Published in final edited form as:

Gastroenterology. 2011 November ; 141(5): 1943–1945. doi:10.1053/j.gastro.2011.09.027.

Hepatocellular carcinoma enters the sequencing era

Augusto Villanueva, M.D.¹, Yujin Hoshida, M.D., Ph.D.^{2,3}, and Josep M. Llovet, M.D.^{1,4,5}

¹HCC Translational Research Laboratory, Barcelona-Clínic Liver Cancer (BCLC) Group, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Liver Unit, Institut d'Investigacions Biomediques Augusto Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain

²Cancer Program, Broad Institute, Cambridge, USA

³Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA

⁴Liver Cancer Program. Division of Liver Diseases, Mount Sinai School of Medicine, New York, USA

⁵Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

Article commented

Totoki Y, Tatsuno K, Yamamoto S, et al. (Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan). High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 2011; 43(5):464–9.

Liver cancer is the fifth most frequently diagnosed cancer worldwide, and the second cause of cancer death in men (**CA Cancer J Clin** 2011;61:69–90). Hepatocellular carcinoma (HCC) is the predominant histological subtype, representing more than 80% of all cases. Most HCCs are still diagnosed at intermediate/advanced stages when prognosis is dismal and potentially curative therapies are unfeasible (**Lancet** 2003;362:1907–17). However, positive results of targeted molecular agents like sorafenib (**New Engl J Med** 2008;359:378–90) have brought new excitement in the field, emphasizing the importance of identifying the molecular alterations that govern HCC development and progression. Despite recent advances related to molecular classification, microRNA profiling, oncogene discovery and sophistication of animal models, the high molecular complexity of this malignancy is complicating this task (**Semin Liver Dis** 2007;27:55–76). Part of this heterogeneity is due to the different etiologies responsible for the underlying liver damage (e.g., viral hepatitis B/C, alcohol abuse) that facilitates HCC onset ('field effect').

The study by Totoki et al (**Nat Genet** 2011;43:464–9) is the first published report applying next-generation sequencing technologies in HCC. The study was performed on a single hepatitis-C related HCC sample, and it also included sequencing of lymphocytes from the same individual to distinguish germ-line from somatic changes. Authors followed 2 parallel approaches using short-insert libraries, one entailed analysis of the entire genome (whole-genome sequencing, WGS) covering >99% of the human reference DNA sequence, while the other was restricted to coding regions (whole-exome sequencing, WES). Comparison of WGS data between tumor and lymphocyte genomes revealed 11,731 somatically acquired changes in the tumor. Genic regions (i.e., introns and exons) were significantly less prone to

harbor these changes when compared to intergenic areas. There was a dominance of T>C/A>G and C>T/G>A transitions. Interestingly, this pattern of substitutions is different from other cancers with other environmental risk factors such as smoking or ultraviolet light.

When focusing on protein-coding regions, the study found 90 somatic substitutions (81 validated with Sanger sequencing), 63 of them representing non-synonymous changes. Additionally, authors identified 670 somatic small insertions and deletions, 7 located in protein-coding regions. Somatic alterations included well-known tumor suppressors in HCC (*TP53* and *AXIN1*), as well as other genes previously described in other malignancies but unknown to HCC (*ADAM22*, *JAK2*, *KHDRBS2*, *NEK8* and *TRRAP*). Gene annotation enrichment analysis revealed overrepresentation of genes encoding phosphoproteins and those with bipartite nuclear localization signals. Larger structural rearrangements occurred in 33 regions (22 validated by Sanger sequencing of breakpoints). Interestingly, nine occurred in the region 11q13, a previously reported hotspot of DNA amplifications (**Cancer Res 2008;68:6779–88**, **Cancer Cell 2011;19:347–58**). Somatic rearrangements also generated four fusion transcripts (*BCORL1-ELF4*, *CTNND1-STX5*, *VCL-ADK*, *CABP2-LOC645332*), but unexpectedly, none was validated in an additional set of 47 primary HCCs analyzed with RT-PCR. Regarding WES analysis, it revealed 47 non-synonymous somatic substitutions, 40 were validated using Sanger technique. Mutated genes included *TSC1*, a negative regulator of *MTORC1* signaling that is frequently mutated in a subgroup of tuberous sclerosis patients. This change went undetected by WGS, and was heterogeneously distributed in tumor alleles, affecting around 13%. Deeper coverage obtained with WES probably enabled the detection of mutations harbored in rare clones and depicted molecular heterogeneity within a tumor nodule. Overall, WES missed 25 non-synonymous somatic substitutions detected by WGS.

Comment

This decade has witnessed a revolution in biomedical research, mainly thanks to the development of technologies that allows massive analysis of molecular data. The new paradigm primes unbiased high-throughput data exploration, extensive validation and integrative genomic analyses. In this sense, second-generation (or next-generation) sequencing has enabled very efficient and comprehensive determination of genomic DNA alterations, including nucleotide substitution, small insertion and deletion, and DNA copy number changes (**Nat Rev Genet 2010;11:685–96**). This technology has also allowed for the first time a systematic identification of chromosomal rearrangements and genomic integration of pathogens. Both features are particularly relevant to HCC since genomic alterations in HCC are enormously complex, and also because DNA viral integrating fragments, especially hepatitis B, are believed to play a role in HCC pathogenesis.

A series of initial studies of second-generation sequencing have assessed single patients in various types of cancer, including acute myeloid leukemia (**Nature 2008;456:66–72**, **N Engl J Med 2009;361:1058–66**), lung cancer (**Nature 2010;463:184–90**), breast cancer (**Nature 2009;461:809–13**, **Nature 2010;464:999–1005**), and melanoma (**Nature 2010;463:191–6**). These studies have revealed comprehensive catalogs of genome-wide DNA structural alterations and demonstrated the characteristic patterns for each cancer type. Furthermore, sequencing of transcriptome (RNA-seq) has revealed fusion oncogenic transcripts in leukemia, prostate cancer, gastric cancer, and melanoma (**Nature 2009;458:97–101**, **Nat Med 2010;16:793–8**). When compared to these studies, Totoki et al. (**Nat Genet 2011;43:464–9**) found more non-synonymous somatic substitutions in HCC than in other malignancies (e.g., acute myeloid leukemia, pancreatic cancer, glioblastoma), even though larger samples sizes are still needed to determine relative mutation rates across human cancer.

Rapid technology development and cost reduction have recently allowed expansion of study size to elucidate recurrent and pathogenic alterations (**Nature 2009;462:1005–10**, **Nature 2011;471:467–72**). For large datasets, national and international initiatives such as The Cancer Genome Atlas (<http://cancergenome.nih.gov/>) or the International Cancer Genome Consortium (<http://www.icgc.org/>) will be instrumental. In addition, targeted sequencing (e.g., WES) will facilitate flexible application of the technology to address more specific biological questions such as clonal evolution of cancer cells (**Nat Biotechnol 2009;27:182–9**) Implementation of deep-sequencing approaches will provide large amounts of information. Data analysis, which requires unprecedented computational resources in biomedical research, has posed a big challenge, and massive effort is undertaken to establish the analysis methodologies and pipeline (**Nat Rev Genet. 2010;11:685–96**).

Previous reports have shown that TP53 and CTNNB1, the codes for β catenin, are the genes more frequently mutated in HCC. Mutation rate is highly variable among studies and, for example, TP53 mutations range between 0–67% (**Semin Liver Dis 2007;27:55–76**). Differences in etiology, stage and tumor heterogeneity may account for these discrepancies. Unlike other solid tumors (**Nat Genet 2007;39:347–51**), HCC lacks a thorough evaluation of mutations in known oncogenes or tumor suppressors using first-generation sequencing technologies. Upcoming next-generation sequencing projects will overcome this limitation. It will be of particular interest to study mutations in targetable molecules, such as kinases. Some of them, as happens with the BCR-ABL fusion transcript in chronic myeloid leukemia, may act as oncogenic addiction loops. A significant number of drugs under development for HCC (**Gastroenterology 2011;140:1410–26**) are able to inhibit kinase activity and block these loops.

The study by Totoki et al (**Nat Genet 2011;43:464–9**) provides several take home messages: 1. This first-in-class report represents the dawn of a new area of sequencing in HCC, 2. Several mutations plague HCC development, in this case the authors confirmed around 80 somatic mutations, 7 small insertions/deletions and 22 rearrangements. This fact is in line with what has been reported in other complex solid tumors, but large series of samples are needed to define the true frequencies. 3. Mutations are not homogeneously distributed among given tumors, meaning that specific clones within large tumors harbor mutations not shared by all neoplastic cells; 4. Identification of driver genes as opposed to bystanders requires additional efforts aimed to integrate all genomic and epigenetic data available of a given tumor; and 5. WES and WGS are complementary gene sequencing approaches. Thus, there are still many issues that need to be worked out: further optimization of WES to rescue uncovered regions, the biological significance of mutation heterogeneity within tumors, the pathogenic role of fusion transcripts, or the impact of substitutions in non-coding regions; but still, this study puts HCC at the starting point of next-generation technologies in biomedical research.

Acknowledgments

Grant Support: Josep M Llovet is supported by grants from the US National Institutes of Diabetes and Digestive and Kidney Diseases (1R01DK076986), the European Commission's Framework Programme 7 (HEPTROMIC; 259744), the Samuel Waxman Cancer Research Foundation, the Spanish National Health Institute (SAF-2010-16055), and the Landon Foundation-American Association for Cancer Research.