Clustering of Minimal Deleted Regions Reveals Distinct Genetic Pathways of Human Hepatocellular Carcinoma

Yuh-Shan Jou,1 Chih-Shia Lee,1 Ya-Hui Chang,1 Chin-Fu Hsiao,2 Chian-Feng Chen,1 Chuan-Chuan Chao,1 Lawrence S.H. Wu,1 Shiou-Hwei Yeh,1 Ding-Shinn Chen,3 and Pei-Jer Chen3

1Division of Molecular and Genomic Medicines and 2Division of Biostatistics and Bioinformatics, National Health Research Institute; and 3Hepatitis Research Center, Department of Internal Medicine, National Taiwan University Hospital, Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

ABSTRACT

Systematic scan and statistical analysis of loss of heterozygosity (LOH) has been widely used to define chromosomal aberrations in various cancers for cloning of tumor suppressor genes and for development of prognostic markers. However, the establishment of novel strategies is needed, so that the nonrandom but heterogeneous chromosomal aberration data could provide significant insights into our understanding of molecular pathogenesis of cancers. After comprehensive allelotyping of recurrent allelic losses with 441 highly informative microsatellite markers and overlapping LOH regions on human hepatocellular carcinoma (HCC) chromosomes, 33 minimal deleted regions (MDRs) were revealed. Five and 13 of the 33 MDRs have physical intervals in less than 5 and 10 Mb, respectively, with the smallest MDR9p1 of 2.2 Mb located at 9p21.3-p21.2. Statistical and Kaplan-Meier survival analysis revealed a significant association between the loss of MDR15q1 (15q21.1-q22.2) and the HCC patient survival (adjusted \(P = 0.033\)). After cluster analysis of 33 MDRs that represented LOH profiles of each HCC tissue based on clinicopathological features and p53 mutations, two major genetic pathways, low-stage and advanced-stage HCC, were uncovered based on high concordance of MDR clusters. We propose that the definition of genome-wide MDRs on the cancer genome not only narrows down the location of existing tumor suppressor genes to facilitate positional candidate cloning and develop potential prognostic markers after statistical association of MDRs with clinicopathological features but also dissect genetic interactions and pathways of chromosomal aberrations in tumorigenesis.

INTRODUCTION

Human hepatocellular carcinoma (HCC) is one of the most prevalent malignancies worldwide, especially in parts of Asia and Africa, with an estimated 0.5 million new cases and around 1 million deaths annually (1). The average incidence of HCC in Asia is about 10–30 versus 1–3 per 100,000 person-year in western countries (2). In an endemic area like Taiwan, the incidence of HCC rises to 200–820 cases per 100,000 person-year with patients suffering chronic hepatitis and around 4,500 persons dying of HCC every year (3–5). Recent epidemiological studies have also projected an alarming increase of HCC in Japan and western countries for the next decade (6, 7). The etiology of HCC is a long-term process of chronic hepatitis and cirrhosis resulting from prolonged exposure of risk factors such as persistent hepatitis virus infection, dietary exposure of mycotoxins, and alcohol abuse (2, 8). Although early HCC is curable by surgical resection, the asymptomatic feature of HCC progression results in poor prognosis and a low 5-year survival rate (12–15%). Therefore, early diagnosis developed from molecular genetic studies of tumorigenesis should improve the clinical management and treatment of HCC.

Recent studies have identified numerous genetic and/or epigenetic changes, even though these likely represent only a fraction of multifactorial and multistage features of HCC progression (4, 8). Among them, mutations of p53 and \(\beta\)-catenin genes (9–11) and inactivation of Rhl and INK4a-ARF genes by aberrant methylation or deletion (12–14) occur most frequently and are likely to contribute to major genetic mechanisms of hepatocarcinogenesis. Other genetic and epigenetic changes include mutations of M6P/IGF2R (15), BRCA2 (16), Smad2/4 (17), LPTS (18), HCCSI (19), DLC-1 (20), PTEN (21), Axin1 (22), and TCF1 (23) genes; gene amplification of c-myc (24), gankyrin (25), and cyclin D1 (26); and gene silencing by hypermethylation of SOCS-1 (27), GSTP1 (28), and 14-3-3 sigma (29) genes. Over the last few years, techniques that allow systematic analysis of chromosomal aberrations at a genome-wide level were applied to HCC (30, 31). Notably, results obtained from loss-of-heterozygosity (LOH) analysis using microsatellite markers are well correlated with those from comparative genomic hybridization analysis and point out common chromosome losses on 1p, 4q, 6q, 8p, 13q, 16q, and 17p.

Because nonrandom but heterogeneous chromosomal aberrations are attributed to mutation accumulation and clonal selection in tumor progression, statistical analysis of chromosomal aberrations associated with various clinical features such as viral infection, cirrhosis, tumor stage and size, tumor cell differentiation, and patient survival were applied to develop prognostic markers for early diagnosis and therapeutic treatment (32). For example, McGlynn et al. (33) applied a phylogenetic approach on genome-wide LOH data to conclude the heterogeneity of HCC with nonrandom allelic loss and cluster four definable branches characterized by distinctive rates of loss in hepatocarcinogenesis. Laurent-Puig et al. (33) used mutations in specific genes (p53, Axin 1, and \(\beta\)-catenin), the stability of chromosome indexed by fraction allelic loss, and allelic losses on a specific chromosome arm to classify HCC into two major genetic pathways in relation with HCC clinical features. To further improve our understanding of molecular pathogenesis of HCC, cloning of unidentified tumor suppressor genes and their interactions based on recurrent chromosomal aberrations in HCC are critical to dissect heterogeneous HCC pathways (31, 32). However, cloning of candidate cancer genes was hampered by the difficulty of further refining these regions due to the low resolution of chromosome aberrations detected by comparative genomic hybridization and the large LOH deletions detected by microsatellite markers. We, therefore, carried out genome-wide minimal deleted region (MDR) analysis on the HCC genome and discovered 33 MDRs that have pinpointed many small intervals to aid in future cloning of putative tumor suppressor genes. Furthermore, cluster analysis of these MDRs not only revealed interesting interactions but also elucidated potential genetic pathways that they participate in the hepatocarcinogenesis. Our systematic strategy of revealing genome-wide MDRs and their potential interactions involved in distinct tumorigenic pathways of HCC should be applicable to genomic studies of other cancers.
MATERIALS AND METHODS

HCC Tissue Samples and p53 Mutation Screening. Primary HCCs and their corresponding nonneoplastic liver tissues were obtained from 48 patients receiving surgical section in National Taiwan University Hospital. The HCC tumor tissues were carefully separated for histopathological confirmation of tumor margin and tumor staging. Both tumor and nontumorous tissues were snap-frozen and stored in liquid nitrogen, and the Institutional Review Board of National Taiwan University Hospital approved the use of these archived tissues. Genomic DNAs were extracted and purified with phenol-chloroform extraction followed by ethanol precipitation. The infection of hepatitis viruses (hepatitis C virus and hepatitis B virus) and serum α-fetoprotein (AFP) level were determined by immunomasysys or radioimmunosays (Abott Laborato-
ries). The mutations on coding regions of the p53 gene (11 exons) were determined by PCR amplification and DNA sequencing.

Allelotyping by Microsatellite Markers. A total of 484 polymorphic microsatellite markers were used for genome-wide allelotyping on 48 primary HCCs and their matched nontumor liver tissues. The microsatellite markers were purchased from PE Applied Biosystems (Foster City, CA) and from the set of Multi-Colored Fluorescent Human MapPairs Markers (version 8) of Research Genetics (Huntsville, Alabama). The protocols of PCR reaction, markers pooling, and gel electrophoresis were based on the protocols provided by the manufacturers with slight modifications.

 Determination of LOH and MDR. For increasing reliability of allelotyping data and statistical analysis, each gel image was independently read by two persons for lane tracking and allele sizing of markers. In addition, only when a microsatellite marker was able to detect informative alleles in more than 50% (>15 cases) of the HCC samples was it considered to be a successful allelotyping. For each marker, the allelic loss ratio was calculated by the formula of (T1/T2)(N1/N2), where T1 and T2 are the values of two peaks derived from tumor and N1 and N2 are those from normal tissue. In this study, we used stringent criteria to define LOH loci, to determine the maximal contiguous LOH loci in the chromosomes of each patient, and to assign MDR based on overlapping contiguous LOH loci in multiple HCC tissues. At first, the peak values of both allele areas and allele heights were used to calculate allelic loss ratio (Fig. 1A). When both sets of ratios were greater than 2-fold, the marker region was considered to be a stringent LOH locus. For determination of contiguous LOH loci in a HCC chromosome, the conventional definition of LOH by determining the difference of allelic ratio in 1.5-fold was applied. To examine the distribution of LOH frequency, a biphasic distribution was ob-
served by plotting number of markers as a function of LOH percentage in allelotyping data (Fig. 1B). The genome alterations detected by markers located in the low LOH frequency are considered to be random insults, whereas those in the high LOH frequency are considered to be nonrandom losses (35), which are significant for the hepatocarcinogenesis. All markers determined as nonrandom losses were chosen to extend the allelic loss region on HCC chromosomes. The noninformative alleles of the internal and surrounding markers in the regions of nonrandom losses on a HCC chromosome were considered to be a continuation of LOH region. It is generally believed that noninformative markers surrounded by LOHs on a given cancer chromo-
some are statistically more likely to be LOHs and, conversely, that those surrounded by retention markers are more likely to be retention (36, 37). Finally, the alignment of continuous LOH regions from multiple HCC tissues allowed us to overlap and minimize the deleted region on cancer chromo-
somes.

Statistical Methods. To study the association of MDRs with clinicopath-
ological features of HCC tissues, we established a cross-tabulation showing the numbers of cases in each classification with or without MDR. Because the HCC samples used in our study contained features of only 21% (10/48) hepatitis B virus negative, 15% (7/48) hepatitis C virus positive, and 15% (7/48) female patients, we excluded features of gender and hepatitis B virus and hepatitis C virus infections for association studies (Supplement 1). We used the two-tailed Fisher’s exact test to examine the association between MDR and HCC clinicopathological features. Because we performed many hypothesis tests on the same data set, we would increase the probability of declaring false significances. We used the permutation adjustments to adjust the P values from multiple hypothesis tests. The permutation adjustments resample the data without replacement to approximate the distribution of the minimum P value of all tests. This distribution is then used to adjust the individual raw P values. Unlike the other methods, the adjusted P values incorporate all correlations and distributional characteristics (38, 39). The SAS PROC MULTTEST was performed to conduct the permutation adjustments. An adjusted P value <0.05 was considered to be statistically significant. The NCSS 2001 program of NCSS statistical software (Kaysville, UT) was applied for analysis of Kaplan-Meier survival curves.

RESULTS

LOH Allelotyping and MDRs Identification. Allelotyping analysis was conducted using 441 (91% successful rate) highly informative microsatellite markers with an average inter-marker distance of 7.53 cM and only one marker displaying an interval greater than 20 cM (21.36 cM). Based on the stringent LOH definition (Fig. 1A), the average genome-wide LOH frequency is 34.01%. LOH analysis by chromosome arms demonstrated that 13 arms, 17p (67.12%), 4q (57.84%), 16q (54.67%), 13q (54.51%), 8p (49.93%), 10q (45.10%), 1q (43.06%), 9p (42.95%), 14q (42.17%), 18q (41.63%), 22q (40.22%), 16p (35.16%), and 15q (35.15%), displayed a LOH frequency above average. When the numbers of markers were plotted with the frequencies of LOH, we observed a biphasic distribution, in which two peaks were separated at a frequency of ∼50% (Fig. 1B). According to previous LOH studies on other tumors (35), markers appearing in the peak of low LOH frequencies detect mostly random changes reflecting the genetic instability of tumors, whereas markers in another peak are likely to show LOHs associated with cancer-specific phenotypes. Therefore, only markers with a LOH frequency >50% were used to be annotated onto chromosomes of individual tumor tissues based on the comprehensive human genetic map published by the Center for Medical Genetics at Marshfield (41). After aligning LOH data of all cancer tissues on each chromosome, 33 MDRs were defined, and Fig. 1C shows the annotation of these MDR data along each chromosome (Supplement 2). The average genetic and physical intervals of 33 MDRs are 18.08 cM and 12.82 Mb, respectively (Table 1). Among 33 MDRs, 23 were reported by conventional LOH analysis on HCC previously and are located on 1p (MDR1p1), 1q (MDRs 1q1 and 1q2), 4q (MDRs 4q1 and 4q2), 6q (MDRs 6q1) 7p (MDR7p1), 8p (MDRs 8p1 to 8p3), 9p (MDR9p1), 10q (MDR10q1), 13q (MDRs 13q1 to 13q5), 14q (MDRs 14q2 and 14q3), 16q (MDRs 16q1 and 16q2), and 17p (MDRs 17p1 and 17p2). In addition, we identified 10 novel MDRs including 2q1, 5q1, 10q1, 12p1, 14q1, 15q1, 17q1, 18q1, 18q2, and 21q1. Using the annotation
of Ensembl database to estimate the physical distances of MDRs, we found that five MDRs are within 5 Mb and 15 are within 10 Mb. Many newly discovered candidate tumor suppressor genes (TSGs) identified in various human tumors were located in these MDRs.

**Loss of MDR15q1 Associated with HCC Patient Survival.** We next examined correlation of MDRs with clinicopathological parameters for developing potential HCC prognostic markers. To avoid false association from multiple hypothesis tests, permutation adjustments were conducted to adjust the $P$ values (38, 39). Except for a marginal association of MDR1p1 (1p36.23–1p36.21) loss with serum AFP concentration ($P = 0.0560$) in HCC patients, we observed that a loss of MDR15q1 (15q21.1–15q22.2) correlated significantly with poor survival ($P = 0.0330$) of HCC patients. Kaplan-Meier survival analysis indicated that the loss of MDR15q1 is significantly associated with 2-year poor survival ($P = 0.0024$), and all three markers within MDR15q1 (6.71 cM) are associated with the tendency of poor survival, especially for allelic loss of $D15S126$ ($P = 0.0054$; Fig. 2).

**Cluster Analysis Identifies Interactions of MDRs and Their Contributions to HCC Pathways.** In addition to identifying the association of particular MDRs with the clinicopathological features, our MDR data provide an opportunity to determine whether MDR in one chromosomal region is linked to MDRs in other chromosomal regions. We approached this by cluster analysis of MDR profiles of the 48 HCC tissues. Cluster analysis has been commonly used in microarray studies to identify functionally related genes that display similar expression patterns (40). Except for the neighboring MDR17p1 and MDR17p2 with a concordance score greater than 0.75, there is no genetic interaction from different chromosomal regions among 31 of 33 MDRs by clustering MDR profiles of 48 HCC tissues. To investigate whether interactions of MDRs could exist in more defined subgroups of HCC, we performed cluster analysis on subclassified HCC tissues according to their clinicopathological features. Fig. 3 shows a representative TreeView presentation of the clustering results of stage II and stage III/IV tissues, based on the AJCC American Joint Committee on Cancer (AJCC) cancer stage grouping of Tumor-Node-Metastasis (TNM) classification. In each case, we observed a number of MDR clusters in which MDRs in distinct chromosomal regions are lost or retained in a highly coordinated fashion. For instance, clusters MDR16q2-MDR13q5 ($C = 0.79$) and MDR13q4-MDR13q2 ($C = 0.79$) were found in stage II tissues, and clusters MDR4q2-MDR1p1 ($C = 0.79$), MDR16q2-MDR8p1 ($C = 0.89$), MDR17p2-MDR17p1-MDR18q2 ($C = 0.76$), MDR16q1-MDR13q2 ($C = 0.79$), and MDR8p3-MDR8p2 ($C = 0.79$) were found in stage III/IV tissues. Additional cluster analysis on various HCC subgroups classified by other clinicopathological features, e.g., invasion, metastasis, $p53$ mutation status, recurrence, serum AFP level, and $T_3/T_4$ classification, revealed the existence of highly concordant MDR clusters in almost every subgroup of HCC, suggesting that genetic interactions between these regions occur during the development of specific feature of HCC (Supplement 3).

By listing the MDR cluster profiles of each HCC subgroup, we
revealed that, based on the similarity of these profiles, they can be categorized into two groups, which are remarkably correlated with the low- and advanced-stage HCC, respectively (Fig. 4). Several features distinguish the profiles of these two groups. First, advanced-stage HCC exhibits many more MDR clusters than low-stage HCC, reflecting the accumulations of genetic changes during the development of advanced tumors. Second, distinct MDR clusters were observed in the two groups. In the low-stage HCC, two MDR clusters, MDR16q2-MDR16q1. On the contrary, in the advanced-stage HCC, the MDR clustering patterns were more complex and diverse. Nevertheless, frequent MDR clusters were still seen, such as the cluster in advanced stage suggests that genetic interactions between chromosome deletions in these regions participate in the low and advanced stage of HCC tumorigenesis, respectively. Notably, both p53 mutation and high AFP level were categorized into advanced-stage HCC according to their MDR clustering patterns. In conclusion, our study of MDR clustering has identified two pathways, in which distinct sets of genetic interactions are associated with the development of low- and advanced-stage HCC.

DISCUSSION

Toward identifying genetic interactions of chromosomal aberrations and dissecting genetic pathways in tumorigenesis, we applied several novel strategies to analyze the HCC genome-wide allelotyping data. First, we distinguished LOH markers that detect mostly tumorigenic allelic losses from those detecting mostly random deletions based on the frequencies of LOH revealed by the marker, thus decreasing the noise generated by these random changes caused by genetic instability of tumors. Second, instead of analyzing LOH data per se, we aligned continuous LOH regions from multiple HCC tissues and defined MDRs. Compared with conventional LOH analysis, this MDR approach allows a more accurate determination of chromosomal losses and evaluation of the contributions of these losses in HCC progression. To our knowledge, our study provides the first genome-wide MDR data in cancer. Furthermore, a large fraction of MDRs identified (20/33; >60%) carries a physical distance of <10 Mb, which should facilitate positional candidate cloning of TSGs. Third, we applied cluster analysis based on MDR profile of each HCC tissue to examine potential interactions of MDRs involved in various clinicopathological features of HCC. Notably, this analysis not only revealed many interactions of MDRs, but also dissected distinct genetic pathways in the progression of HCC.

HCC is rapidly fatal after diagnosis; the survival rate over a 5-year interval is only an average of 15% and ranges from 22 to 73% without and with surgery, respectively (42). Development of a DNA marker for HCC prognosis will be important for HCC patient management. Our statistical analysis revealed a significant association between the loss of MDR15q1 (15q21.1–15q22.2) and HCC patient survival. Recent studies provide increasing evidence that high LOH frequency of the 15q21.1–15q22.2 region was detected in several types of human tumor, including carcinomas of the colorectum, breast, lung, cervix, and bladder (35, 43–46). Interestingly, the frequent LOH observed on chromosome 15q21 in carcinomas of cervical and breast was statistically correlated with advanced stage of tumors (44, 45). We speculated that the loss of MDR15q1 might inactivate a putative tumor suppressor gene involved in malignant tumor progression and provide a novel prognostic marker for various cancers.
The current theory of multistep tumorigenesis indicates that progression into cancer cells requires six essential alterations in cell physiology that collectively perturb regulatory circuits of normal cell proliferation and homeostasis, leading to malignant growth (47). Because linkage analysis reveals no obvious familiar predisposition of HCC, identification of critical genes located in recurrent genomic aberrant regions has been the rate-limiting step for understanding the molecular mechanism of multistage HCC. However, this task is hampered by the highly heterogeneous tumorigenesis process and the large number of genetic alterations accumulating in tumor cells as a result of genome instability. To better understand the mechanism of hepatocarcinogenesis, various models of pathways of HCC have been proposed recently for dissecting tumorigenesis of HCC based on their differences in the clinical characteristics, their aberrations of chromosome arms, and their mutations of known cancer genes (p53, Axin 1, and β-catenin; Refs. 8, 30–33, 48, and 49).

The Kaplan-Meier survival plots of MDR15q1. The survival curves represent the survival fraction against survived months of HCC patients. The red and green lines represent loss (LOSS: MDR or LOH) and retention (RET: without MDR or RET) of the region, respectively. A, the survival plot of MDR15q1 and its cytogenetic location. B, the survival plots of microsatellite markers defined MDR15q1.

Fig. 2. The Kaplan-Meier survival plots of MDR15q1. The survival curves represent the survival fraction against survived months of HCC patients. The red and green lines represent loss (LOSS: MDR or LOH) and retention (RET: without MDR or RET) of the region, respectively. A, the survival plot of MDR15q1 and its cytogenetic location. B, the survival plots of microsatellite markers defined MDR15q1.

The American Joint Committee on Cancer (AJCC) cancer stage grouping of Tumor-Node-Metastasis (TNM) classification and displayed by TreeView program. The cluster analysis is based on the profiles of 33 MDRs either with (red) or without (green) MDR in the subgroup of HCC patients (top row). The MDR subclusters with concordance score over C = 0.75 of stage II (left cluster) and stage III & IV (right cluster) suggested potential interactions of MDRs in HCC progression.

Fig. 3. A representative example of cluster analyses of MDR profiles in HCC tissues based on American Joint Committee on Cancer (AJCC) cancer stage grouping of Tumor-Node-Metastasis (TNM) classification and displayed by TreeView program. The cluster analysis is based on the profiles of 33 MDRs either with (red) or without (green) MDR in the subgroup of HCC patients (top row). The MDR subclusters with concordance score over C = 0.75 of stage II (left cluster) and stage III & IV (right cluster) suggested potential interactions of MDRs in HCC progression.
In our studies, we took cancer genomic approaches to pinpoint 33 recurrent deletion loci in small intervals for positional candidate cloning of TSGs in HCC. Cluster analysis of the 33 MDRs was used to reveal the interactions of these MDRs involved in the highly heterogeneous pathways of hepatocarcinogenesis. A preliminary cluster analysis on all HCC samples did not reveal any concordance except for two neighboring MDRs on chromosome 17p. Nevertheless, subsequent cluster analysis on subgroups of HCC samples classified based on their clinicopathological features has successfully identified many associated MDRs. The existence of these associated MDRs might represent different TSGs, the synergistic loss of which would provide a growth advantage during HCC development. In addition, our cluster analysis also reveals the existence of common MDR clusters in most subgroups of low-stage HCC and a distinct MDR cluster in advanced-stage subgroups. The presence of these “signature” MDR clusters not only suggests that the synergistic interactions between putative TSGs in these loci would be important in the development of low- and advanced-stage HCC but also facilitates the assignment of a given subgroup of HCC into low or advanced stage. Indeed, HCC subgroups with the p53 mutation and an aberrant serum level of AFP were both classified in the advanced stage. On the other hand, the low-stage HCCs show similar concordance of clinicopathological features with normal level of serum AFP and without mutation, the low-stage HCCs show similar concordance of clinicopathological features with normal level of serum AFP and without mutation.


Clustering of Minimal Deleted Regions Reveals Distinct Genetic Pathways of Human Hepatocellular Carcinoma

Yuh-Shan Jou, Chih-Shia Lee, Ya-Hui Chang, et al.


Updated version Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/9/3030

Cited articles This article cites 48 articles, 8 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/9/3030.full.html#ref-list-1

Citing articles This article has been cited by 9 HighWire-hosted articles. Access the articles at:
/content/64/9/3030.full.html#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.