Commentary

Whole-genome mutational analysis: cause and effect of noncoding and structural mutations in liver cancer

Suchandrima Ghosh, Soma Banerjee

Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education and Research, West Bengal, India
Correspondence to: Dr. Soma Banerjee, Associate Professor. Centre for Liver Research, School of Digestive and Liver Diseases, Institute of Post Graduate Medical Education and Research, 244, Acharya Jagadish Chandra Bose Road, Kolkata-700020, West Bengal, India. Email: somabanerjee70@gmail.com.
Provenance: This is an invited Commentary commissioned by Editor-in-Chief Yilei Mao (Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China).
Comment on: Fujimoto A, Furuta M, Totoki Y, *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016;48:500-9.

Submitted Nov 08, 2016. Accepted for publication Nov 14, 2016.
doi: 10.21037/hbsn.2017.01.03
View this article at: http://dx.doi.org/10.21037/hbsn.2017.01.03

Hepatocellular carcinoma (HCC) is rapidly becoming the most prevalent cancer worldwide (1). Heterogeneity among etiological factors, clinical presentations, variations in genetic and epigenetic alterations probably pose the major difficulties in identification of early diagnostic marker, selection of effective therapeutics and prevention of its recurrence. Recent multiple deep sequence analyses identified several discrete factors associated with the development of the diseases such as coding mutations in *TP53*, *CTNNB1*, *AXIN1*, *ARID1A*, *ARID2*, *BRD7*; over representation of components of *JAK*, *IDH*, *FLT* and *mTOR* pathways; somatic mutations in noncoding regions and mutations in the *TERT* promoter (2-8). However, integrated approaches to determine whole-genome mutational landscape, characterization of noncoding and coding mutations in large number of liver cancer tissues could be the effective approaches in combination with the data analysed from whole genome sequencing and the RNAseq to investigate the cause and effect of the mutations.

As the diverse mechanisms result in the development of liver cancers depending on either different cell origin causing hepatocellular carcinomas (HCCs), or intrahepatic cholangiocarcinomas (ICCs) or different etiologies like hepatitis viruses (Hepatitis B or C), alcoholic, cryptogenic or combined type; hence mutation landscapes might not always be similar in all the causes. Thus categorizing the accumulation of mutations in particular type might give an insight to the cause of the differential expression of the oncogenes and tumor suppressor genes.

Fujimoto *et al.* has enrolled 89% HCC patients (268/300), 8% (24/300) ICC and 3% (8/300) combined type (HCC and ICC). HCV and HBV monoinfected patients were 53% (159/300) and 27% (82/300) respectively; while only 1% (4/300) was combined HCC (cHCC) and 18% (55/300) nonB-nonC group. Thus mutational signatures only in HCV and/or HBV related HCC and their association with disease have been clearly elucidated in this study (2).

On average large number of somatic substitutions [9,718], indels [271], structural variations (STVs) [40.6] and HBV integrations [2.5] were found in HBV-HCC samples. This huge number of point mutations were correlated well with the increasing age of the patients, diameter of the tumors, smoking status and intratumoral heterogeneity. It is well known fact that cancer-related genes are often involved in fundamental cellular processes such as development, cell cycle, and DNA repair, and these genes are generally in early replicating regions; while late-replicating regions tend to contain tissue-specific or lowly expressed genes. Mutations in the latter regions would be more likely to be tolerated than cancer related genes. Point mutations were more common in late replicating region which is probably not accessible to DNA repair enzyme and has less DHSs (9). Again tandem gene duplications were observed mostly in early replicating region while deletion events were associated with late replication as confirmed by *in vitro* cell culture assays in this study (2).

Comparing nucleotide substitutions in sample groups HCC versus ICC and cHCC/ICC has revealed seven

mutational signatures in HCC, which have been further validated with mutations reported in previous studies and in Current Catalogue of Somatic Mutations in Cancer (COSMIC) data of HCC. Among them W6, which was characterized by T>A mutations, was a novel signature found in this Japanese population. Signature W5, which was not observed in other cancers, was found frequently in Japanese alcoholic HCC patients. Signature W4 was found to be associated with *TP53* mutation and smoking status of the patients. W2 showed connotation with mutations in ARID family members, while W3 and W5 were found with *TERT* promoter mutations.

Assembly of the RNA seq data of HBV-HCC samples revealed that 28% (74/268) of the samples had more than 10 HBV integrations in the human genome and HBV genotype C showed highest rate of integration among Japanese patients which has been reported recently from India too (10). Interestingly, majority of the integrations were observed either in protein coding region causing structural variations or in distal promoter region (>10 Kb upstream region), altering gene expression level. Recurrent integration of HBx of HBV genome in *TERT* and *KMT2B* (MLL4) genes have been reported in several previous studies (4,5) including this study (2) and number of integrations were directly proportional to the double strand DNA break points. Similar to a recent French study, adeno associated virus (AAV) integration event has been found in liver cancer samples of this Japanese cohort. These HBV and AAV integration often caused overexpression of *KMT2B* and cell cycle regulator *CCNE1* (4). In addition to the viral integration and mutations in *TERT* promoter region in HCC, ten other promoters including *TFPI2*, *MED16*, *WDR74*; nine UTRs including *BCL6* and *AFF4* and six long intergenic non-coding RNA (lincRNA) comprising *NEAT1* and *MALAT1* were identified as rich with multiple mutations. Additionally, for the first time this study has revealed the mutations in 4 consecutive regions on chromosome 18, 3 CTCF-binding sites, 3 regions in an ABHD5 intron, and chromosome 17 and 14 noncoding regions, by considering the mutational clusters in non-coding regions, present in more than 3% of the samples and located in DHSs and/or ChiP-seq transcription binding sites. Indels in chromosome 18 were found to be ~1kb downstream of miR-122, though no change in its expression has been found. Using ChIP-qPCR and competitive electrophoretic mobility shift assays cluster of mutations have been identified in CTCF transcription factor binding sites in chromosome 2, 3, 18 and 20 which were located downstream of *APOH* and upstream of *PRKCA* and showed inverse relation with their expressions in liver cancers.

Apart from the mutations in regulatory or intergenic regions, mutations in the coding region could be the possible driver genes in liver cancer. Hence protein altering mutations have been found in multiple genes involved in chromatin remodeling (*ARID2*, *ARID1A*), transcription factors (*NFE2L2*, *HNF4A*), tumor suppressors or like proteins (*TP53*, *RB1*, *AXIN1*, *BAP1*) and oncogene (beta-catenin *CTNNB1*). Similar to the Chinese data (3), highest rate of mutation was observed in *CTNNB1* gene in HCC compared to ICC. Loss-of functions was also observed in a myriad of genes such as *ARID2*, *ARID1A*, *AXIN1*, *TP53*, *BRD7*, *RPS6KA3*, *RB* and *HNF4A*. Interestingly, detailed sequence analysis of break points to see the effect of STVs in coding regions considering 500kb genomic span revealed that breakpoints contained either tumor suppressors (*CDKN2A*, *LRP1B*, *RB1*), DNA repair genes (*MACROD2*) or oncogenes (*CCND1*). Overall analysis showed that recurrent mutations were enriched in regulators of chromatin structure (*SETDB1*, *ASH1L*, *NCOR1*) and cell cycle (*CCNE1*) in HCC. These structural variations also alter gene expression through change in copy numbers (CNAs) and structure disruption as found in several known cancer related genes including *MET*, *APC*, *AMER1*, *PTPN11* etc. Most interestingly, in addition to point mutation and HBV integration in *TERT*, structural variations which cause translocations and inversions in this region were exclusively found in HCC.

Finally, patient's survival and prognosis were correlated with the mutations and STVs in driver genes. *TERT* promoter mutations were associated with smoking while *CTNNB1* and *BRD7* were linked to HCV and HBV respectively. Mutations in *LRP1B* was found in HCV with alcohol. Overall analysis showed that mutations in *TP53*, *CTNNB1*, *CDKN2A*, *MACROD2* and *ARID2* cluster are correlated to disease free survival after adjustment of age, sex, tumor stage and pre-surgery or adjuvant therapy. While mutations in *TP53* and *ARID2* alone tend to show worse prognosis, mutations cluster in *MACROD2* showed better prognosis.

Thus, the high throughput advanced integrated methodologies used in this study to identify driver mutations, structural variations in non-coding and coding region are not only a powerful tool to study disease biology but also provide the opportunities to translate the scientific discoveries into patient care improvement.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Fujimoto A, Totoki Y, Abe T, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012;44:760-4.
3. Kan Z, Zheng H, Liu X, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 2013;23:1422-33.
4. Sung WK, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765-9.
5. Jiang Z, Jhunjhunwala S, Liu J, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res 2012;22:593-601.
6. Schulze K, Imbeaud S, Letouzé E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505-11.
7. Weinhold N, Jacobsen A, Schultz N, et al. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 2014;46:1160-5.
8. Nault JC, Mallet M, Pilati C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat Commun 2013;4:2218.
9. Woo YH, Li WH. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. Nat Commun 2012;3:1004.
10. Datta S, Roychoudhury S, Ghosh A, et al. Distinct distribution pattern of hepatitis B virus genotype C and D in liver tissue and serum of dual genotype infected liver cirrhosis and hepatocellular carcinoma patients. PLoS One 2014;9:e102573.

Cite this article as: Ghosh S, Banerjee S. Whole-genome mutational analysis: cause and effect of noncoding and structural mutations in liver cancer. HepatoBiliary Surg Nutr 2017;6(1):57-59. doi: 10.21037/hbsn.2017.01.03

