An Overview of the Genomic Characterization of Hepatocellular Carcinoma

Sun Young Yim 1
Ju-Seog Lee 2

1 Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea; 2 Department of Systems Biology, The University of Texas M. D. Anderson Cancer Center, Houston, TX, 77030, USA

Abstract: Tumor classifications based on alterations in the genome, epigenome, or proteome have revealed distinct tumor subgroups that are associated with clinical outcomes. Several landmark studies have demonstrated that such classifications can significantly improve patient outcomes by enabling tailoring of therapy to specific alterations in cancer cells. Since cancer cells accumulate numerous alterations in many cancer-related genes, it is a daunting task to find and confirm important cancer-promoting alterations as therapeutic targets or biomarkers that can predict clinical outcomes such as survival and response to treatments. To aid further advances, we provide here an overview of the current understanding of molecular and genomic subtypes of hepatocellular carcinoma (HCC). System-level integration of data from multiple studies and development of new technical platforms for analyzing patient samples hold great promise for the discovery of new targets for treatment and correlated biomarkers, leading to personalized medicine for treatment of HCC patients.

Keywords: gene expression profile, RNA-seq, proteomics, mutations, genomic subtypes

Introduction

The development of hepatocellular carcinoma (HCC) is a highly complex process involving the accumulation of epigenetic and genetic alterations in numerous genomic loci, leading to dramatic changes in signaling pathways. Adding further complexity, genetic aberrations of HCC show considerable heterogeneity derived from differences in ethnicity, etiology, and environmental exposures.1–4 The complex linkages between various risk factors and hepatocarcinogenesis were not well understood until the recent developments in genomic technologies that have enabled us to gain a systemic overview of genomic profiles, providing insight into the pathogenesis of HCC. Substantial efforts have been carried out to uncover molecular subtypes of HCC with differences in clinical outcomes.

HCC arises in the background of a fibrotic or cirrhotic liver following infection with hepatitis C or B viruses; it less commonly occurs with other conditions such as alcohol abuse, hemochromatosis, and non-alcoholic steatohepatitis.5,6 These conditions trigger a series of genetic and epigenetic alterations that lead to transformation of normal cells into pre-malignant cells.7 Rapid progress in the use of genomic and proteomic data for understanding HCC development at the molecular level has brought the promise of new clinical utility for treating HCC patients. In analyzing genomic and proteomic data, there are two approaches for discovering subtypes correlated with clinical characteristics.8 In a supervised
approach, the data are arranged to uncover a set of variables, such as genetic alterations or expression of genes, that are significantly associated with clinical outcomes such as response to treatment, overall survival, and recurrence. An unsupervised approach seeks to discover new subtypes of HCC patients that have not previously been recognized. The ultimate goal of both of these approaches is to uncover the underlying biology of disease that is associated with clinical outcomes and identify reliable markers for classifying HCC patients for better management in the clinic.

While a wide range of options, including several multi-targeted tyrosine kinase inhibitors (TKIs) and immune checkpoint inhibitors (ICIs), can be offered for the treatment of HCC, not all patients benefit from these treatments, and response rates to these systemic treatments are typically lower than those observed in other cancer types. However, recent success in a clinical trial of a combination of a targeted drug (bevacizumab) and an ICI (atezolizumab) encourages us to continue seeking better combination treatments in future studies. Although the effects of this new combination treatment were very promising, many patients still did not respond, further emphasizing that the identification and validation of biomarkers or genetically homogenous subtypes that can help to identify better responders to treatment remain of critical importance.

**Mutational Landscape of HCC**

There have been many genomic studies for cataloguing driver genes activated in HCC (Table 1). The *TERT* gene encodes an enzymatic subunit of telomerase that is required for preserving telomere length and plays important roles in cancer. TERT expression is largely suppressed in most somatic cells, except for stem cells and regenerating hepatocytes. However, cancer cells express TERT stably, which is reactivated and essential for limitless growth of cancer cells. *TERT* is regulated transcriptionally, and the promoter region has several binding sites for transcription factors such as MYC and Sp1. Unlike other oncogenes, in which most mutations are in coding regions, the most frequent mutations in *TERT* were found in the promoter sequences providing novel insight into the means of tumor-specific activation of *TERT*. In fact, more than 50% of HCC tumors have mutations in the *TERT* promoter, making it one of the most frequently mutated sites in HCC. Mutated sequences become a potential binding site for ETS/TCF transcription factors which can increase promoter activity and expression of TERT.

**Table 1 Most Frequently Mutated Genes in HCC (from TCGA Study)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>MutSig (q-value)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>3.12E-12</td>
<td>30.80%</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>3.12E-12</td>
<td>26.00%</td>
</tr>
<tr>
<td>ALB</td>
<td>3.37E-10</td>
<td>11.50%</td>
</tr>
<tr>
<td>APOB</td>
<td>5.60E-05</td>
<td>10.50%</td>
</tr>
<tr>
<td>ARID1A</td>
<td>1.99E-08</td>
<td>8.60%</td>
</tr>
<tr>
<td>AXIN1</td>
<td>3.12E-12</td>
<td>6.40%</td>
</tr>
<tr>
<td>ALMS1</td>
<td>5.03E-02</td>
<td>6.40%</td>
</tr>
<tr>
<td>ARID2</td>
<td>1.85E-04</td>
<td>5.90%</td>
</tr>
<tr>
<td>RBL1</td>
<td>2.07E-10</td>
<td>5.60%</td>
</tr>
<tr>
<td>BAP1</td>
<td>1.05E-09</td>
<td>5.60%</td>
</tr>
<tr>
<td>PTPRQ</td>
<td>5.13E-02</td>
<td>5.10%</td>
</tr>
<tr>
<td>FMN2</td>
<td>5.72E-02</td>
<td>4.80%</td>
</tr>
<tr>
<td>KEAP1</td>
<td>1.06E-07</td>
<td>4.60%</td>
</tr>
<tr>
<td>CDC27</td>
<td>5.97E-07</td>
<td>4.00%</td>
</tr>
<tr>
<td>NEFH</td>
<td>1.82E-02</td>
<td>4.00%</td>
</tr>
<tr>
<td>NRDC</td>
<td>1.34E-04</td>
<td>3.80%</td>
</tr>
<tr>
<td>RPS6KA3</td>
<td>2.03E-04</td>
<td>3.80%</td>
</tr>
<tr>
<td>JAK1</td>
<td>2.45E-02</td>
<td>3.80%</td>
</tr>
<tr>
<td>ADCY2</td>
<td>4.23E-02</td>
<td>3.80%</td>
</tr>
<tr>
<td>NFE2L2</td>
<td>3.12E-12</td>
<td>3.50%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>1.07E-03</td>
<td>3.50%</td>
</tr>
<tr>
<td>PCDH16</td>
<td>8.06E-03</td>
<td>3.50%</td>
</tr>
<tr>
<td>BCLAF1</td>
<td>1.24E-02</td>
<td>3.50%</td>
</tr>
<tr>
<td>KCNN3</td>
<td>7.66E-06</td>
<td>3.20%</td>
</tr>
<tr>
<td>IL6ST</td>
<td>8.35E-05</td>
<td>3.20%</td>
</tr>
<tr>
<td>KRT2</td>
<td>1.15E-04</td>
<td>3.20%</td>
</tr>
<tr>
<td>PTEN</td>
<td>2.03E-04</td>
<td>2.90%</td>
</tr>
<tr>
<td>ACVR2A</td>
<td>4.72E-04</td>
<td>2.90%</td>
</tr>
<tr>
<td>CNGA3</td>
<td>1.84E-03</td>
<td>2.90%</td>
</tr>
<tr>
<td>IRX1</td>
<td>8.40E-03</td>
<td>2.90%</td>
</tr>
<tr>
<td>FAM47A</td>
<td>1.18E-02</td>
<td>2.90%</td>
</tr>
<tr>
<td>NLGN1</td>
<td>2.32E-02</td>
<td>2.90%</td>
</tr>
<tr>
<td>KCNB2</td>
<td>7.08E-02</td>
<td>2.90%</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>1.86E-09</td>
<td>2.70%</td>
</tr>
<tr>
<td>EEF1A1</td>
<td>1.24E-03</td>
<td>2.70%</td>
</tr>
<tr>
<td>ATXN1</td>
<td>1.76E-03</td>
<td>2.70%</td>
</tr>
<tr>
<td>BRD7</td>
<td>2.40E-03</td>
<td>2.70%</td>
</tr>
<tr>
<td>KRT10</td>
<td>1.85E-06</td>
<td>2.40%</td>
</tr>
<tr>
<td>ZNF714</td>
<td>4.51E-03</td>
<td>2.40%</td>
</tr>
<tr>
<td>ADAM12</td>
<td>4.42E-02</td>
<td>2.40%</td>
</tr>
<tr>
<td>IDH1</td>
<td>3.11E-03</td>
<td>2.10%</td>
</tr>
<tr>
<td>OR8H2</td>
<td>2.45E-02</td>
<td>2.10%</td>
</tr>
<tr>
<td>HNF1A</td>
<td>5.72E-02</td>
<td>2.10%</td>
</tr>
<tr>
<td>BCL11B</td>
<td>7.72E-02</td>
<td>2.10%</td>
</tr>
</tbody>
</table>

**Notes:** Frequently mutated genes were selected from mutation data of The Cancer Genome Atlas (TCGA) study mutation frequency represents percentage from 373 tumors. Genes with mutation frequency more than 2% and q-value of mutation significance < 0.05 were selected.

**Abbreviation:** MutSig, mutation significance.
suppressor gene that, to preserve the genome integrity, can initiate apoptosis, cell-cycle arrest, senescence, and ferroptosis in response to cellular stresses such as genomic damage, activation of oncopgenes, and hypoxia.\(^4\)\(^2\)\(^-\)\(^4\(^5\) TP53 can function both as an activator and a repressor of gene expression, controlling expression of a large number of genes involved in ferroptosis, apoptosis, cell cycle arrest, and senescence. However, recent studies showed that TP53 also regulates expression of genes playing key roles in stem cell maintenance, oxidative stress, autophagy, and cellular metabolism.\(^4\(^6\) The vast majority of TP53 mutations are located in its DNA-binding region, causing a weak interaction of TP53 with the promoters of target genes. However, some mutations are gain-of-function alterations that give rise to new molecular activity that are not seen with wild-type TP53, such as metastasis, angiogenesis, and resistance to standard therapies.\(^4\(^6\)\(^7\) As would be expected, mutations in TP53 are significantly correlated with poor survival of HCC patients.\(^4\(^1\)

**CTNNB1** (coding β-catenin) is a subunit of the cellular cadherin complex acting as a downstream effector of the WNT pathway.\(^4\(^8\)\(^4\(^9\) In the absence of WNT signaling, the intracellular level of β-catenin is low due to phosphorylation-dependent ubiquitination/degradation that is regulated by the APC-AXIN-CK1-GSK3 complex. When WNT signaling is active, stabilized β-catenin is translocated to the nucleus and forms a complex with TCF/LEF transcription factors to turn on the expression of genes regulating cell proliferation. Aberrant activation of β-catenin due to mutations has been observed in more than 20% of HCC tumors.\(^2\(^1\)\(^-\)\(^2\(^3\)\(^5\(^0\) Most mutations are located in or near phosphorylation sites,\(^2\(^1\)\(^-\)\(^2\(^3\)\(^4\(^9\)\(^5\(^1\) enabling β-catenin to escape phosphorylation-dependent degradation. Thus, mutated β-catenin is constitutively activated. These observations indicate that tumors with mutated β-catenin may represent a clinically different subtype of HCC. In addition to β-catenin, **AXIN1** is another frequently mutated gene in the WNT pathway in HCC (occurring in 5–10% of tumors).\(^1\(^6\)\(^-\)\(^2\(^0\)\(^-\)\(^2\(^3\)\(^5\(^2\) Intriguingly, **CTNNB1** mutations are most significantly associated with **TERC** mutations, but are almost mutually exclusive from mutations in **TP53**,\(^1\(^6\)\(^1\)\(^4\)\(^1\)\(^9\)\(^3\(^9\) suggesting a potential working interface among these three genes in HCC development.

Mutations in **ARID1A** and **ARID2** are also frequently observed in HCC (10% of tumors).\(^1\(^4\)\(^2\(^0\)\(^-\)\(^2\(^2\) Both genes are members of the ARID family.\(^5\(^4\) By interacting with several other proteins, they make DNA accessible to cellular complexes regulating DNA replication, DNA repair, and transcription. Mutations in **ARID1A** are also frequently observed in many other cancer types, such as gastric, esophageal, breast, gynecologic, prostate, bladder, and pancreatic cancers.\(^3\(^5\)\(^-\)\(^5\(^4\)\(^-\)\(^5\(^8\) The vast majority of mutations in **ARID1A** appear to be loss-of-function mutations, strongly indicating that **ARID1A** is a tumor suppressor. Intriguingly, **Arid1a**-knockout mice showed a surprising and complex phenotype: unexpectedly, HCC development was significantly reduced in these mice,\(^5\(^9\) indicating that **Arid1a** may play roles in transformation of hepatocytes. In contrast to the earlier stage of HCC development, in the later stage **Arid1a** knockout enhanced progression and metastasis of HCC tumors, also suggesting that the roles of **Arid1a** in HCC development are complicated and context dependent.\(^5\(^9\) Tumor-suppressive activity of **ARID1A** in the liver was further supported by recent analysis of genomic data from tumors. Inactivation of **ARID1A** was significantly correlated with shorter overall survival of patients and consistently related to poor-prognostic genomic subtypes such as high-recurrence subtypes, hepatic stem cell, highly proliferative, and inactivation of the Hippo pathway.\(^6\(^0\) Intriguingly, The Cancer Genome Atlas (TCGA) project showed that **ARID1A** and **ARID2** mutations were mutually exclusive from TP53 mutations, suggesting that these genes may interact.\(^2\(^2\) However, more studies are needed to understand how **TP53** interacts with **ARID1A** or whether **TP53** can modulate **ARID1A** or **ARID2** activity.

Reactive oxygen species (ROS) are typically generated by cellular metabolism and can cause mutations leading to the initiation of tumorigenesis by interacting with genomic DNA.\(^6\(^1\) **NRF2** (NFE2L2), a member of the leucine zipper transcription factor family, is a major defense gene against cellular oxidative stress.\(^6\(^2\)\(^-\)\(^6\(^4\) In normal conditions, cellular NRF2 is maintained at a low level by KEAP1, which negatively regulates NRF2 by targeting it for E3 ligase–mediated ubiquitination and proteasomal degradation. Under oxidative conditions from metabolic stress, NRF2 is stabilized by inactivation of KEAP1 and translocated to the nucleus, in which it binds to the promoters of the downstream genes and increases their expression. Because NRF2 can prevent damage of genomic DNA and mutations, it was previously considered a tumor suppressor. However, recent studies showed that NRF2 can protect not only untransformed cells but also transformed cancer cells, resulting in protection of cancer cells from oxidative stress–mediated cell death.\(^6\(^5\)\(^-\)\(^6\(^7\) While **KEAP1** and **NRF2** are not highly mutated genes in HCC (3–5% of
NRF2 is significantly amplified and overexpressed in HCC. Importantly, mutations in NRF2 are located in the KEAP1-binding sites, which are necessary for degradation of NRF2, showing that mutations make NRF2 constitutively active in HCC. In good agreement with this, several KEAP1 mutations cause loss of its interaction with NRF2, making NRF2 constitutively active. During tumor progression, high metabolic activity in cancer cells substantially increases the burden of oxidative stress that causes apoptotic cell death or growth arrest. Thus, activation of NRF2 would deliver great benefits to cancer cells under oxidative stress. Several studies showed that NRF2 activation is significantly associated with shorter survival as a result of high cellular level of cytoprotective proteins in cancer cells, further supporting the notion that NRF2 is an important gene for protecting cancer cells. Once cells are transformed, cancer cells steal the NRF2 system to decrease cellular stress and have an advantage in growth. In addition, NRF2 can translocate to the nucleus to form dimers with sMAF proteins and bind antioxidant response elements for the transcription of its target genes. Interestingly, a recent drug screening study showed that NRF2 activity is potentially regulated by the oncosuppressor GNMT. Taken together, these observations show that NRF2 is a main regulator of the antioxidative and detoxification responses and propels the progression of cancer, formation of metastases, and resistance to therapy.

In addition to mutations in cancer-related genes, HCC tumors harbor mutations in serum proteins such as albumin (ALB) and apolipoprotein B (APOB), which are not frequently mutated in other cancer types. Because many mutations appear to confer loss of function, they may play roles in the development of HCC. However, the functional consequences of these mutations have not yet been examined. Interestingly, mutations in APOB were also discovered in familial hypobetalipoproteinemia (FHBL) with reduced level of low-density lipoprotein cholesterol in plasma. High incidence of hepatic steatosis, liver cirrhosis, and hepatocarcinoma have been observed in individuals with FHBL attributed to mutations in APOB. In a recent study from our lab, we carried out an analysis with systematic integration of genomic data from mouse models and human HCC tumors and found that APOB ablation in HCC is associated with poorer survival of HCC patients. Interestingly, silencing of APOB expression with siRNAs significantly increased cell growth in multiple HCC cells, indicating that APOB may have weak tumor-suppressive activity. However, the underlying biology of APOB-mediated growth suppression is currently unknown.

Genomic Subtypes of HCC with Prognostic Association

Recurrence of HCC is a very serious problem after curative removal of the primary HCC and occurs in around half of cases within 3 years after the resection. Many studies have used supervised or unsupervised approaches to genomic data to identify genes or subgroups associated with recurrence or overall survival.

Iizuka et al analyzed gene profiling data from a small number of HCC tumors as a training set; they identified 12 recurrence-associated genes in a supervised learning manner and generated a prediction model. When the prediction model was tested in an independent cohort, it identified 93% of patients with early recurrence. This was the first study that demonstrated the prognostic association of genomic markers in HCC. Similarly, Kurokawa et al identified genes associated with recurrence by analyzing gene expression in HCC tumors from 100 patients. Their trained model with 20 genes also successfully identified patients with early recurrence in an independent test cohort.

By applying a supervised machine-learning algorithm, Ye et al identified 153 genes significantly associated with intrahepatic metastasis of HCC. They showed that primary tumors with metastatic nodules had gene expression patterns similar to those of their surrounding metastatic tumors, suggesting that activation of genes favoring invasion began in the primary HCC tumors. In a functional validation experiment, an antibody specific to osteopontin, a leading gene product in the signature, successfully inhibited invasion of HCC cells and suppressed metastasis of HCC tumors in a mouse model. This genomic metastasis predictor was further validated in an independent cohort. In a Cox analysis that included clinical features, the metastasis signature was a significant prognostic indicator, particularly for early recurrence after treatment, but not for late recurrence, which is mainly attributed to the field effect in remnant livers.

Another study applied a similar approach to selected genes associated with early recurrence after curative treatment. Around 600 genes were selected from hepatitis B–associated HCC tumors, and their prognostic significance was tested in a validation cohort of HCC patients.
Network mapping analysis with selected genes discovered that transcription factor SP1 is a potential upstream regulator of genes correlated with high risk of early recurrence.

By using an unbiased approach for genomic data from primary tumors, investigators at the National Cancer Institute (NCI) identified two subtypes that are distinctive in terms of overall survival of the patients. Their statistical analysis yielded 947 genes correlated with overall survival. As gene set enrichment analysis of selected genes revealed that proliferation of cancer cells is most characteristic of a poor-prognostic subtype, it was named the NCI Proliferation (NCIP) signature. The poor-prognostic subtype showed elevated expression of genes related to histone modification and ubiquitination. The robustness of the signature in identifying patients with poor prognosis was further validated in multiple studies.

As part of a follow-up study of the NCIP signature, Kim et al identified a small number of genes that were significantly correlated with patients’ prognosis by using two previously discovered prognostic signatures. This meta-analysis selected 65 genes that were conserved in both signatures. The authors developed recurrence-risk scoring with a dynamic range from 0 to 100 for better translation of the signature to the clinic. In validation cohorts, the score was a highly significant indicator of both overall survival and recurrence-free survival. Gene network analysis with integration of mutation data revealed that mutations in β-catenin were significantly associated with low risk scores and activations of the AKT and IGF1R pathways were significantly associated with high risk scores.

Vascular invasion is significantly correlated with prognosis, as reflected in high recurrence after treatment. A large-scale analysis with >1000 cases clearly demonstrated that hepatic vascular invasion should be a critical criterion in selecting patients for curative treatment. Minguez et al identified 35 genes whose expression is significantly correlated with hepatic vascular invasion and developed a model that can predict HCC patients with vascular invasion. One of the best correlated genes is CD24, which is expressed in endothelial cells. Interestingly, metabolic genes were negatively associated with vascular invasion.

Systematic analysis of HCC genomic data identified five genes (RAMP3, TAF9, KRT19, HN1, and RAN) with strong prognostic association and developed predictive scoring. Because these five genes well reflect dysregulation of key signaling pathways in HCC, the authors proposed that the new score can be used for improving selection of patients for liver transplantation by extending the Milan criteria for better-prognostic tumors even if the size is larger than 5 cm.

**Genomic Subtypes of HCC Associated with Hepatic Stem Cell Features**

The HCC subtype with hepatic stem cell features was discovered by co-analysis of genomic data from primary HCC and fetal liver stem cells. Intriguingly, this subtype is a subset of the previously discovered high-proliferation subtypes in the NCIP classification. Informatics analysis of the signature predicted that JUNB, FOS, and FOSL2 are highly active in the hepatic stem cell subtype. High similarity of the signature to stem cells indicates that this subtype might be derived from stem or progenitor cells in the liver. In agreement with this notion, expression of marker genes for hepatic progenitor cells, such as VIM, KRT19, and KRT7, is highly elevated in this subtype.

EpCAM is expressed predominantly in hepatic stem or progenitor cells. In an effort to discover a subgroup of HCC tumors with hepatic stem cell features, investigators selected EpCAM-positive HCC and identified 70 genes whose expression correlated with EpCAM. With use of the EpCAM signature reflecting different hepatic cell lineages, authors suggested classifying HCC tumors into four subtypes: EpCAM+/AFP+ (hepatic stem cell–like), EpCAM+/AFP− (bile duct epithelium–like), EpCAM−/AFP+ (hepatocytic progenitor–like), and EpCAM−/AFP− (mature hepatocyte–like). Expression of marker genes for hepatic stem cells, such as KIT and KRT19, is elevated in hepatic stem cell–like HCC. In contrast, expression of marker genes for mature hepatocytes, such as CYP3A4, is higher in mature hepatocyte–like HCC. Not surprisingly, EpCAM+ tumors are highly invasive, and EpCAM increased motility and invasion of HCC cells in functional assays.

By analyzing genomic data from HCC and cholangiocarcinoma tumors together, Woo et al discovered that a fraction of HCC tumors shared gene expression patterns specific to cholangiocarcinoma and called the new subtype cholangiocarcinoma-like HCC (CLHCC). The most distinct features of these tumors are elevated expression of hepatic progenitor cell markers such as PROM1, EpCAM,
and KRT19. Not surprisingly, validation of the CLHCC subtype in several independent cohorts showed that the subtype is significantly associated with poor survival of patients. Comparison of the signature with other studies showed that it is highly similar to multiple embryonic stem cell signatures and the previously discovered hepatic stem cell signature.\(^{86,99}\)

Another group of investigators used a similar approach for finding genes correlated with hepatic progenitor or stem cells and developed a stem cell predictive signature.\(^{100}\) The signature was then tested in independent cohorts by using real-time RT-PCR with RNA from tumor tissues first and later by immunostaining paraffin-embedded tissues.

### Genomic Subtypes of HCC Associated with Underlying Liver Biology

Hoshida et al found in a systemic review of multiple genomic studies that three subtypes of HCC tumors—namely, S1, S2, and S3—could be identified. They showed substantial association with conventional clinical parameters such as blood AFP level, cellular differentiation status, and number of tumors.\(^{101}\) Subtypes S1 and S2 are associated with poor prognosis, and subtype S3 is associated with better prognosis, as reflected in smaller and more differentiated tumors and preserved liver function. Gene set analysis showed that the TGF-β pathway is activated in subtype S1, while the WNT/β-catenin pathway is activated in subtype S3. Most interestingly, subtype S2 is associated with characteristics of hepatic stem cells as reflected by high expression of hepatic stem cell markers such as AFP, EPCAM, and GPC3. Moreover, subtype S2 is strongly associated with a vascular invasion signature and hepatoblastoma-like signature.\(^{89,96,97}\)

Six genomic subtypes of HCC with distinct characteristics of genetic alterations—namely, G1 to G6—were uncovered by an unsupervised approach.\(^{102}\) Subtypes G1, G2, and G3 were associated with low chromosomal stability compared to the rest of the subtypes. \(CTNNB1\) is the most frequently mutated gene in subtypes G5 and G6, while \(TP53\) is the most frequently mutated gene in subtypes G2 and G3. \(PIK3CA\) is the most frequently mutated gene in subtype G2. Promoter regions of \(CDKN2A\) and \(CDH1\) are most frequently methylated in subtypes G3, G5, and G6.

Co-analysis of gene expression data with genomic copy number alteration from the same tumors revealed five subtypes: namely, poly 7, proliferation, INF, CTNNB1, and unannotated subtypes.\(^{103}\) Subtype Poly 7 showed unique features: chromosome 7 polysomy and the concomitant high expression of genes on that chromosome. Interestingly, these tumors are further characterized by lack of chromosome 8q amplification, which is one of the most frequent gains of chromosomes in HCC.\(^{104,105}\)

A subtype with activated β-catenin was characterized by activation mutations in \(CTNNB1\). IGFBP and RPS6 phosphorylation is enriched in the subtype with proliferation features. Interestingly, tumors in the INF-related subtype are smaller than those of other subtypes.

Analysis with pooled genomic data from HCC tumors revealed four molecular subtypes that are well correlated with liver zonation: namely, the perivenous (PV) subtype, perportal (PP) subtype, stem cell (STEM) subtype, and extracellular matrix (ECM) subtype.\(^{106}\) Subtype PV is associated with frequent mutations in \(CTNNB1\) and shows elevated expression of genes typical for hepatic zone 3 such as \(HAL\), \(VNII\), and \(GLUL\). Similarly, tumors in subtype PP express genes typical for hepatic zone 1, such as \(GLS2\) and \(ARG1\). The STEM subtype is highly similar to the previously discovered hepatic stem cell subtype.\(^{86}\) The ECM subtype is characterized by high expression of genes involved in extracellular matrix remodeling and epithelial-mesenchymal transition.

Recent analysis of TCGA HCC genome data uncovered interesting subtypes that were not previously recognized. They discovered that tumors with IDH1 and IDH2 mutations have a highly distinct gene signature.\(^{22}\) Intriguingly, a substantial fraction of tumors lacking IDH mutations have a gene signature similar to those with IDH mutations. When directly compared with previously identified genomic subtypes, the IDH-like subtype showed the highest resemblance to the poor-prognostic hepatic stem cell subtype.\(^{86}\) Likewise, these tumors exhibited resemblance to the S2 subtype of Hoshida’s classification\(^{101}\) and the CLHCC subtype,\(^{98}\) and had very high scores using the 65-gene-based recurrence risk score classifier.\(^{35}\)

Unlike previous studies, they used the iCluster strategy, which uses most of the collected genomic data, including mRNA expression, mutations, miRNA expression, copy number alteration, and DNA methylation data.\(^{107}\) This new approach uncovered three subtypes: namely, iC1, iC2, and iC3.\(^{22}\) Subtype iC1 is associated with female sex, Asian ethnicity, and younger age. These tumors also
show poor-prognostic characteristics, such as macrovascular invasion and higher tumor grade. Moreover, subtype iC1 is associated with low frequency of CTNNB1 mutations, CDKN2A silencing, and mutations in the TERT promoter. Unlike iC1, subtypes iC2 and iC3 are high in CTNNB1 mutations, CDKN2A silencing, HNF1A mutation, and TERT promoter mutations. Subtype iC2 is correlated with clinical variables such as less microvascular invasion and low-grade tumors. Subtype iC3 is associated with higher mutations of TP53, low chromosomal stability, and under-methylation. In comparison with Hoshida’s classification, iC1 is highly similar to subtype S2, whereas iC3 is highly similar to subtype S3.

**Proteomic Subtype**

Although genomic approaches uncovered many HCC subtypes with significant association with clinical features and unique underlying biology at the molecular level, these studies are limited by lack of insight on protein activity that is the functional end product of all genetic elements coupled with phenotypes. Moreover, post-translational modifications of gene products such as ubiquitination, phosphorylation, and methylation that are critical in regulating cellular activity cannot be faithfully represented solely by profiling the genome.

By applying an unsupervised approach to HCC proteomic data generated with use of mass spectrometry, a recent study uncovered three proteomic subtypes: namely, subtypes S-I, S-II, and S-III.\(^23\) Subtype S-I is characterized by high expression of metabolism-related proteins typical for retaining liver function. Not surprisingly, among the three proteomic subtypes, S-I has the best prognosis. As seen in the genomic subtype with good prognosis, tumors with subtype S-I showed very high genome stability. Subtype S-III is characterized by increased expression of proteins regulating cell proliferation, such as MKI-67, PCNA, TOP2A, and PARP1, and is associated with poor prognostic clinical features such as larger tumor size, tumor thrombus, and advanced tumor stage. In agreement with the expression patterns of proliferation proteins, patients with subtype S-III showed the poorest survival after surgery. Unlike S-I and S-III, subtype S-II lacks a specific proteomic signature, as reflected in having an intermediate expression pattern of metabolic and proliferative proteins. However, S-II is characterized by downregulation of immunity-related proteins such as CD4, ITGB3, and CD8A.

Not surprisingly, some of the proteomic subtypes were subsets of previously recognized genomic subtypes (Figure 1). Subtype S-III is highly similar to high-proliferation subtype of the NCIP classification, while S-I and S-II are subsets of low-proliferation subtype. Likewise, subtype S-III is similar to EpCAM-high HCC tumors, while S-I and S-II are subsets of EpCAM-low tumors. Interestingly, subtype S-III is split into S1 and S2 of Hoshida’s classification, indicating that genomic data can help to identify subsets of proteomic subtypes. More interestingly, no proteomic subtype showed similarity to the TCGA iC3 subtype, indicating that the biological features in subtype iC3 may not be well reflected in proteomic data. Together, these relationships suggest that no single-platform data set can fully represent clinical and biological features of HCC tumors. As data from both platforms are complementary to each other, we should analyze them together to achieve a full spectrum of tumor features.

**Summary and Discussion**

Genomic, epigenomic, and proteomic analyses of HCC tumors have discovered clinically distinct molecular subtypes, characteristic molecular changes correlated with subtypes, and prediction models that can identify patients in each subtype. These new analyses have begun to advance our understanding of oncogenesis as well as clinical decision-making for treatment of patients.

Although some of these findings are very promising, there are considerable barriers in moving discovered subtypes to clinical use. Even though these genomic or proteomic subtypes are significantly associated with clinical outcomes, their clinical usefulness is hindered by lack of concordance, which is possibly because of difference in classification algorithms, etiology of patients, collection and storage of samples, and technological platforms used. However, several subtypes have been rediscovered multiple times. For example, the hepatic stem cell subtype is a subset of the high-proliferation NCIP subtype and similar to EpCAM\(^+\) tumors, IDH-like HCC, and the CLHCC subtype.\(^22\) The S3 subtype in Hoshida’s classification is highly similar to the CTNNB1 subtype from Barcelona’s classification and the PV subtype.\(^103\) Despite substantial differences among molecular classifications, similarities among subsets of tumors strongly indicate that it is possible to uncover good consensus from independently defined classification of HCC tumors.
A major limitation of these subtypes is lack of clinically useful knowledge that is necessary for treating HCC patients. For example, although sorafenib has been used for treatment of HCC patients for a long time, no markers and subtypes have yet been found to be useful in predicting response to sorafenib treatment. Similarly, many other treatments for HCC also lack predictive markers associated with molecular subtypes. The development of antibodies blocking ICIs such as CTLA-4 or PD-1/PD-L1 represents a major breakthrough for treatment of cancer patients. While the response rates for ICIs or their combination with targeted drugs for HCC are significantly better than response rates for standard systemic treatment, not all patients benefit from these treatments, and those with greatest potential benefit from such treatments are currently unknown. Therefore, there is an urgent need for identifying biomarkers or molecular subtypes that can help to predict response or resistance to ICI therapy or related immunotherapies.

There are several issues that need to be addressed in the near future to fill our knowledge gap in optimizing currently available treatments for HCC patients. First, molecular subtypes that are homogeneous in underlying biology as well as having significant association with clinical outcomes need to be clearly defined. Although many different genomic or molecular subtypes have been discovered from previous studies, they lack consensus, and

---

**Figure 1** Circos plots of the relationship of proteomic subtypes with genomic subtypes in Zhongshan HCC cohort. (A) Comparison of proteomic subtypes with NCIP subtype. (B) Comparison of proteomic subtypes with EpCAM subtype. (C) Comparison of proteomic subtypes with Hoshida subtype. (D) Comparison of proteomic subtypes with TCGA subtype. Circular ideogram of pairwise comparison of two classifications. Proteomic subtypes are displayed in left and matched subtypes from previous study are displayed in right. Ribbons indicate matched samples between two classifications. Graphic is built using Circos (http://circos.ca/).

**Abbreviations:** NCIP, National Cancer Institute Proliferation; HP (also known as (A)), high proliferation; LP (also known as (B)), low proliferation; EpCAM, Epithelial Cell Adhesion Molecule subtype; Hoshida, Hoshida classification; TCGA, The Cancer Genome Atlas.
robustness of subtypes has yet to be properly tested and validated. Integration and co-analysis of discovered subtypes together might be a first step for finding subtypes with good consensus that can be later validated in pooled large data sets from previous studies. Second, biomarkers tightly associated with subtypes or durable response to standard treatments should be validated in prospective cohorts. As HCC is a heterogeneous disease, it is unlikely that a single biomarker will be good enough to identify patients with durable response. Therefore, non-redundant multiple biomarkers such as a mutation profile from cell-free DNA, serum markers, or metabolites need to be analyzed together and validated in prospective cohorts of patients. Finally, finding the best preclinical models that mimic the underlying biology of particular subtypes would be necessary for rational preclinical studies evaluating new treatments in the future.

Funding
This study was supported by National Cancer Institute grant R01-CA237327, P50-CA217674 (HCC SPoRE), P30-CA016672 (Cancer Center Support Grant), the Duncan Cancer Prevention Research Seed Funding Program at MD Anderson Cancer Center, the MD Anderson Sister Institute Network Fund, and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07040523).

Disclosure
The authors report no conflicts of interest in this work.

References


47. Sporn MB, Liby KT, Nrf2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer. 2012;12:564–571. doi:10.1038/nrc3278


95. Okamoto H. PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated. Cancer Res. 2008;68:499–507. doi:10.1158/0008-5472.CAN-07-6013


