



Published in final edited form as:

Gastroenterology. 2021 June ; 160(7): 2572–2584. doi:10.1053/j.gastro.2021.01.233.

International Liver Cancer Association (ILCA) White Paper on Biomarker Development for Hepatocellular Carcinoma

Amit G. Singal¹, Yujin Hoshida¹, David J. Pinato², Jorge Marrero¹, Jean-Charles Nault^{3,4,5}, Valerie Paradis⁶, Nabihah Tayob⁷, Morris Sherman⁸, Young Suk Lim⁹, Ziding Feng¹⁰, Anna S. Lok¹¹, Jo Ann Rinaudo¹², Sudhir Srivastava¹², Josep Llovet^{13,14,15}, Augusto Villanueva¹³

¹Division of Digestive and Liver Diseases, Department of Internal Medicine, UT Southwestern Medical Center, Dallas, Texas, Texas ²Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital, London, UK ³Service d'hépatologie, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, Bondy, France ⁴Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris 13, Paris, France ⁵Centre de Recherche des Cordeliers, Inserm, Sorbonne Université, Université Paris, INSERM UMR 118 Functional Genomics of Solid Tumors Laboratory, F-75006, Paris, France ⁶Pathology Department, Beaujon hospital, Clichy, University Paris, France ⁷Department of Data Science, Dana Farber Cancer Institute; Department of Medicine, Harvard Medical School, Boston, Massachusetts ⁸University Health Network, Toronto, Ontario, Canada ⁹Department of Gastroenterology, Liver Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea ¹⁰Biostatistics Program, Public Health Science Division, Fred Hutchinson Cancer Research Center, Seattle, Washington ¹¹Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor, Michigan ¹²Cancer Biomarker Research Group, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Rockville, Maryland ¹³Division of Liver Diseases and Hematology/Medical Oncology, Liver Cancer Program, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York ¹⁴Translational Research in Hepatic Oncology, Liver Unit, IDIBAPS, Hospital Clinic, University of Barcelona, Catalonia, Spain ¹⁵Institució Catalana d'Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

Keywords

Liver Cancer; Biomarkers; Cirrhosis

Biomarkers are key components of the clinical management of patients with cancer, as they have contributed to major survival improvements in these patients.¹ According to the National Cancer Institute, a biomarker is a biological molecule found in blood, other body fluids (eg, urine), or tissues that is a sign of a normal or abnormal process, or of a condition

Address correspondence to: Amit G. Singal, MD, MS, Division of Digestive and Liver Diseases, UT Southwestern Medical Center, 5959 Harry Hines Boulevard, POB 1, Suite 420, Dallas, Texas 75390-8887. amit.singal@utsouthwestern.edu; fax: 214-645-6114; Augusto Villanueva, MD, PhD, 1425 Madison Avenue, Box 1123, Room 11-70E, New York, New York 10029. augusto.villanueva@mssm.edu.

or disease. They allow classification of patients based on common features and facilitate risk stratification, early detection, diagnosis, and prediction of prognosis or treatment response. In hepatocellular carcinoma (HCC), there are few biomarkers incorporated in clinical practice despite a need to better stratify patients at different steps of clinical management. However, this has been an extensive area of research in recent years, with increasing efforts to identify biomarkers across the cancer care continuum from risk stratification to early detection to prognostication and treatment response (Table 1, Figure 1).

One of the first systematic sets of recommendations dealing with biomarkers in cancer was introduced in 1996 by Hayes et al,² known as the Tumor Marker Utility Grading System (TMGUS). These recommendations covered not only technical aspects of assay development, but also issues related to clinical utility and levels of evidence. TMGUS was later expanded by the Reporting recommendations for tumor marker prognostic studies (REMARK) guidelines,³ a more focused approach on recommendations for reporting prognostic biomarkers in oncology. There have also been specific initiatives to describe study design thoroughly for cancer biomarkers in specific clinical scenarios, such as the Early Detection Research Network (EDRN) publishing a framework for 5 phases of biomarker development and validation for cancer screening.⁴ In brief, these phases extend from biomarker discovery (phase I) to evaluation of biomarker performance (phases II-III) and clinical benefits and harms (phase IV-V) (Table 2).

Although these guidelines provide a useful general framework of data elements required at each step, deviations from this framework may be possible or necessary in specific circumstances. Further, modifications may be required when applied to other clinical scenarios, including risk stratification and treatment response assessment. Finally, singularities unique to HCC, particularly the coexistence of chronic liver disease in most patients, lead to necessary considerations when designing biomarker studies. To address these issues, the International Liver Cancer Association has assembled a group of experts on biomarker development to provide a framework on best practices to design, execute, and interpret biomarker studies for risk stratification, early detection, diagnosis, prognostication, and treatment response assessment in HCC.

HCC Risk Stratification

Rationale

Risk of HCC is elevated in patients with chronic liver disease, particularly those with cirrhosis from any etiology, and HCC is one of the leading causes of death in these patients.⁵ Contemporary cohorts, which have higher numbers of patients with hepatitis C post-sustained viral response and those with nonalcoholic steatohepatitis (NASH), have demonstrated an annual HCC risk of 1% to 3%, substantially lower than the traditional annual HCC risk of 3% to 5% from older cohorts.^{6,7} However, patient characteristics, such as age, sex, race/ethnicity, and degree of fibrosis, introduce heterogeneity in HCC risk between patients.⁸ Subgroups of patients with chronic hepatitis B virus (HBV) infection without cirrhosis, such as Asian men older than 40 years and Asian women older than 50 years, have an annual HCC incidence of 0.4% to 0.6%, whereas younger individuals have a lower HCC incidence of 0.2%.

Refined risk stratification could have several implications for clinical practice, such as tailoring of HCC surveillance intensity in the future and targeting chemoprevention efforts (eg, coffee, lipophilic statins, aspirin⁹) to high-risk persons. For example, the “one-size-fits-all” surveillance strategy of semi-annual ultrasound with or without alpha-fetoprotein (AFP) could be tailored based on the individual HCC risk in each patient.¹⁰ Further, risk stratification biomarkers could help select patients for chemoprevention clinical trials, thereby reducing necessary sample sizes and making these trials more feasible.

Target Population

In addition to patients with cirrhosis, risk stratification is also needed for patients with chronic HBV infection or with advanced fibrosis, particularly in the setting of NASH or post-sustained viral response. Although HCC can occur in the absence of cirrhosis,¹¹ patients without cirrhosis have a low annual HCC rate and surveillance is not cost-effective.^{12,13} Effective risk stratification biomarkers may identify subgroups of those with advanced fibrosis but not cirrhosis who would benefit from surveillance.

Study Design and Outcomes

The EDRN framework for phases of biomarker evaluation for early detection do not directly apply to risk stratification, although several of the concepts are similar. Derivation of risk stratification biomarkers can be done in a retrospective cohort and/or nested case-control study, in which the biomarker assay would be assessed at baseline, cases develop HCC during follow-up, and controls remain HCC-free over equivalent or longer follow-up. The time frame between biomarker assessment and HCC diagnosis should be long enough (eg, > 2 years) to minimize the likelihood of detecting preexisting undetected tumoral disease. A 2-year time frame is recommended given the low sensitivity of imaging to detect very early-stage HCC and some patients with HCC exhibiting indolent growth patterns, thereby potentially remaining subclinical for many months.^{14,15} Therefore, case-control studies with shorter periods between biomarker assessment and HCC diagnosis may confound a biomarker’s performance for risk stratification versus early detection (Figure 2).

When selecting patients for a nested case-control design, cases and controls should be matched for known clinical risk factors, such as age, gender, and etiology and severity of liver disease. When available, archived biospecimens from a prospective cohort study (ie, prospective-retrospective design) would allow more reliable estimation of biomarker performance based on time-to-event data analysis.¹⁶ As needed, early-phase validation can be conducted in independent nested case-control studies or cohorts, where model parameters and/or thresholds to define risk groups could be further optimized. Late-phase validation is then conducted in an adequately powered independent prospective-retrospective or prospective cohort to compare the HCC incidence rate between high-risk and low-risk groups or assess association of the biomarker and HCC incidence.

Biomarker performance should be evaluated using a combination of several metrics. Overall model performance (degree of variation explained by the biomarker panel) is typically assessed by R^2 or Brier scores. Discrimination (ability to distinguish between patients who develop versus do not develop HCC) is assessed by the magnitude of risk separation

between high-risk and low-risk groups measured by fitness of the model determined by time-dependent area under the receiver operator characteristic curve, concordance index (c-index), and/or Akaike's Information Criterion. Calibration (difference between observed and predicted event rates) is often assessed by the Hosmer-Lemeshow statistic and improvement of prediction (or reclassification) is assessed via net reclassification index misclassification tables or standardized net benefit.

Following validation of risk-predictive capability, the implementation phase would evaluate the benefit of incorporating risk stratification biomarkers in an experimental system, for example, individual-risk-stratified HCC screening, for outcomes including early detection, overall survival (OS), and cost-effectiveness. This is a complex problem, for which simulation analyses such as cost-effectiveness models can be helpful.⁷ Relevant variables to consider include performance of the biomarker to stratify HCC risk groups, HCC incidence in each risk group, costs for medical care including the biomarker, and benefits and harms of HCC screening in each risk group. These models can also provide information about desired biomarker performance and costs to meet the cost-effectiveness threshold.

There are several analytical issues that should be considered including the following: (1) use of a continuous risk measure versus assigning risk categories, with the former providing more granular information but the latter being easier to interpret by providers; (2) evaluating a biomarker in isolation vs combining with clinical variables as an integrative risk score, with the latter likely being needed in light of the heterogeneity of cancer pathogenesis; (3) accounting for changes in risk over time due to natural disease progression or clinical interventions (eg, antiviral therapy); and (4) optimizing risk stratification models in specific clinical contexts defined by clinical characteristics such as liver disease etiology. Decisions regarding these model optimization issues must balance clinically meaningful benefit vs practical feasibility.

Early Detection

Rationale

HCC fulfills all of the World Health Organization criteria for a cancer screening program, including high morbidity and mortality, an identifiable target population, a recognizable preclinical stage, accepted recall procedures, and efficacious treatment. Therefore, professional society guidelines recommend semi-annual surveillance using abdominal ultrasound with or without AFP among at-risk patients, including subgroups of patients with chronic HBV infection and those with cirrhosis from any etiology.^{17,18}

Ultrasound and AFP are the only 2 surveillance tests recommended in guidelines and have been the longstanding cornerstone of HCC surveillance. However, ultrasound can have highly variable performance given its operator-dependent nature and the sensitivity of ultrasound with AFP, at a cutoff of 20 ng/mL, for early HCC detection is suboptimal, at only 63%.¹⁹ Evolving data also highlight the potential for poor ultrasound visualization, particularly among obese patients and those with NASH, as well as false positive or indeterminate results causing screening-related harms.^{20,21} Finally, ultrasound-based surveillance programs often require a separate appointment, creating potential barriers to

adherence, contributing to underuse of HCC surveillance in clinical practice, occurring in less than 50% across geographic regions.^{22,23} Overall, the limitations of our current strategy highlight a strong need for alternative surveillance tests, particularly highly accurate blood-based biomarkers.

Target Population

Patients with cirrhosis from any etiology comprise the group with the highest risk for developing HCC and account for >90% of all cases in the United States and Europe, whereas chronic HBV infection remains the most common target population globally. Although contemporary cohorts suggest a lower annual HCC incidence of ~2%,²⁴⁻²⁶ HCC incidence among patients with cirrhosis still exceeds the cost-effectiveness threshold of 1.5% per year.^{17,18,27} Similarly, annual HCC incidence in subgroups with chronic HBV infection exceeds the cost-effectiveness threshold of 0.2%. Therefore, patients with cirrhosis or chronic HBV infection should be the target population for studies examining HCC surveillance biomarkers.¹⁷ Further, surveillance studies should be restricted to those who would potentially benefit from a therapeutic intervention. For example, surveillance is only of benefit in patients with preserved liver function, as those with significant hepatic decompensation (eg, Child Pugh C cirrhosis if not eligible for liver transplantation) or comorbidity have higher competing risk of non-HCC mortality.

Patients with advanced fibrosis but not cirrhosis are known to develop HCC, but with incidence rates well below the threshold recommended for surveillance. Biomarker studies in HCC should not combine patients with cirrhosis and those with advanced fibrosis, as this will lead to underestimating sample size calculations, longer accrual to identify a sufficient number of HCC cases, and ultimately lead to biased conclusions and uncertainty if there is a benefit of the new biomarker. If an accurate risk stratification biomarker can identify a subset of patients with advanced fibrosis and similar HCC risk as those with cirrhosis, these patients may be included in early detection trials in the future.

Study Design and Outcomes

A frequent misconception in the field is the confusion between early detection and diagnostic biomarkers. As tools for cancer surveillance, early detection biomarkers will trigger a confirmatory diagnostic procedure, but per se, they are not sufficient to assign an HCC diagnosis. The appropriate study design and outcomes will be dependent on the phase of biomarker development (Table 2).⁴

Phase 1 and 2 biomarker studies.—Phase 1 studies aim to identify biomarkers and determine how well they distinguish HCC and non-HCC controls, that is, the true positive rate (TPR) and false positive rate (FPR). These studies can include genes (single or array), proteins, and radiologic tests and may start with measurement of the biomarker at the tissue level, with or without correlation with serum or plasma. Phase 2 is the clinical assay development based on a specimen that can be obtained non-invasively. Outcomes at this phase are estimation of TPR and FPR or the receiver operating characteristics curve for the biomarker to distinguish subjects with HCC from those with cirrhosis but without HCC.

The same analytic principles comparing cases and controls apply to phase 1 and phase 2 studies. However, phase 2 studies should be appropriately powered to not only estimate TPR, FPR, and area under the receiver operator characteristic curve but also determine the impact of covariates such as age, sex, etiology of liver disease, and degree of liver dysfunction on biomarker performance. These covariates of interest cannot be used as matching variables, as doing so would render their effects toward null. In a phase II study, HCC cases should be ideally restricted to those at an early stage, either defined by Barcelona Clinic Liver Cancer (BCLC) staging system or Milan Criteria, because the goal of HCC is detection of early-stage disease and biomarker performance would otherwise be over-estimated.^{17,18} It is at this stage that comparison with the current standard is performed and the study should be powered to compare the new biomarker to ultrasound with or without AFP, although it should be noted this comparison may have bias given ultrasound and AFP were likely to trigger diagnostic testing in a subset of the cohort. Depending on the biomarker, one may consider subgroup analyses of particular interest, such as among those with NASH cirrhosis. Similarly, one could consider subgroup analyses by baseline HCC risk if accurate risk stratification biomarkers are available in the future.

Phase 3 biomarker studies.—Phase 3 studies leverage prospective cohort studies, in which samples are collected at regular semi-annual intervals, with the main outcome to evaluate the ability of the biomarker to detect preclinical HCC. The samples are stored initially and then analyzed using the PRoBE design (prospective specimen collection, retrospective blinded evaluation), allowing a nested case-control analysis.²⁸ Adequate sample size to facilitate a sufficient number of incident HCC and provide strong conclusions is critical, including facilitating subgroup analyses in subpopulations of interest to help determine the impact of covariates on the biomarker's accuracy. Protocols should detail specimen handling (including collection, processing, storage, and retrieval), and there should be strict definitions for incident HCC, per guidelines,^{17,18} including the use of a multidisciplinary tumor board or adjudication committee. Phase 3 studies should also incorporate end-of-study imaging or a follow-up period among non-HCC patients to minimize risk of ascertainment bias. It is important to identify if the biomarker will be used alone or combined with other markers or demographic information (such as age and sex, as has been done with GALAD).²⁹ Based on ultrasound and AFP performance, minimally acceptable TPR and FPR rates for new biomarkers for early HCC detection are approximately 65% and 10%, respectively, and these should be measured at preclinical lag times of interest (eg, at diagnosis, or 6 to 12 months before HCC diagnosis). Thresholds for TPR and FPR vary by population-level HCC risk in the local area, so strategies with higher TPR may be desired in areas with higher HCC risk populations.

Phase 4 biomarker studies.—Phase 4 studies are prospective cohort studies in which the biomarker of interest is applied to individuals in real-time and diagnostic procedures are performed for those with a positive test. For such studies, the assay must be reliable and reproducible, and readily available to clinicians to make decisions for diagnosis and treatment. There are 4 potential outcomes at each surveillance interval: (1) the biomarker is positive and HCC is confirmed (true positive), (2) the biomarker is positive and HCC is not confirmed (false positive), (3) the biomarker is negative and HCC is discovered (false

negative), and (4) the biomarker is negative and HCC is absent (true negative). Without adequate follow-up to confirm false negatives or true negatives, positive predictive value can be calculated after workup for test positives but sensitivity or specificity could not be calculated. Therefore, measures such as an additional follow-up period of 6 to 12 months with ultrasound-based surveillance, or diagnostic computed tomography (CT) or magnetic resonance imaging (MRI) in patients with negative surveillance tests, are needed to exclude HCC and minimize risk of ascertainment bias. Outcomes of interest from a phase IV biomarker study include the detection rate, that is, the proportion of cirrhotic subjects who tested positive and have HCC, and false-referral rate, that is, the proportion who have a positive surveillance test but do not have HCC on diagnostic imaging. Tumor characteristics including stage and any features of tumor biology should be collected to inform power calculations for a subsequent phase 5 study.

An issue with phase 4 studies is that they can be costly and time-consuming, thereby delaying availability of new biomarkers for early detection of HCC. A well-performed phase 3 study may obviate the need for a phase 4 study if the following are present: (1) the biomarker is readily available and reproducible, (2) phase III study accounted for potential ascertainment bias in controls without HCC, (3) phase III study performed longitudinal evaluation of the biomarker in non-HCC patients to characterize FPR over time, (4) phase III study assessed biomarker performance at early detectable time points among HCC patients to characterize TPR, and (5) phase III study provided an estimate for mortality reduction to inform sample size calculations for a phase 5 study. However, this approach assumes that a positive biomarker result would have prompted a diagnostic evaluation and the CT or MRI would have detected HCC if present. Therefore, proceeding directly from a phase 3 to phase 5 study should be performed with some caution, particularly given the expense associated with phase 5 studies. Alternatively, an adaptive design could be considered, in which a randomized phase 4 study transitions to a phase 5 study if significant differences in tumor stage are observed at a prespecified interim analysis. If no differences are seen at the time of this interim analysis, the study would be terminated given low likelihood of detecting a difference in HCC-related mortality with continued follow-up.

Phase 5 biomarker studies.—Finally, a phase 5 study would address the question of whether surveillance using the new biomarker reduces HCC-related mortality. Even when HCC is detected at early stages, surveillance may not reduce HCC-related mortality due to poor compliance and utilization, lack of access to curative treatments, and overdiagnosis. The design should include central randomization to the new biomarker vs standard of care (ie, ultrasound with or without AFP). Designing a randomized controlled study of surveillance versus no surveillance among patients with cirrhosis is not well accepted by providers or patients.³⁰ The sample should be robust to allow evaluating the primary endpoint of HCC-related mortality among all patients. Concerns about statistical power and required sample size make all-cause mortality reduction a difficult primary outcome. Secondary outcomes of interest include cost, acceptance, extent of overdiagnosis and overtreatment, and utilization. At this time, the ability to further stratify HCC risk among cirrhosis to identify those at highest risk is not available; however, one can consider using features such as male sex and family history of HCC to enrich the study cohort. The analysis

would involve using survival analysis methods for censored data to compare study arms. One can consider computer modeling methods as a preliminary step for assessing the necessity for these randomized studies by evaluating cost-effectiveness and quality of life of new surveillance strategies. Performing a phase 5 study without robust phase 3 and/or phase 4 data is not recommended given the high cost and low likelihood of success.

Diagnosis

Rationale

As opposed to biomarkers for early HCC detection, diagnostic biomarkers are applied in patients with a suspected nodule to rule-in/out HCC and treat the patient accordingly. The diagnosis of HCC in cirrhosis has improved considerably, thanks to the wide implementation of noninvasive imaging criteria based on contrast-enhanced CT and MRI.^{17,18} In essence, patients with cirrhosis and nodules larger than 1 cm with a specific appearance on dynamic imaging can be confidently diagnosed with HCC, with a Liver Reporting and Data System 5 lesion having >95% positive predictive value for presence of HCC.³¹ These criteria have reduced the need to perform invasive tissue biopsies for diagnostic purposes. However, there are still clinical niches where diagnostic biomarkers can help improve the clinical management of patients with HCC, including those with indeterminate nodules in the setting of cirrhosis as well as suspicious nodules in patients in whom radiologic diagnosis is not possible, such as those without cirrhosis or chronic hepatitis B infection.

Target Population

Three target populations would significantly benefit from new diagnostic biomarkers: First, patients with cirrhosis with indeterminate liver nodules without the hallmark radiological features on imaging (eg, LR-3 or LR-4) have HCC risk that varies between 30% and 75% and currently requires either a biopsy or enhanced follow-up^{17,18}; second, hepatic nodules detected in patients without cirrhosis cannot be diagnosed radiographically and a confirmatory tissue biopsy is required; last, patients with cirrhosis and small liver nodules (<2 cm) in whom the histological diagnosis of HCC can be challenging, particularly differentiating well-differentiated HCC from high-grade dysplastic nodules or intrahepatic cholangiocarcinoma.³² Differentiating HCC from intrahepatic cholangiocarcinoma is increasingly important given expanded treatment options for the latter, including biomarker-selected molecular targeted therapies.

Study Design and Outcomes

Despite most HCC diagnostic biomarker studies using retrospective “convenience” samples,³³ better study designs should be considered, including cross-sectional, case-control, and prospective cohorts, particularly for newly identified hepatic nodules in cirrhosis not conclusive for HCC on imaging. It is crucial that the gold standard for diagnosis is clearly specified in the study design of a novel diagnostic biomarker. Primary outcomes should be sensitivity and specificity of the candidate biomarker to diagnose HCC. For most studies testing novel diagnostic biomarkers, the gold standard should be either histology or noninvasive criteria using dynamic imaging. Although patients with typical HCC on imaging can be included to better assess sensitivity and specificity of the new diagnostic biomarker,

histology should be the gold standard for biomarkers tested in patients with atypical imaging features or in the absence of cirrhosis. Notably, histology or imaging are not sufficient to exclude the diagnosis of HCC if the tumor biopsy is negative for malignancy or if the imaging is inconclusive given both having imperfect negative predictive value. If the nodule is stable in size at imaging during follow-up, there is no clear-cut length of follow-up to exclude HCC. We recommend the use of a 2-year cutoff to exclude the diagnosis of HCC, as up to one-third of HCC may have a tumor volume doubling time of more than 1 year.^{14,15}

In the case of biomarkers to discriminate between well-differentiated HCC and high-grade dysplastic nodules, histological analysis may require surgical specimens, as needle biopsies are sometimes suboptimal to address this diagnostic challenge. In these cases, independent review by at least 2 expert liver pathologists may be required to reach a consensus diagnosis. Glypican 3, glutamine synthase, and HSP70 immunostaining also can be used as an aid in the evaluation of these difficult-to-diagnose lesions.^{34,35}

Besides conventional demographic data (eg, sex and age), additional covariates should be included in the description of the study population, such as the degree of liver dysfunction and inflammation (aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, bilirubin, albumin, and international normalized ratio), portal hypertension (platelet count, elastography), and the etiology of the underlying liver disease. Studies that evaluate diagnostic biomarkers should clearly describe when and how the analyte (eg, blood or tissue) was collected, as this is critical to understand which clinical situation best suits the new biomarker.

Prognosis

Rationale

Prognosis prediction is a crucial component in the clinical management of HCC. It helps classify patients based on survival and informs tumor staging systems. Prognostic factors are also fundamental tools to stratify patients in clinical trials and allow fair comparisons between the treatment arms. As most patients with HCC have concomitant liver disease, their prognosis not only depends on tumor features, but also on the severity of the underlying liver disease. Thus, HCC prognosis can be described in terms of 2 sets of variables: those derived from the tumor (eg, microvascular invasion, satellite nodules, AFP levels) and those from the nontumor liver (eg, presence of cirrhosis, portal hypertension, and severity of liver failure).³⁶ Some of these features (eg, microvascular invasion, tumor satellites) can be assessed only in surgical specimens. Currently, only serum AFP levels have enough high-level supporting evidence to be used as a biomarker in clinical practice to predict prognosis of patients with HCC.³⁷ Higher AFP levels are associated with poorer prognosis in patients at different stages of the disease. For example, an AFP threshold of 400 ng/mL has been used to stratify patients in randomized clinical trials, and a threshold of 1000 ng/mL has been used to select patients unsuitable for liver transplantation in some countries, although AFP's prognostic value is also observed when interpreted in a continuous manner.^{38,39} However, the prognostic value of serum AFP is limited, as only a small proportion of patients have increased AFP levels (10% of patients at early stages have AFP >400 ng/dL scaling up to 40% in advanced stages⁴⁰). There are several prognostic gene

signatures associated with tumor recurrence or survival, using both information from the tumor and the nontumor adjacent liver.⁴¹⁻⁴⁴ Although some have been independently validated,^{41,43,45} few have been externally evaluated in the setting of prospective studies.⁴⁶ Thus, no prognostic gene signature is applied in clinical practice and development of new prognostic biomarkers is a clinical need.

Target Population

It is critical to have a homogeneous HCC patient population when developing novel prognostic biomarkers. Therefore, studies exploring these biomarkers should follow the BCLC staging system⁴⁷ or similarly validated staging system to select patients within a single stage and they should receive the same treatment. This will avoid mixing different treatment modalities and confounding the prognostic performance of the biomarker. A valid alternative would be to have enough patients in each treatment group to adequately power subgroup analysis for biomarker validation.

Study Design and Outcomes

Study design and data analysis should be in consonance with the REMARK recommendations for reporting prognostic biomarkers in oncology.^{3,48} The main steps in the development of novel prognostic biomarkers start with the identification of the biomarker using a training and validation set. The proposed new biomarkers should retain independent prognostic value compared with known clinical and pathological features. The performance of the biomarker should be validated in an external cohort, ideally from an independent team. It is critical that the study design fits the intended clinical use of the biomarker. There are several unique features related to HCC and the underlying liver disease that should be considered during study design and analysis. The best way to control for features that are known to predict outcomes is to adjust for them in the models, and the new biomarker should predict outcomes independently of these factors. Variables to consider include characteristics of the tumor (size, number, degree of cell differentiation, microvascular invasion, satellite nodules, presence of metastasis and macrovascular tumor invasion, and serum AFP), the underlying liver disease (etiology, hepatitis C virus eradication, HBV therapy, alcohol use cessation) and degree of fibrosis, severity of liver disease and portal hypertension (platelet count, bilirubin, international normalized ratio). Ideally, the methods used to assess these variables should be prespecified in the protocol. It is crucial that biomarker studies collect the analyte (eg, blood or tissue) at the time and clinical setting when the biomarker is intended to be used in patient decisionmaking. Any modifications should be described in detail so that the prognostic performance is not biased because of protocol deviations.

The primary outcomes of new prognostic biomarkers vary according to the tumor stage for which the biomarker is designed. For instance, the primary outcome for patients at early stage (BCLC 0/A) should be recurrence-free survival (RFS) and OS, whereas time to recurrence could be a secondary outcome. In those patients at intermediate stage (BCLC B) or advanced stage (BCLC C), the primary outcomes should be OS and progression-free survival (PFS).⁴⁹ The assessment of some outcomes, especially time to recurrence and PFS, is often biased in retrospective studies because they are based on heterogeneous clinical

practice, specifically in the frequency and interval when imaging studies are conducted. A prespecified imaging protocol (typically performed every 6–8 weeks) and a plan for calling a tumor recurrence are key in prospective cohort studies evaluating novel prognostic biomarkers.

Treatment Response Prediction

Rationale

The development of predictive biomarkers in HCC has been traditionally restricted by the inherent complexity of its pathophysiology⁵⁰ and a limited therapeutic landscape. With 7 drugs showing survival benefits in phase III clinical trials (eg, atezolizumab plus bevacizumab, sorafenib, lenvatinib, regorafenib, cabozantinib, ramucirumab), predictive correlates of response and survival advantage to systemic therapy hold the potential to improve the quality of hypothesis testing in clinical trials, optimize drug development by reducing attrition and guide clinical decision-making in routine practice. Similarly, an expanding landscape of locoregional therapy options now include transarterial radioembolization and stereotactic body radiation therapy in addition to transarterial chemoembolization. Recognizing the growing need for molecularly based prediction of clinical outcomes, this section highlights the challenges of predictive biomarker qualification and validation and provides a series of consensus recommendations to accelerate the clinical translation of predictive biomarkers in HCC across molecularly targeted agents and immune-based therapies. The lack of need for histological confirmation for an HCC diagnosis has limited the use of diagnostic biopsies for biomarker development.⁵¹ Although noninvasive biomarkers are expanding in HCC,⁵² evidence of low complication rates from tumor biopsies in clinical trials⁵³ has promoted acknowledgment of their importance in trial design and clinical practice guidelines.^{17,18} Upfront molecular stratification has been attempted in clinical trials, but with limited success in most cases (eg, tivantinib,⁵³ refametinib⁵⁴). Retrospective analyses of samples collected in the context of clinical trials did not result in robust biomarkers of response to either sorafenib³⁷ or regorafenib.⁵⁵ Currently, the only validated treatment biomarker in HCC is high AFP level (>400 ng/mL), as it predicts survival benefit to ramucirumab after a biomarker enriched phase III trial.⁵⁶

Target Population

Similar to any other clinical scenarios, studies on treatment response predictive biomarkers need to be evaluated in a homogeneous patient population, and this consensus endorses the use of the BCLC staging system to select patients within a single tumor stage and homogeneous treatment type. For predictive biomarkers, the intervention needs to be well defined and described in the study protocol.

Study Design and Outcomes

One of the common errors when defining a predictive biomarker is to mistakenly assign treatment response properties to a prognostic biomarker. This is generally the case when the biomarker is shown to correlate with a specific outcome in treated patients. However, it is key to show the performance of the biomarker in untreated patients and calculate the significance of interaction between the biomarker and treatment (interaction *P* value).

Irrespective of the source of human biological material considered, the roadmap for predictive biomarker development includes verification of analytical validity (capacity to reproducibly detect the desired biologic trait), determination of clinical validity (adequate linkage with desired clinical outcome), and clinical utility (capacity of a biomarker testing strategy to improve patient outcomes compared with a biomarker-unselected strategy).⁵⁷

When designing studies to evaluate predictive biomarkers, it is crucial to have a solid methodological framework that acknowledges the prognostic heterogeneity across BCLC stage and controls for key confounders (ie, known prognostic factors across stage), including liver functional reserve, AFP levels, macrovascular invasion, and extrahepatic spread. Predictive biomarker development in HCC has largely followed the adoption of targeted clinical trial design, where biomarker-positivity is a prerequisite for exposure to experimental therapies (eg, AFP and ramucirumab^{56,58}). Other biomarker-driven strategies are preferred when confidence on the stratifying potential of the biomarker of choice is lower, as summarized in Figure 4.⁵⁹ The biomarker-stratified design, for instance, allows randomization to treatment stratified by marker status, therefore enabling evaluation of the marker/treatment interaction across the experimental and control arms. The interaction *P* value is typically a secondary objective given limited statistical power for this outcome. In the biomarker-based strategy design, randomization happens before biomarker testing, allowing a comparison between a biomarker-selected population with the experimental therapy versus a biomarker-unselected control population treated with standard of care therapy. This design is optimal to determine the clinical utility of the biomarker as a selection tool for a specific therapy. Differences in costs and sample size affect the feasibility of such approaches, with key determinants of success being prevalence of the biomarker and predicted magnitude of treatment effects in the experimental and control arms. To overcome these limitations, umbrella studies allow parallel enrollment of patients of predefined molecular subtypes, each arm matched with the most appropriate molecular therapy.⁶⁰ However, patient randomization in umbrella trials has a significant impact on feasibility, and this approach has yet to be tested in HCC. Although exceptions are emerging in HCC (eg, FGF19 and fisogatinib⁶¹) the lack of clear actionable drivers typical of the genomics of HCC represents an intrinsic challenge to the delivery of precision medicine in liver cancer.⁶²

Similar to prognostic biomarkers, the choice of the most appropriate outcome to qualify the clinical validity of a biomarker is specific to stage and treatment modality and ideally should match the primary outcome required for regulatory approval of novel therapies.⁴⁹ For early stage (BCLC 0/A), the primary outcomes should be RFS and OS, with secondary outcomes time to recurrence and health-related quality of life. The development of molecular predictors of RFS after resection or ablation is expected to facilitate risk stratification for adjuvant therapy, mirroring paradigmatic, clinically available examples in breast cancer.⁶³ For studies testing interventions at intermediate or advanced stages (BCLC B or C), the primary outcomes should be OS and PFS, whereas objective response rate and health-related quality of life should be secondary outcomes. OS has been traditionally recommended as endpoint of choice in patients at advanced stages because of the cytostatic effect of most tyrosine kinase inhibitors, which led to low objective response rates.⁶⁴ However, the capacity of immune-based therapies to induce higher objective response rates (between 17%

and 28%⁶⁵⁻⁶⁷) has led to a growing interest in PFS as a primary endpoint for therapeutic trials in these patients. A threshold of 0.6 for the hazard ratio of PFS has been proposed as a surrogate for OS in HCC.^{49,63}

Rules for Incorporating New Biomarkers in Clinical Practice

It is key to rank the level of evidence to support the introduction of new biomarkers in clinical practice. We endorse a modified version of the levels of evidence proposed by Simon et al.¹⁶ in 2009. We have adapted these criteria to meet the needs for biomarker development across indications in HCC (Figure 3). Level 1 includes prospective studies designed to assess the value of the biomarker with a prespecified power calculation based on this outcome. The biomarker of interest can be assessed prospectively or retrospectively using archived samples, but in retrospective studies, the analysis of the biomarker should be prespecified in the original study protocol. In this scenario, the study must demonstrate that biomarker validity (bias and type-1 error in reporting biomarker performance) is not compromised using archived samples. In other words, one would expect to get the same conclusion, in terms of accuracy, if a prospective study would be conducted specifically designed to answer the study hypothesis for this biomarker. The key considerations are (1) the original study protocol defines the target population for which the new diagnostic biomarker is designed, (2) the biomarker assay is not affected by using archived vs new specimens, and (3) the data generated from the original study did not inform and bias this new diagnostic biomarker evaluation. To be considered level 1, studies using archived samples need to fulfill some conditions: (1) significant representation of samples in relation to the patients included in the trial, (2) robustness and reproducibility of the assay, and (3) study design must address the intended clinical use of the biomarker (prespecified data analysis, and assays to be conducted blinded to clinical data). There are examples of using retrospective samples to recommend predictive biomarkers in oncology (ie, hormone receptor status in breast cancer⁶⁸ or KRAS mutation in colorectal cancer⁶⁹).

For level 2, the study should be prospective but does not need to be designed to evaluate the biomarker. Prospective cohort studies evaluating risk stratification, early detection, or diagnostic biomarkers can be considered as level 1. Level 3 includes multiple retrospective observational studies, from independent investigators, using archived samples (traditionally known as “convenience samples”). These are specimens available to the researcher without any predetermined enrollment criteria. The study is designed to estimate clinical validity of the biomarker and follows a robust assay reproducibility framework. This also includes retrospective sample collections without prespecified selection criteria. Unfortunately, the vast majority of biomarker research in HCC derives from observational retrospective studies that test biomarker accuracy using convenience samples.⁷⁰ The target population may be heterogeneous due to the lack of a prespecified management protocol or standard follow-up plan. It is unlikely that such studies will comply with the requirements described previously when archived samples are used. In this context, the likelihood of potential bias and spurious associations increases.

The panel agrees on the need for level 1 evidence studies for the incorporation of new biomarkers for each setting in clinical practice guidelines. For studies evaluating prognostic

biomarkers, we recommend the use of the REMARK guidelines for reporting prognostic biomarkers in oncology.^{3,48} For treatment prediction biomarkers, besides rigorous qualification and validation, a number of additional considerations apply to the successful clinical implementation including cost-effectiveness and practicality of use.⁷¹ Full regulatory approval is mandatory for qualification as companion diagnostics, that is, biomarkers that are codeveloped with a drug to define its label, such as the PD-L1 IHC 22c3 pharmDx assay in lung cancer.⁷²

Conclusions

There is a clear need for biomarkers in HCC risk stratification, early detection, diagnosis, prognosis, and treatment response. However, rigorous evaluation is required, extending beyond level 3 evidence, for these biomarkers to be used in clinical practice and level 1 evidence is required to be incorporated into practice guidelines. Ongoing efforts such as the Hepatocellular Carcinoma Early Detection Strategy Study and Translational Liver Cancer Consortium are evaluating biomarkers for risk stratification and early detection, whereas advances in the HCC treatment landscape are facilitating biomarker evaluation for prognostication and treatment response. Our proposal provides guidance for these efforts on how best to perform rigorous biomarker evaluation in each of these areas.

Acknowledgments

Funding

Amit G. Singal's research is supported in part by National Cancer Institute U01 CA230694, R01 CA222900, and U01 CA226052. Yujin Hoshida's research is supported by the National Institutes of Health (NIH) (CA233794, CA226052), European Commission (ERC-2014-AdG-671231), Cancer Prevention and Research Institute of Texas (RR180016). David J. Pinato is supported by the Wellcome Trust Strategic Fund (PS3416), Cancer Research UK (Postdoctoral Bursary Grant C57701/A26137), and the Conquer Cancer, the ASCO Foundation (Young Investigator Award 2019) and acknowledges infrastructural support by the Cancer Research UK Imperial Centre and the Experimental Cancer Medicine Centre. Ziding Feng is supported in part by NIH U24CA230144 and U24CA086368. Nabihah Tayob is supported in part by NIH R01CA230503, U24CA230144, R01CA222900, and U01CA230694. Anna S. Lok is supported in part by NIH U01CA230669 and R01CA230503. Josep Llovet acknowledges funding from the Accelerator Award (CRUCK, AECC, AIRC) (HUNTER, Ref. C9380/A26813), National Cancer Institute (P30-CA196521), US Department of Defense (CA150272P3), Samuel Waxman Cancer Research Foundation, Spanish National Health Institute (PID2019-105378RB-100), and the Generalitat de Catalunya/AGAUR (SGR-1358). Augusto Villanueva is supported by the US Department of Defense (CA150272). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The funding agency had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation of the manuscript.

Conflict of interest

Amit G. Singal has served as a consultant for Wako Diagnostics, Glycotest, Exact Sciences, Roche, GRAIL, Genentech, Bayer, Eisai, Exelixis, AstraZeneca, BMS, and TARGET Pharmsolutions. Yujin Hoshida has served as advisory board member for Helio Health, consultant for Ferring Pharmaceuticals, research funding from Morphic Therapeutics, and shareholder of Alentis Therapeutics. David J. Pinato received lecture fees from ViiV Healthcare, Bayer Healthcare and travel expenses from BMS and Bayer Healthcare; consulting fees for Mina Therapeutics, EISAI, Roche, Astra Zeneca; received research funding (to institution) from MSD, BMS. Jorge Marrero has served as a consultant for Glycotest. Jean-Charles Nault received a research grant from Bayer for Inserm UMR1148. Young Suk Lim is an advisory board member of Bayer Healthcare and Gilead Sciences and received investigator-initiated research funding from Bayer Healthcare and Gilead Sciences. Ziding Feng received research funding from Exact Sciences. Anna S. Lok serves on the advisory board of Epigenomics. Josep Llovet receives research support from Bayer Healthcare Pharmaceuticals, Bristol-Myers Squibb, Boehringer Ingelheim, Eisai Inc, and Ipsen; and received consulting fees from AstraZeneca, Bayer Healthcare Pharmaceuticals, Bristol-Myers Squibb, Can-Fite Biopharma, Celsion Corporation, Eli Lilly, Eisai Inc, Exelixis, Genentech, Glycotest, Merck, Nucleix, and Roche. Augusto Villanueva receives consulting fees from Guidepoint, Fujifilm, Boehringer Ingelheim, FirstWord, and

MHLife Sciences; advisory board fees from Exact Sciences, Nucleix, Gilead, and NGM Pharmaceuticals; and research support from Eisai. The remaining authors disclose no conflicts.

Abbreviations used in this paper:

AFP	alpha-fetoprotein
BCLC	Barcelona Clinic Liver Cancer staging system
CT	computed tomography
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
FPR	false positive rate
MRI	magnetic resonance imaging
NASH	nonalcoholic steatohepatitis
OS	overall survival
PFS	progression-free survival
REMARK	Reporting recommendations for tumor marker prognostic studies
RFS	recurrence-free survival
TPR	true positive rate.

References

1. Robertson DJ, Lee JK, Boland CR, et al. Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2017;152:1217–1237.e3. [PubMed: 27769517]
2. Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–1466. [PubMed: 8841020]
3. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005;97:1180–1184. [PubMed: 16106022]
4. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001;93:1054–1061. [PubMed: 11459866]
5. Fujiwara N, Friedman SL, Goossens N, et al. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* 2018;68:526–549. [PubMed: 28989095]
6. Moon AM, Singal AG, Tapper EB. Contemporary epidemiology of chronic liver disease and cirrhosis. *Clin Gastroenterol Hepatol* 2020;18:2650–2666. [PubMed: 31401364]
7. Goossens N, Bian CB, Hoshida Y. Tailored algorithms for hepatocellular carcinoma surveillance: Is one-size-fits-all strategy outdated? *Curr Hepatol Rep* 2017;16:64–71. [PubMed: 28337405]
8. Rich NE, Hester C, Odewole M, et al. Racial and ethnic differences in presentation and outcomes of hepatocellular carcinoma. *Clin Gastroenterol Hepatol* 2019;17:551–559.e1. [PubMed: 29859983]
9. Athuluri-Divakar SK, Hoshida Y. Generic chemoprevention of hepatocellular carcinoma. *Ann N Y Acad Sci* 2019;1440:23–35. [PubMed: 30221358]
10. Goossens N, Singal AG, King LY, et al. Cost-effectiveness of risk score-stratified hepatocellular carcinoma screening in patients with cirrhosis. *Clin Transl Gastroenterol* 2017;8:e101. [PubMed: 28640287]

11. Mittal S, El-Serag HB, Sada YH, et al. Hepatocellular carcinoma in the absence of cirrhosis in United States veterans is associated with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2016;14:124–131.e1. [PubMed: 26196445]
12. Loomba R, Lim JK, Patton H, et al. AGA clinical practice update on screening and surveillance for hepatocellular carcinoma in patients with nonalcoholic fatty liver disease: expert review. *Gastroenterology* 2020;158:1822–1830. [PubMed: 32006545]
13. Farhang Zangneh H, Wong WWL, Sander B, et al. Cost effectiveness of hepatocellular carcinoma surveillance after a sustained virologic response to therapy in patients with hepatitis C virus infection and advanced fibrosis. *Clin Gastroenterol Hepatol* 2019;17:1840–1849.e16. [PubMed: 30580095]
14. Nathani P, Gopal P, Rich N, et al. Hepatocellular carcinoma tumour volume doubling time: a systematic review and meta-analysis. *Gut* 2021;70:401–407. [PubMed: 32398224]
15. Rich NE, John BV, Parikh ND, et al. Hepatocellular carcinoma demonstrates heterogeneous growth patterns in a multi-center cohort of patients with cirrhosis. *Hepatology* 2020;72:1664–1665.
16. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446–1452. [PubMed: 19815849]
17. Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 2018;68:723–750. [PubMed: 29624699]
18. Galle PR, Forner A, Llovet JM, et al. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2018;69:182–236. [PubMed: 29628281]
19. Tzartzeva K, Obi J, Rich NE, et al. Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. *Gastroenterology* 2018;154:1706–1718.e1. [PubMed: 29425931]
20. Simmons O, Fetzer DT, Yokoo T, et al. Predictors of adequate ultrasound quality for hepatocellular carcinoma surveillance in patients with cirrhosis. *Aliment Pharmacol Ther* 2017;45:169–177. [PubMed: 27862091]
21. Atiq O, Tiro J, Yopp AC, et al. An assessment of benefits and harms of hepatocellular carcinoma surveillance in patients with cirrhosis. *Hepatology* 2017;65:1196–1205. [PubMed: 27775821]
22. Singal AG, Tiro JA, Murphy CC, et al. Patient-reported barriers are associated with receipt of hepatocellular carcinoma surveillance in a multi-center cohort of patients with cirrhosis. *Clin Gastroenterol Hepatol* 2021;19:987–995.e1. [PubMed: 32629122]
23. Wolf E, Rich NE, Marrero JA, et al. Use of hepatocellular carcinoma surveillance in patients with cirrhosis: a systematic review and meta-analysis. *Hepatology* 2021;73:713–725. [PubMed: 32383272]
24. Ioannou GN, Beste LA, Green PK, et al. Increased risk for hepatocellular carcinoma persists up to 10 years after HCV eradication in patients with baseline cirrhosis or high FIB-4 Scores. *Gastroenterology* 2019;157:1264–1278.e4. [PubMed: 31356807]
25. Bertot LC, Adams LA. Trends in hepatocellular carcinoma due to non-alcoholic fatty liver disease. *Expert Rev Gastroenterol Hepatol* 2019;13:179–187. [PubMed: 30791782]
26. Ganne-Carrie N, Chaffaut C, Bourcier V, et al. Estimate of hepatocellular carcinoma incidence in patients with alcoholic cirrhosis. *J Hepatol* 2018;69:1274–1283. [PubMed: 30092234]
27. Sarasin FP, Giostra E, Hadengue A. Cost-effectiveness of screening for detection of small hepatocellular carcinoma in western patients with Child-Pugh class A cirrhosis. *Am J Med* 1996;101:422–434. [PubMed: 8873514]
28. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–1438. [PubMed: 18840817]
29. Berhane S, Toyoda H, Tada T, et al. Role of the GALAD and BALAD-2 serologic models in diagnosis of hepatocellular carcinoma and prediction of survival in patients. *Clin Gastroenterol Hepatol* 2016;14:875–886.e6. [PubMed: 26775025]
30. Poustchi H, Farrell GC, Strasser SI, et al. Feasibility of conducting a randomized control trial for liver cancer screening: is a randomized controlled trial for liver cancer screening feasible or still needed? *Hepatology* 2011;54:1998–2004. [PubMed: 21800340]

31. Villanueva A Hepatocellular carcinoma. *N Engl J Med* 2019;380:1450–1462. [PubMed: 30970190]
32. International Consensus Group for Hepatocellular Neoplasia. The International Consensus Group for Hepatocellular Neoplasia. Pathologic diagnosis of early hepatocellular carcinoma: a report of the International Consensus Group for Hepatocellular Neoplasia. *Hepatology* 2009;49:658–664. [PubMed: 19177576]
33. Llovet JM, Chen Y, Wurmbach E, et al. A molecular signature to discriminate dysplastic nodules from early hepatocellular carcinoma in HCV cirrhosis. *Gastroenterology* 2006;131:1758–1767. [PubMed: 17087938]
34. Tremosini S, Forner A, Boix L, et al. Prospective validation of an immunohistochemical panel (glypican 3, heat shock protein 70 and glutamine synthetase) in liver biopsies for diagnosis of very early hepatocellular carcinoma. *Gut* 2012;61:1481–1487. [PubMed: 22287594]
35. Di Tommaso L, Franchi G, Park YN, et al. Diagnostic value of HSP70, glypican 3, and glutamine synthetase in hepatocellular nodules in cirrhosis. *Hepatology* 2007;45:725–734. [PubMed: 17326147]
36. Kluger MD, Salceda JA, Laurent A, et al. Liver resection for hepatocellular carcinoma in 313 Western patients: tumor biology and underlying liver rather than tumor size drive prognosis. *J Hepatol* 2015;62:1131–1140. [PubMed: 25529622]
37. Llovet JM, Peña CEA, Lathia CD, et al. Plasma biomarkers as predictors of outcome in patients with advanced hepatocellular carcinoma. *Clin Cancer Res* 2012;18:2290–2300. [PubMed: 22374331]
38. Duvoux C, Roudot-Thoraval F, Decaens T, et al. Liver transplantation for hepatocellular carcinoma: a model including alpha-fetoprotein improves the performance of Milan criteria. *Gastroenterology* 2012;143:986–994.e3; quiz e14–5. [PubMed: 22750200]
39. Raoul J-L, Bruix J, Greten TF, et al. Relationship between baseline hepatic status and outcome, and effect of sorafenib on liver function: SHARP trial subanalyses. *J Hepatol* 2012;56:1080–1088. [PubMed: 22245896]
40. Montal R, Andreu-Oller C, Bassaganyas L, et al. Molecular portrait of high alpha-fetoprotein in hepatocellular carcinoma: implications for biomarker-driven clinical trials. *Br J Cancer* 2019;121:340–343. [PubMed: 31285588]
41. Villanueva A, Hoshida Y, Battiston C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011;140:1501–1512.e2. [PubMed: 21320499]
42. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 2008;359:1995–2004. [PubMed: 18923165]
43. Nault J-C, De Reyniès A, Villanueva A, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology* 2013;145:176–187. [PubMed: 23567350]
44. Zucman-Rossi J, Villanueva A, Nault J-C, et al. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* 2015;149:1226–1239.e4. [PubMed: 26099527]
45. Yamashita T, Forgues M, Wang W, et al. EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. *Cancer Res* 2008;68:1451–1461. [PubMed: 18316609]
46. Pinyol R, Montal R, Bassaganyas L, et al. Molecular predictors of prevention of recurrence in HCC with sorafenib as adjuvant treatment and prognostic factors in the phase 3 STORM trial. *Gut* 2019;68:1065–1075. [PubMed: 30108162]
47. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018;391:1301–1314. [PubMed: 29307467]
48. Sauerbrei W, Taube SE, McShane LM, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): an abridged explanation and elaboration. *J Natl Cancer Inst* 2018;110:803–811. [PubMed: 29873743]
49. Llovet JM, Villanueva A, Marrero JA, et al. Trial design and endpoints in hepatocellular carcinoma: AASLD Consensus Conference. *Hepatology* 2021;73(Suppl 1):158–191. [PubMed: 32430997]

50. Rebouissou S, Nault J-C. Advances in molecular classification and precision oncology in hepatocellular carcinoma. *J Hepatol* 2020;72:215–229. [PubMed: 31954487]
51. Rimassa L, Reig M, Abbadessa G, et al. Tumor biopsy and patient enrollment in clinical trials for advanced hepatocellular carcinoma. *World J Gastroenterol* 2017;23:2448–2452. [PubMed: 28428725]
52. Labgaa I, Villacorta-Martin C, D'Avola D, et al. A pilot study of ultra-deep targeted sequencing of plasma DNA identifies driver mutations in hepatocellular carcinoma. *Oncogene* 2018;37:3740–3752. [PubMed: 29628508]
53. Rimassa L, Assenat E, Peck-Radosavljevic M, et al. Tivantinib for second-line treatment of MET-high, advanced hepatocellular carcinoma (METIV-HCC): a final analysis of a phase 3, randomised, placebo-controlled study. *Lancet Oncol* 2018;19:682–693. [PubMed: 29625879]
54. Lim HY, Merle P, Weiss KH, et al. Phase II studies with refametinib or refametinib plus sorafenib in patients with RAS-mutated hepatocellular carcinoma. *Clin Cancer Res* 2018;24:4650–4661. [PubMed: 29950351]
55. Teufel M, Seidel H, Köchert K, et al. Biomarkers associated with response to regorafenib in patients with hepatocellular carcinoma. *Gastroenterology* 2019;156:1731–1741. [PubMed: 30738047]
56. Zhu AX, Kang Y-K, Yen C-J, et al. Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α -fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2019;20:282–296. [PubMed: 30665869]
57. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:3–14. [PubMed: 18813139]
58. Zhu AX, Baron AD, Malfertheiner P, et al. Ramucirumab as second-line treatment in patients with advanced hepatocellular carcinoma: analysis of REACH trial results by Child-Pugh score. *JAMA Oncol* 2017;3:235–243. [PubMed: 27657674]
59. Sargent DJ, Conley BA, Allegra C, et al. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005;23:2020–2027. [PubMed: 15774793]
60. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* 2011;1:44–53. [PubMed: 22586319]
61. Kim RD, Sarker D, Meyer T, et al. First-in-human Phase I study of fisogatinib (BLU-554) validates aberrant fibroblast growth factor 19 signaling as a driver event in hepatocellular carcinoma. *Cancer Discov* 2019;9:1696–1707. [PubMed: 31575541]
62. Schulze K, Nault J C, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J Hepatol* 2016;65:1031–1042. [PubMed: 27262756]
63. Duffy MJ, Harbeck N, Nap M, et al. Clinical use of biomarkers in breast cancer: updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer* 2017;75:284–298. [PubMed: 28259011]
64. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008;359:378–390. [PubMed: 18650514]
65. Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med* 2020;382:1894–1905. [PubMed: 32402160]
66. El-Khoueiry AB, Sangro B, Yau T, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 2017;389:2492–2502. [PubMed: 28434648]
67. Zhu AX, Finn RS, Edeline J, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol* 2018;19:940–952. [PubMed: 29875066]
68. Jensen EV, Block GE, Smith S, et al. Estrogen receptors and breast cancer response to adrenalectomy. *Natl Cancer Inst Monogr* 1971;34:55–70. [PubMed: 5140877]
69. Lievre A, Bachet JB, Le Corre D, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res* 2006;66:3992–3995. [PubMed: 16618717]

70. Hoshida Y, Moeini A, Alsinet C, et al. Gene signatures in the management of hepatocellular carcinoma. *Semin Oncol* 2012;39:473–485. [PubMed: 22846864]
71. Nalejska E, Maczy ska E, Lewandowska MA. Prognostic and predictive biomarkers: tools in personalized oncology. *Mol Diagn Ther* 2014;18:273–284. [PubMed: 24385403]
72. Sul J, Blumenthal GM, Jiang X, et al. FDA approval summary: pembrolizumab for the treatment of patients with metastatic non-small cell lung cancer whose tumors express programmed death-ligand 1. *Oncologist* 2016;21:643–650. [PubMed: 27026676]

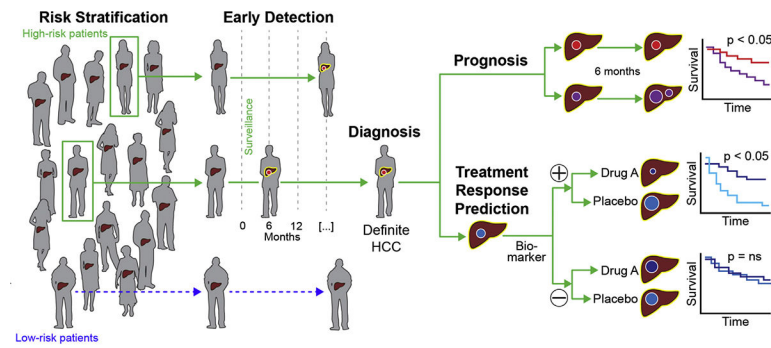


Figure 1.

Summary figure of the main clinical indications where biomarkers are needed to improve outcomes in patients with HCC. A risk stratification biomarker aims to predict future development of HCC and can differentiate high- and low-risk patients. An early detection biomarker aims to detect HCC at an early stage. A diagnostic biomarker can confirm or exclude presence of HCC in patients with clinical concern of HCC. A prognostic biomarker predicts cancer outcomes and can differentiate favorable vs poor prognosis (eg, survival). A treatment response biomarker aims to predict favorable or unfavorable response to treatment. Used with permission from ©Mount Sinai Health System.

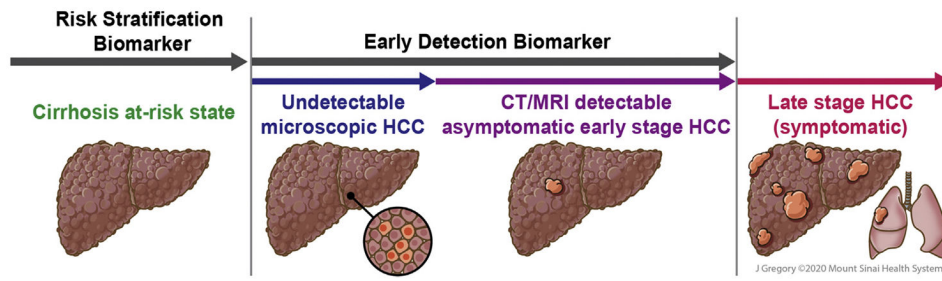


Figure 2.

Visual summary of the natural history of HCC that focus on key differences and challenges when designing studies to evaluate risk stratification vs early detection biomarkers. For risk stratification biomarkers, it is crucial to ensure that the patients are free of HCC. This includes a 2-year period to minimize the risk of undetectable microscopic HCC, which is a potential confounder in these studies. Used with permission from ©Mount Sinai Health System.

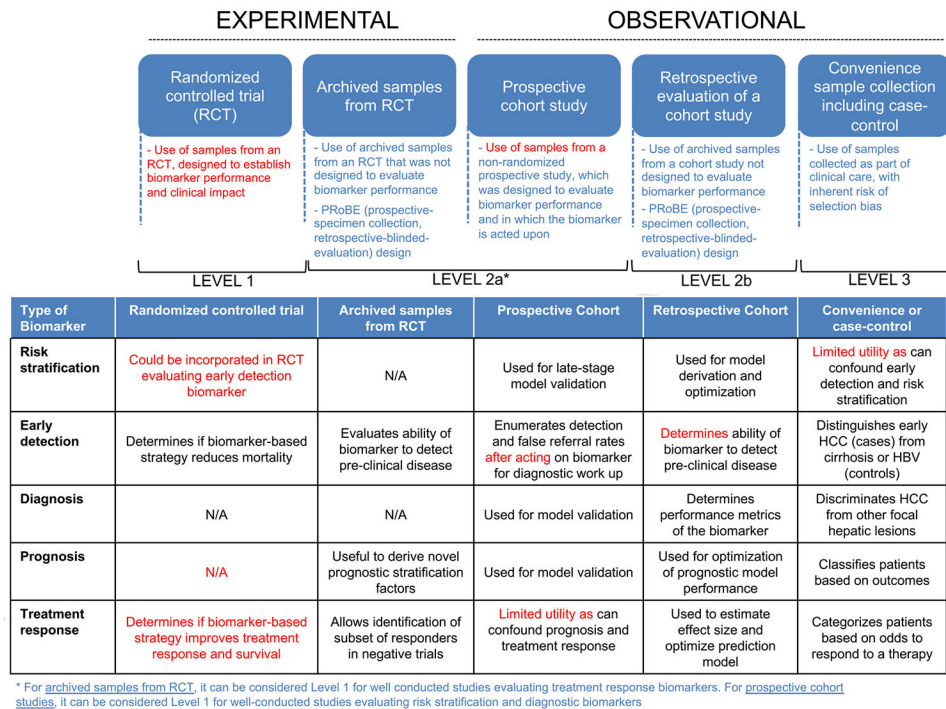


Figure 3. Levels of evidence in biomarker studies. (*Top*) Main study designs, key characteristics, and levels of evidence assigned for each design. (*Bottom*) Highlights the main outcomes and some key limitations of each study design.

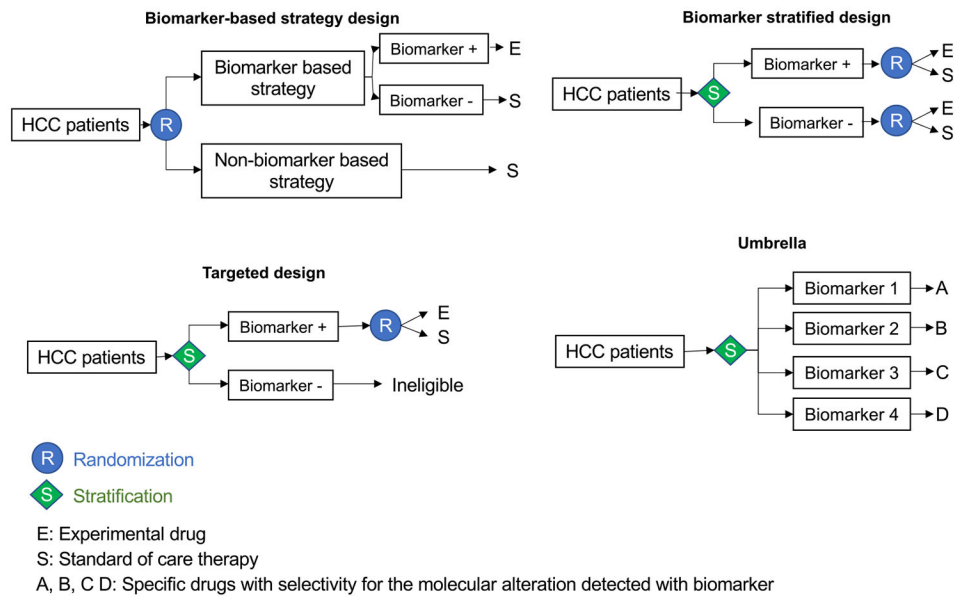


Figure 4. Biomarker-based clinical trial designs testing new therapies. Depiction of the 4 most frequently used clinical trial designs to test the utility of biomarkers as a selection tool for new therapies. In the biomarker-based strategy, patients are randomized to a biomarker-based strategy or no stratification before they are allocated to the treatment arm. In the biomarker-stratified design, patients are stratified based on the biomarker and then randomized to receive the experimental drug or placebo. Both, patients with the biomarker and those without are included in the study. In the targeted design, only the patients with the biomarker are randomized to the drug trial (eg, trial of ramucirumab in patients with high AFP levels⁵⁶). The umbrella trial design allocates specific therapies based on predicted molecular alterations detected with the biomarker, which generally includes an array of different molecular alterations.

Table 1.

Types of Biomarkers

Type of biomarker	Definition
Risk stratification	Biomarker that predicts the future development of HCC in an at-risk patient
Early detection	Biomarker that detects HCC at an early tumor stage, when therapy can be delivered which would reduce HCC-related mortality
Diagnosis	Biomarker that can confirm or exclude presence of HCC in patients with a suspicious nodule or clinical concern for HCC
Prognosis	Biomarker that predicts overall cancer outcome, regardless of treatment
Treatment response	Biomarker that predicts favorable or unfavorable response to a specific treatment

HCC, hepatocellular carcinoma.

Table 2.

Outcomes, Study Design, and Analysis for Early Detection of HCC

Phases	Outcomes	Study design	Analysis
1	Exploratory TPR and FPR	Preclinical case-control study	• TPR, FPR, ROC curves to identify potential biomarkers
2	Clinical assay TPR and FPR	Case-control study	• TPR, FPR, ROC curves • Comparison with AFP \pm ultrasound • Combinations of markers • Impact of covariates on biomarker performance
3	Detection of preclinical HCC	PRoBE (prospective specimen collection, retrospective blinded evaluation) design	• TPR, FPR, ROC curves • Comparison with AFP \pm ultrasound • Time between outcome and biomarker measurement • Combination of markers
4	HCC detection rates	Prospective cohort	• PPV • HCC Stage distribution • TPR, FPR, ROC curves • Benefits (early HCC detection) and harms (false positive rate) • Comparison with AFP \pm ultrasound
5	Decrease in HCC mortality	Randomized study	• Survival analysis and Cox regression • Cost and quality of life • Overdiagnosis and overtreatment

AFP, alpha-fetoprotein; FPR, false positive rate; HCC, hepatocellular carcinoma; PPV, positive predictive value; ROC, receiver operating characteristics curve; TPR, true positive rate.