Mini-review

Comprehensive genome sequencing of the liver cancer genome

Hidewaki Nakagawa [a],*, Tatsuhiro Shibata [b],*

[a] Center for Genomic Medicine, RIKEN, Yokohama, Japan
[b] Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan

ARTICLE INFO

ABSTRACT

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide. Recently, comprehensive whole genome and exome sequencing analyses for HCC revealed new cancer-associated genes and a variety of genomic alterations. In particular, frequent genetic alterations of the chromatin remodeling genes were observed, suggesting a new potential therapeutic target for HCC. Sequencing analysis has further identified the molecular complexities of multicentric lesions and intratumoral heterogeneity. Detailed analyses of the somatic substitution pattern of the cancer genome and the HBV virus genome integration sites by using whole-genome sequencing will elucidate the molecular basis and diverse etiological factors involved in liver cancer development.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide [1]. Prevalence of this cancer is different among ethnic groups; HCC is more frequent in East Asian and African populations [1,2]. The number of cases has also been rapidly increasing in Western countries, and globally, more than 700,000 new patients are diagnosed annually [1]. Despite the recent progress in diagnostic (computed tomography and ultrasound imaging) and therapeutic (transcatheter arterial chemo-embolization, radio frequency ablation, and transplantation) modalities, the 5-year survival rate is still low [3]. Characteristically, HCC shows a high rate of recurrence or multicentric occurrence against strong carcinogenic backgrounds such as chronic hepatitis and liver cirrhosis. Patients occasionally develop recurrent or secondary tumors with reduced liver function and cannot tolerate surgery. Therefore, new treatments, including molecular therapies for inoperable cases, and preventive strategies have been intensively sought.

Multiple etiological factors for HCC have been identified [2]. The most important risk factor is infection with hepatitis viruses, mainly HBV and HCV. Alcohol-induced liver damage also ranks high, especially in Western countries. Recently, metabolic diseases such as obesity and diabetes mellitus have become well recognized as risk factors for HCC [3]. Virus infection or metabolic stress results in liver damage including fatty change, hepatitis, and cirrhosis, which set premalignant conditions for HCC. Chronic inflammation, virus infection, and liver regeneration in cirrhosis have been reported to induce genetic and epigenetic damage to the host genome [4]. Most HCCs gradually develop from these premalignant stages by the accumulation of these alterations. Accordingly, highly damaged livers are extremely susceptible to multiple tumors.

Recent innovative developments in next-generation sequencing (NGS) technologies, especially massively parallel sequencing and paired-end sequencing, have enabled whole-genome or exome sequencing of the cancer genome within a reasonable time frame (~10 days) and cost (less than $10,000 for a tumor/normal pair). This approach can rapidly and exhaustively identify potential key genetic events, including potential molecular therapeutic targets, in cancer [5]. Along with the application of NGS technologies in cancer research, international or domestic networks of cancer genome research have been initiated to effectively promote cancer genome sequencing and share high-quality data among scientists. Among these initiatives are the International Cancer Genome Consortium (ICGC) [6] and The Cancer Genome Atlas (TCGA) in the US [7]. Currently, three national projects focusing on HCC have been launched by the ICGC/TCGA: the virus-associated HCC project in Japan, the alcohol-associated HCC project in France, and another HCC project in the US (as part of TCGA) (http://www.icgc.org/). Each HCC genome project will analyze 500 HCC cases by using high-resolution genome sequencing and deposit the data on the ICGC or TCGA database, which can be accessed by the public. These projects will collect samples with distinct epidemiological (virus

* Corresponding authors. Addresses: 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan (H. Nakagawa), 5-1-1 Tsukiji, Chuoh-ku, Tokyo 104-0045, Japan (T. Shibata).
E-mail addresses: hidewaki@ims.u-tokyo.ac.jp (H. Nakagawa), tashibat@ncc.go.jp (T. Shibata).

and alcohol) and ethnic (Asian and Caucasian) backgrounds; thus, future integration of these huge datasets will provide a much clearer understanding of the association between molecular genetics and epidemiological, environmental, and ethnic factors in HCC.

Here we summarize recent genome-wide sequencing (whole genome and whole exome) studies of the liver cancer genomes which cover mutational signatures with distinctive etiological backgrounds, identification of new cancer genes, multicentric cancer genomes and HBV genome integration. Based on these sequencing data, we also discuss about intratumoral heterogeneity and evolution of liver cancer genomes

2. Whole genome and exome sequencing analyses for HCC

Several comprehensive genome sequencing analyses, including exome and whole genome sequencing, were conducted for HCC. The results of these analyses are summarized in Table 1. Whole exome sequencing (WES) can efficiently detect mutations in protein-coding exons, which are much more easily interpretable than mutations or variants in non-coding regions; this is now the most common platform for ICGC/TCGA projects. This approach involves target-enrichment of whole protein-coding exons of the human genome (30–40 Mb, approximately 1% of the whole human genome) using in-solution RNA or oligonucleotide DNA probe hybridization technologies [8]. The accuracy of sequencing analysis by using NGS is basically dependent on the depth or coverage of target regions, and WES following target-enrichment can usually cover more than ×80 or ×100 depth at a much lower cost. This enables more accurate detection of single nucleotide changes, which is an advantage of WES. On the other hand, whole genome sequencing (WGS) can cover almost all the human genome sequences (approximately 3 Gb) and detect variants in non-coding regions, genomic rearrangements, copy number alterations, and virus genome integrations in addition to single nucleotide changes. This strategy is more comprehensive and suitable for analyzing the cancer genome, because cancer has been recognized as a "disease of the genome," where a variety of complicated genomic alterations plays a central role in carcinogenesis. Hence, analysis of structural alterations is required to fully understand its genomic signature. Following the evolutionary progress of NGS technology and informatics, the cost and labor of WGS is rapidly dropping down and it is becoming a central technology for human and cancer genome analysis.

Initial exome sequencing analysis for HCC, which did not cover the whole exome region but analyzed exons of ~18,000 genes, identified recurrent mutations of ARID2, which was further validated in more than 120 additional cases [9]. The whole genome of one HCV-related HCC was unveiled for the first time by using a massive parallel sequencing approach and demonstrated its genomic features in high resolution and intratumoral heterogeneity [10]. Whole genomes of multiple regions of one HBV-related HCC and its intrahepatic metastases were sequenced by using a massive parallel sequencing approach with relatively low depth [11], and these genomes were analyzed in terms of cell-population genetic aspects. Jiang et al. analyzed whole genomes of four HBV-related HCC genomes by using the low-cost Complete Genomics technology [12] and detected a number of HBV genome integrations. Recently, WES analysis combined with copy-number analysis using SNP arrays for mainly alcohol-related HCCs [13] showed that several pathways were involved in liver carcinogenesis. WGS analyses for 27 HCCs [14] and 88 HBV-related HCCs [15] detected candidates of driver mutations and/or HBV integrations comprehensively. Another WES analysis of ten advanced (with portal vein invasion) or metastatic HCCs with large-scale functional analysis by siRNA identified several driver genes in liver carcinogenesis and liver cancer progression [16]. WGS usually requires more than ×30 depth, but in the near future "deep" WGS or WGS with ultra-long sequencing reads will be able to detect more mutations, rearrangements, and virus integrations even in minor cell populations more accurately. However, the biological significance of somatic alterations that occur in minor tumor populations is controversial.

3. Mutation signatures of HCC

3.1. Substitution signature and its association with etiological factors

Somatic mutations are caused by environmental factors (mutagens) and deficiencies in the DNA repair systems. Previous studies revealed specific somatic substitution patterns in cancer genes [17]. For example, C > A/G > T transversions are the most frequent substitutions in the TP53 gene in liver cancer. However, these analyses may be biased because positive or negative selection for cancer-related genes could exist. In contrast, WGS can identify thousands of neutral (unselected) mutations in the cancer genome, whose profile is more reliable to statistically evaluate the mutation signature.

Totoki et al. reported a significant increase in C > T/G > A and T > C/A > G somatic transitions in the HCC genome [10]. Using more than 10,000 somatic mutations in this case, the authors attempted to elucidate the characteristic sequence contexts surrounding each somatic substitution. They found that the C > T/G > A transitions significantly occurred at CpG sites, with T > C/A > G at ApT sites. C > T transitions, caused by deamination of cytosine, are known to frequently occur in nature. However, the molecular mechanism involved in T > C transitions at ApT have not been elucidated. Jiang et al. performed WGS of HBV-infected HCCs and reported a similar substitution pattern [12]. C > T/G > A and T > C/A > G transitions and C > A/G > T transversions are reported to be dominant in both HCV- and HBV-positive HCC cases [14]. Guichard et al. found that C > A/G > T transversions were over-represented in

Table 1
Summary of comprehensive genome sequencing of liver cancers.

Analysis platform	Sample no.	SQ coverage	Main etiology of HCC	Specific focus	Ref.
Exome (~18,000 genes)	10	×98	HBV + HCV+NBNC[a]		Li et al. [9]
WGS + Exome	1	×36	HCV		Totoki et al. [10]
WGS	1	×20	HBV	Multiple metastasis	Tao et al. [11]
WGS + RNA-Seq	4	×80	HBV	HBV integrations	Jiang et al. [12]
Exome + SNP array	24	×73	Alcohol-related		Guichard et al. [13]
WGS	27	×40	HBV + HCV + NBNC	Multicentric tumors	Fujimoto et al. [14]
WGS	88	×38	HBV	HBV integrations	Sung et al. [15]
Exome	10	×28, ×88	HBV	HBVPPVTs[b]	Huang et al. [16]

[a] Non-HBV and non-HCV infection.
[b] Portal vein tumor thromboses.