Clin Endocrinol.* (2004) 60:711–8. doi: 10.1111/j.1365-2265.2004.02040.x

39. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet.* (2013) 45:580–5. doi: 10.1038/ng.2653

40. Russo D, Betterle C, Arturi F, Chiefari E, Girelli ME, Filetti S, et al. novel mutation in the thyrotropin (TSH) receptor gene causing loss of TSH binding but constitutive receptor activation in a family with resistance to TSH. *J Clin Endocrinol Metab.* (2000) 85:4238–42. doi: 10.1210/jc.85.11.4238

41. Biebermann H, Schöneberg T, Hess C, Germak J, Gudermann T, Grüters A. The first activating TSH receptor mutation in transmembrane domain 1 identified in a family with nonautoimmune hyperthyroidism. *J Clin Endocrinol Metab.* (2001) 86:4429–33. doi: 10.1210/jcem.86.9.7888

42. Prejbisz A, Warchol-Celinska E, Lenders JW, Januszewicz A. Cardiovascular risk in primary hyperaldosteronism. *Horm Metab Res.* (2015) 47:973–80. doi: 10.1055/s-0035-1565124

43. Battocchio M, Rebellato A, Grillo A, Dassie F, Maffei P, Bernardi S, et al. Ambulatory arterial stiffness indexes in cushing's syndrome. *Horm Metab Res.* (2017) 49:214–20. doi: 10.1055/s-0043-100385

44. Bano A, Chaker L, Mattace-Raso FUS, van der Lugt A, Ikram MA, Franco OH, et al. Thyroid function and the risk of atherosclerotic cardiovascular morbidity and mortality: the Rotterdam Study. *Circ Res.* (2017) 121:1392–400. doi: 10.1161/CIRCRESAHA.117.311603

45. Bano A, Chaker L, de Maat MPM, Atiq F, Kavousi M, Franco OH, et al. Thyroid function and cardiovascular disease: the mediating role of coagulation factors. *J Clin Endocrinol Metab.* (2019) 104:3203–12. doi: 10.1210/jc.2019-00072

46. Burggraaf J, Lalezari S, Emeis JJ, Vischer UM, de Meyer PH, Pijl H, et al. Endothelial function in patients with hyperthyroidism before and after treatment with propranolol and thiamazol. *Thyroid.* (2001) 11:153–60. doi: 10.1089/105072501300042820

47. Erem C. Coagulation and fibrinolysis in thyroid dysfunction. *Endocrine.* (2009) 36:110–8. doi: 10.1007/s12020-009-9185-z

48. Matsuura E, Atzeni F, Sarzi-Puttini P, Turiel M, Lopez LR, Nurmohamed MT. Is atherosclerosis an autoimmune disease? *BMC Med.* (2014) 12:47. doi: 10.1186/1741-7015-12-47

49. Mehran L, Amouzegar A, Rahimabad PK, Tohidi M, Tahmasebinejad Z, Azizi F. Thyroid function and metabolic syndrome: a population-based thyroid study. *Horm Metab Res.* (2017) 49:192–200. doi: 10.1055/s-0042-117279

50. Delitala AP, Fanciulli G, Pes GM, Maioli M, Delitala G. Thyroid hormones, metabolic syndrome and its components. *Endocr Metab Immune Disord Drug Targets.* (2017) 17:56–62. doi: 10.2174/1871530317666170320105221

51. Danzi S, Klein I. Thyroid hormone and the cardiovascular system. *Med Clin North Am.* (2012) 96:257–68. doi: 10.1016/j.mcna.2012.01.006

52. Osman F, Franklyn JA, Holder RL, Sheppard MC, Gammage MD. Cardiovascular manifestations of hyperthyroidism before and after antithyroid therapy: a matched case-control study. *J Am Coll Cardiol.* (2007) 49:71–81. doi: 10.1016/j.jacc.2006.08.042

53. Kannan L, Kotus-Bart J, Amanullah A. Prevalence of cardiac arrhythmias in hypothyroid and euthyroid patients. *Horm Metab Res.* (2017) 49:430–3. doi: 10.1055/s-0043-105275

54. Delitala AP, Orrù M, Filigheddu F, Pilia MG, Delitala G, Ganau A, et al. Serum free thyroxine levels are positively associated with arterial stiffness in the SardiNIA study. *Clin Endocrinol.* (2015) 82:592–7. doi: 10.1111/cen.12532

55. Völzke H, Robinson DM, Schminke U, Lüdemann J, Rettig R, Felix SB, et al. Thyroid function and carotid wall thickness. *J Clin Endocrinol Metab.* (2004) 89:2145–9. doi: 10.1210/jc.2003-031028

56. Wang J, Zheng X, Sun M, Wang Z, Fu Q, Shi Y, et al. Low serum free thyroxine concentrations associate with increased arterial stiffness in euthyroid subjects: a population-based cross-sectional study. *Endocrine.* (2015) 50:465–73. doi: 10.1007/s12020-015-0602-1

57. Delitala AP, Filigheddu F, Orrù M, AlGhatrif M, Steri M, Pilia MG, et al. No evidence of association between subclinical thyroid disorders and common carotid intima medial thickness or atherosclerotic plaque. *Nutr Metab Cardiovasc Dis.* (2015) 25:1104–10. doi: 10.1016/j.numecd.2015.09.001

58. Peixoto de Miranda ÉJ, Bittencourt MS, Goulart AC, Santos IS, Mill JG, Schmidt MI, et al. Lack of association between subclinical hypothyroidism and carotid-femoral pulse wave velocity in a cross-sectional analysis of the ELSA-Brasil. *Am J Hypertens.* (2017) 30:81–7. doi: 10.1093/ajh/hpw117

59. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation.* (2010) 121:586–613. doi: 10.1161/CIRCULATIONAHA.109.192703

60. Huang ZS, Chiang TL, Lee TK. Stroke prevalence in Taiwan. Findings from the 1994 national health interview survey. *Stroke.* (1997) 28:1579–84. doi: 10.1161/01.STR.28.8.1579

61. Lin HC, Lin YJ, Liu TC, Chen CS, Chiu WT. Urbanization and stroke prevalence in Taiwan: analysis of a nationwide survey. *J Urban Health.* (2007) 84:604–14. doi: 10.1007/s11524-007-9195-1

62. Sheu JJ, Kang JH, Lin HC, Lin HC. Hyperthyroidism and risk of ischemic stroke in young adults: a 5-year follow-up study. *Stroke.* (2010) 41:961–6. doi: 10.1161/STROKEAHA.109.577742

63. Chen PL, Fann CS, Chu CC, Chang CC, Chang SW, Hsieh HY, et al. Comprehensive genotyping in two homogeneous Graves' disease samples reveals major and novel HLA association alleles. *PLoS ONE.* (2011) 6:e16635. doi: 10.1371/journal.pone.0016635

64. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG–HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* (2014) 14:192–200. doi: 10.1038/tpj.2013.18

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact