Published online 10 November 2006 *Nucleic Acids Research, 2007, Vol. 35, Database issue* ***D727–D731***
doi:10.1093/nar/gkl845

OncoDB.HCC: an integrated oncogenomic database of hepatocellular carcinoma revealed aberrant cancer target genes and loci

Wen-Hui Su[1,5], Chuan-Chuan Chao[5], Shiou-Hwei Yeh[2], Ding-Shinn Chen[3,4], Pei-Jer Chen[3,4] and Yuh-Shan Jou[1,5,*]

[1]Graduate Institute of Life Sciences, National Defense Medical Center, National Defense University, Taipei 114, Taiwan, [2]Department of Microbiology and [3]Hepatitis Research Center, Department of Internal Medicine, National Taiwan University Hospital, [4]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan and [5]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Received July 18, 2006; Revised and Accepted October 9, 2006

ABSTRACT

The OncoDB.HCC (http://oncodb.hcc.ibms.sinica.edu.tw) is based on physical maps of rodent and human genomes containing quantitative trait loci of rodent HCC models and various human HCC somatic aberrations including chromosomal data from loss of heterozygosity and comparative genome hybridization analyses, altered expression of genes from microarray and proteomic studies, and finally experimental data of published HCC genes. Comprehensive integration of HCC genomic aberration data avoids potential pitfalls of data inconsistency from single genomic approach and provides lines of evidence to reveal somatic aberrations from levels of DNA, RNA to protein. Twenty-nine of 30 (96.7%) novel HCC genes with significant altered expressions in compared between tumor and adjacent normal tissues were validated by RT–PCR in 45 pairs of HCC tissues and by matching expression profiles in 57 HCC patients of re-analyzed Stanford HCC microarray data. Comparative mapping of HCC loci in between human aberrant chromosomal regions and QTLs of rodent HCC models revealed 12 syntenic HCC regions with 2 loci effectively narrowed down to 2 Mb. Together, OncoDB.HCC graphically presents comprehensive HCC data integration, reveals important HCC genes and loci for positional cloning and functional studies, and discloses potential molecular targets for improving HCC diagnosis and therapy.

INTRODUCTION

Cancer is a heterogeneous genetic disease of somatic cells arising from accumulated genetic changes on cancer genome resulted in alterations of gene expression, unregulated cell growth and triggering formation of malignant neoplasm. Systematic genomic approaches have been applied to dissect tumorigenic pathways for diagnostic and prognostic applications and to search for potential cancer genes as therapeutic targets. These technologies revealed a global view of cancer genomic aberrations including loss of heterozygosity (LOH) and comparative genomic hybridization (CGH) analyses to identify chromosome aberrations as well as microarray and proteomic analyses to profile the alteration of cancer gene expression. It has been proposed that there are four to seven somatic aberrations occurred at the rate-limiting steps during epithelial tumor progression including six categories of essential alterations in cell physiology that collectively perturb regulatory circuits of normal cell proliferation and homeostasis leading to malignant growth (1). Since multiple signaling pathways might be disrupted at different points in different cancers and since aberration of mutator genes could promote the genome instability during cancer development, the patterns of genetic aberrations tend to be nonrandom but differ between cancers of different tissues and of different subtypes from the same tissue.

Toward accelerating our understanding of tumorigenesis for better management of cancer patients, genomic approaches for systematically measuring somatic altered cancer genome and gene expression should be critical for clinical applications such as diagnosis, prognosis, classifying cancer subtypes and options of therapeutic treatment (2,3). However, identification of putative cancer gene is still hampered by the difficulty of further refining the precise aberrant region owing to the low resolution of chromosome alterations detected by CGH and the large deletions of LOH detected by microsatellite markers (4). In addition, the noisy data of chromosome aberrations and inconsistent results of altered gene expression detected by microarray and proteomic experiments, further demonstrated an emergent need of integrating qualified cancer genomic and expression data for developing new and effective cancer therapeutic targets. (5–8)

*To whom correspondence should be addressed. Tel: +886 2 26523521; Fax: +886 2 27827654; Email: jou@ibms.sinica.edu.tw

Human hepatocellular carcinoma (HCC) is the sixth most common cancer worldwide and the third most common cause of cancer death with prevalent areas in Asia and sub-Saharan Africa. (9) Although recent studies suggested increase of HCC incidence in western countries, >80% of the HCC cases occurred in above endemic areas are owing to exposure of major risk factors such as hepatitis viruses, mycotoxins and alcohol abuse. (10) Since HCC progression is usually asymptomatic resulted in poor prognosis and low 5-year survival rate (12–15%), comprehensive molecular genetic studies will be important for improving clinical management of HCC. Previous studies by our group and others have already conducted experiments of genome-wide LOH, CGH, microarray and proteomic analyses (11–14). Comprehensive analysis further allowed us to reveal two major genetic pathways, genome stable and instable pathways, in HCC progression. We therefore selected HCC as a cancer model to construct a public accessible integrated database, OncoDB.HCC, with user-friendly and graphically displayed interfaces as useful resources for facilitating researches in HCC tumorigenesis.

RESULTS

Construction of OncoDB.HCC

As indicated in Figure 1, the chromosome view interface demonstrates the main features of the database. The HCC data are constructed based on physical maps of human and rodent genome sequences from Ensembl and illustrated in several aspects including cancer genomic aberration data of LOH and CGH studies, altered gene expression in transcriptomes and proteomes analyses, genes with experimental data in HCC tissues reported in PubMed articles and the QTLs of rodent HCC models. The interactive interface further allows users to display chromosome regions defined by physical position, cytogenetic bands and sequence-tagged site (STS) markers. In expression view, a total of 9785 genes can be searched by using gene ID and gene description showing individual gene data including 9162 genes displayed the detail altered expression profiles of individual arrays reprocessed from re-analyzed Stanford HCC microarray data and experimental data of the gene extracted from published articles in PubMed. To demonstrate data reliability in our database, *AURKA*, a gene reported to be up-regulated in majority of HCC tissues (15), was selected as positive control to perform semi-quantitative RT–PCR experiments in 45 pairs of HCC tissues. As indicated in Figure 2A, the *AURKA* is highly up-regulated and the expression profiles are almost identical to that of re-analyzed Stanford HCC microarray data.

Selection and experimental validation of HCC genes

Stringent criteria were applied to select a set of 614 HCC genes with significant altered expression in HCC tissues (Supplementary files in OncoDB.HCC). Among them, 446 genes were supported with more than two independent studies in terms of altered expression including 145 concordantly up-regulated, 176 concordantly down-regulated and 125 genes with mixed up-/down-regulated expression in HCC tissues. The concordant expression of HCC genes could serve as potential biomarkers for HCC diagnosis. In addition, there were 234 genes with limited experimental data in HCC and 256 genes located within recurrent chromosome aberration regions. All of them represent potential targets for evaluation of their involvement in HCC tumorigenesis. In addition to *AURKA*, we further selected 30 out of 234 genes with limited wet-lab experiments in HCC for experimental validation. The validated results were demonstrated in terms of concordant expression of gene up- or down-regulation in 45 pairs of HCC tissues by RT–PCR analysis and in comparison to 57 pairs of HCC samples of re-analyzed Stanford HCC microarray data (Figure 2 and supplementary files in OncoDB.HCC). A near perfect concordant result except *CKAP2* (96.7%, 29/30 genes) in altered gene expression of HCC was obtained for 12 genes selected based on three independent microarray and/or proteomic reports and for 18 genes selected based on at least 2-fold expression changes within 70% patients in re-analyzed Stanford HCC microarray data.

Comparative mapping of HCC aberrant regions by using rodent HCC models

To provide additional genetic support and refine the HCC loci for targeting cancer genes, the QTLs of rodent HCC models identified via linkage studies were integrated into OncoDB.HCC based on comparative maps of rodent genomes in Ensembl. The results of 35 rodent HCC QTLs were displayed according to the relative positions of human chromosomes (Figure 1e). Comparative mapping of HCC loci demonstrated that over 45% rodent QTLs (8 of the 12 mouse HCC QTLs and another 8 of the 23 rat HCC QTLs) are located in the major aberrant loci of human HCC (Table 1). Among 16 syntenic QTLs located in 12 HCC loci, 10 QTLs in 6 loci are potentially located in gain/amplified regions and 6 QTLs in 6 loci are located in loss/deleted regions. While the critical regions of HCC loci existed in comparative genomes of human and rodents, the HCC loci could be effectively narrowed down due to the scrambled structure of genomes by comparing human and rodent syntenic regions. We narrowed down two human HCC loci to 2 Mb and another 6 HCC loci in between 4 and 10 Mb. Interestingly, the comparative mapping of HCC loci allowed us to split three human major HCC aberration regions 1q, 4q and 8p21–23 into two smaller regions and to conclude that at least two putative cancer genes located on the same arm of above three human HCC chromosomes.

DISCUSSION

OncoDB.HCC is the first attempt to establish a detail bioinformatic resource of one tumor genome by integrating genomic data of chromosome aberrations, altered gene expression, experimental data of genes in HCC tissues and QTLs of rodent HCC models. Three important advantages were revealed after data integration in OncoDB.HCC: First, data integration from independent studies containing aberrant consequences from levels of DNA, RNA and protein could avoid possible pitfalls of data inconsistency from a single

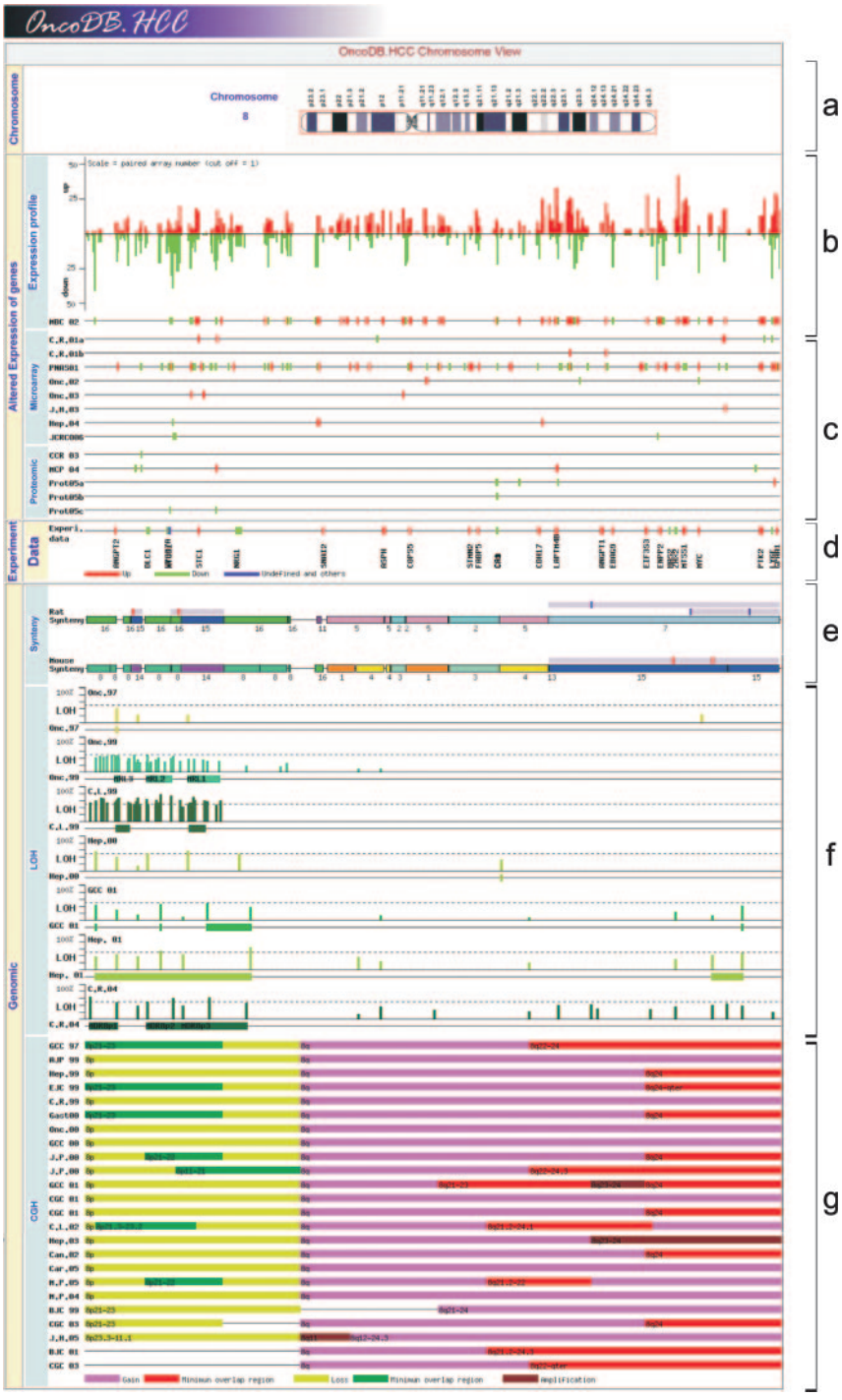

Figure 1. The chromosome view of OncoDB.HCC: (**a**) indicated the chromosomal region displayed in cytogenetic position; (**b**) demonstrated the expression intensity of genes along the physical positions of chromosome in terms of patient number by selecting gene expression cut-off value (default value = 1) in re-analyzed Stanford HCC microarray data; (**c**) showed the microarray/proteomic expression results from references; (**d**) indicated the positions of genes collected from wet-lab experimental results; (**e**) revealed the comparative maps and the syntenic regions of mouse and rat HCC QTLs; (**f**) displayed the LOH frequencies and minimum deletion regions (MDR) in positions of microsatellite markers; and (**g**) the cytogenetic locations of CGH results. All genes and markers were annotated and hyperlinked against physical maps of genomes in Ensembl. The up-regulated genes and gain/amplified chromosomal regions were displayed in red series color. In contrast, the down-regulated genes and loss/deleted chromosome regions were displayed in green series color.

genomic approach and provide lines of evidence to conclude somatic aberrations. Second, due to the heterogeneity nature of HCC tumorigenesis, successful gene validation in OncoDB.HCC is critical for revealing significantly altered HCC genes for 'signatures' of somatic aberrations in dissecting tumorigenic pathways and in clinical applications. Finally, integrated genomic data in OncoDB.HCC could narrow down and prioritize critical cancer genes and regions

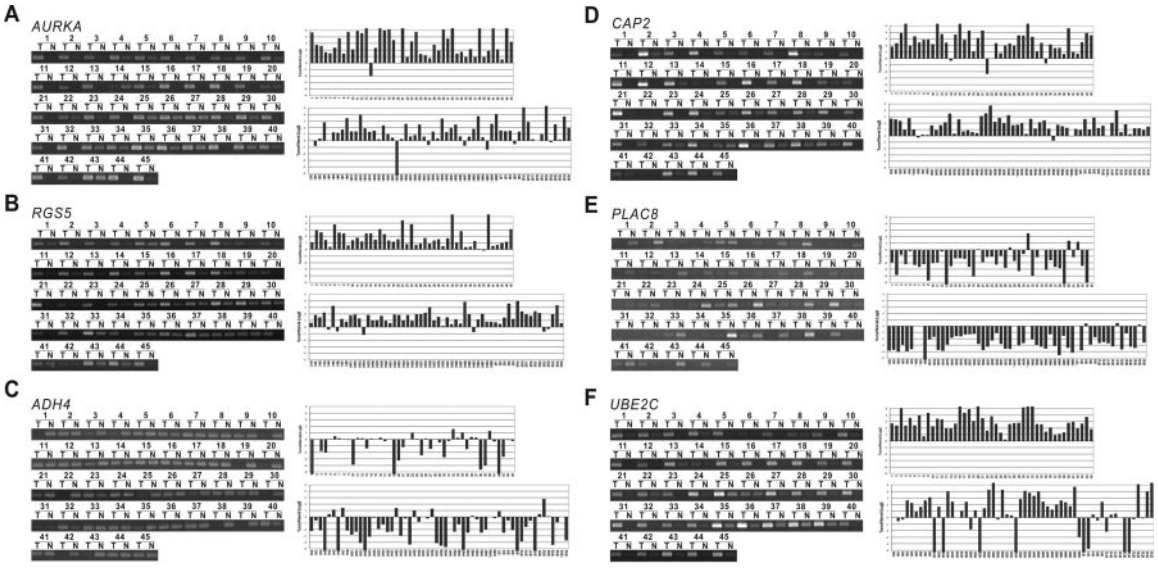

Figure 2. Experimental validation of representative genes in the HCC gene set. For each gene, the semi-quantitative RT–PCR results in 45 HCC pairs are presented in gel images (left), in quantified expression profiles after normalization with β-actin expression as internal control (upper right) and in comparison with expression profiles of the gene in re-analyzed Stanford HCC microarray data from expression view of OncoDB.HCC (lower right). (**A**) *AURKA* is a positive control for HCC data process; (**B** and **C**) genes selected from criteria of significant expression difference in at lease three independent microarray/proteomic studies; and (**D–F**) genes selected with criteria of at least 2-fold expression difference in at least 70% of paired arrays in re-analyzed Stanford HCC microarray data.

Table 1. HCC loci and putative cancer genes by comparative mapping strategy

	Human HCC Interval (Mb)	LOH[a]	CGH[b]	Rodent QTL Rat	Mouse	Putative cancer genes
Hcc1p	8 (59–67)	1	L (9)	—	Hcr1	*CACHD1*
Hcc1q1	49 (157–206)	4	G (25)	Hcrem5 Hcrem6	Hcf2 Hcs7	*COPA, ATF6, RGS5, GLUL, UBE2T, KISS1*
Hcc1q2	6 (227–233)	2	G (25)	—	Hcs2	
Hcc4q1	37 (52–89)	5	L (23)	—	Hcs5	*IGJ, SLC4A4, ALB, AFM, CXCL2, PLAC8, PTPN13, ABCG2*
Hcc4q2	6 (90–96)	3	L (22)	Drh2a	—	
Hcc6p	4 (30–34)	0	G (14)	—	Hcf1	*UBD, HSPA1B*
Hcc8p1	2 (10–12)	1	L (22)	Hcs7	—	*CTSB*
Hcc8p2	2 (18–20)	2	L (22)	Hcs4	—	*NAT2*
Hcc8q	38 (97–135)	2	G (24)	—	Hfib1	*LAPTM4B, PABPC1, ANGPT1, EIF3S6, EBAG9, ENPP2, ATAD2*
Hcc9p	7 (0–7)	3	L (3)	Drh1b	—	
Hcc11q	13 (58–71)	—	A (7), G (1)	Hcs3 Drh1a	—	*FEN1, FADS2, BAD, CDCA5*
Hcc20q	29 (29–58)	—	G (10)	—	Hcs4	*DNMT3B, E2F1, SRC, MYBL2, UBE2C, MMP9, CD40, AURKA*

[a]Number of references with significant LOH in the region defined by original authors.
[b]Number of CGH references with significant alterations covered this region. G: CGH gain; L: CGH loss; and A: CGH amplification.

for positional cloning and molecular studies of cancer genes in HCC. The quality of integrated data in OncoDB.HCC was experimentally supported by successful validation of altered expression in selected genes with limited wet-lab experimental studies in HCC. Therefore, the open access OncoDB.HCC should serve as a valuable resource for HCC research community.

FUTURE DEVELOPMENT

The OncoDB.HCC is the first comprehensive integration of cancer genomic data in one prevalent cancer with experimental validation and available freely to the research community. The future perspectives for OncoDB.HCC are to further integrate other newly emerging tumorigenic factors such as epigenetic modulations, point mutations and microRNA alterations in genome-wide aspects. In addition, commercial available 300K or 500K high density SNP chips are potentially useful to reveal high density novel genomic alterations in HCC genome. In conclusion, the comprehensive OncoDB.HCC is an invaluable resource for better understanding the tumorigenic mechanisms and developing useful information in clinical applications. OncoDB.HCC could serve as a bioinformatics resource that is applicable to other prevalent human cancers for dissecting tumorigenic pathways and the foundation of tumor systems biology.

ACKNOWLEDGEMENTS

This work was supported by the National Research Program for Genomic Medicine of National Science Council (NSC 95-3112-B-001-005) and funded in part by grant from

Thematic Research Program of Academia Sinica of Taiwan (AS-96-TP-B02). Funding to pay the Open Access publication charges for this article was provided by the above grants.

Conflict of interest statement. None declared.

REFERENCES

1. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
2. Alaiya,A., Al-Mohanna,M. and Linder,S. (2005) Clinical cancer proteomics: promises and pitfalls. *J. Proteome Res.*, **4**, 1213–1222.
3. Liang,P. and Pardee,A.B. (2003) Analysing differential gene expression in cancer. *Nature Rev. Cancer*, **3**, 869–876.
4. Zhou,X., Rao,N.P., Cole,S.W., Mok,S.C., Chen,Z. and Wong,D.T. (2005) Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizing high density single nucleotide polymorphism arrays. *Cancer Genet. Cytogenet.*, **159**, 53–57.
5. Hanash,S. (2004) Integrated global profiling of cancer. *Nature Rev. Cancer*, **4**, 638–644.
6. Tomlinson,I.P., Lambros,M.B. and Roylance,R.R. (2002) Loss of heterozygosity analysis: practically and conceptually flawed? *Genes Chromosomes Cancer*, **34**, 349–353.
7. Rhodes,D.R. and Chinnaiyan,A.M. (2005) Integrative analysis of the cancer transcriptome. *Nature Genet.*, **37**, S31–S37.
8. Albertson,D.G., Collins,C., McCormick,F. and Gray,J.W. (2003) Chromosome aberrations in solid tumors. *Nature Genet.*, **34**, 369–376.
9. Parkin,D.M., Bray,F., Ferlay,J. and Pisani,P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
10. Bosch,F.X., Ribes,J., Diaz,M. and Cleries,R. (2004) Primary liver cancer: worldwide incidence and trends. *Gastroenterology*, **127**, S5–S16.
11. Jou,Y.S., Lee,C.S., Chang,Y.H., Hsiao,C.F., Chen,C.F., Chao,C.C., Wu,L.S., Yeh,S.H., Chen,D.S. and Chen,P.J. (2004) Clustering of minimal deleted regions reveals distinct genetic pathways of human hepatocellular carcinoma. *Cancer Res.*, **64**, 3030–3036.
12. Nishida,N., Nishimura,T., Ito,T., Komeda,T., Fukuda,Y. and Nakao,K. (2003) Chromosomal instability and human hepatocarcinogenesis. *Histol. Histopathol.*, **18**, 897–909.
13. Kim,J.W. and Wang,X.W. (2003) Gene expression profiling of preneoplastic liver disease and liver cancer: a new era for improved early detection and treatment of these deadly diseases? *Carcinogenesis*, **24**, 363–369.
14. Thorgeirsson,S.S. and Grisham,J.W. (2002) Molecular pathogenesis of human hepatocellular carcinoma. *Nature Genet.*, **31**, 339–346.
15. Jeng,Y.M., Peng,S.Y., Lin,C.Y. and Hsu,H.C. (2004) Overexpression and amplification of Aurora-A in hepatocellular carcinoma. *Clin. Cancer Res.*, **10**, 2065–2071.