3. APPLICATION OF SAGE FOR CANCER TARGET/BIOMARKER IDENTIFICATION

3.1. Differential Gene Expression Between Normal and Tumor Tissues

SAGE has been used for the analysis of various tumors to decipher the pathogenesis of cancer development by comparing the gene expression patterns with those of corresponding normal tissues [34-36]. The first application of SAGE for comparison of normal and tumor tissues was reported in 1997, analyzing the expression profiles of colon and pancreatic tumors [37]. Although the authors showed the extensive similarity between normal and cancer tissues of the same organ, they discovered approximately 500 transcripts differentially expressed between normal and cancer cells, including many novel genes. Since then, various types of tumors have been analyzed by SAGE including lung, breast, ovarian, brain, gastric, and liver cancers [35, 38-44]. These results have been compiled in the SAGE map (http://www.ncbi.nlm.nih.gov/SAGE/) as a part of the Cancer Genome Anatomy Project (CGAP) and GEO database, and the expression of genes in each tissue type can be visualized in SAGE Anatomic Viewer as Digital Northern Results (http://cgap.nci.nih.gov/SAGE/AnatomicViewer).

These large amounts of data of various cancers have elucidated the principles of gene expression patterns commonly observed in tumors, and major findings have been well documented in recent reviews [32, 45]. The fundamental characteristics of cancer transcriptome are reemphasized here: 1) tumors share major gene expression traits with normal tissues that they originate from, 2) tumors increase the expression of genes associated with cell proliferation signaling, and 3) tumors decrease the expression of genes associated with the function of differentiated normal cells. These characteristics of cancer transcriptome evoke the dedifferentiated cellular phenotypes of cancer cells with an ability to grow rapidly. On the other hand, SAGE has also provided the candidate genes altered in specific type of tumors such as colon, breast, ovarian, brain, liver, and pancreatic cancers [46]. For example, *GPC3,* which encodes a heparan sulfate proteoglycan Glypican 3, was initially considered as a down-regulated gene in various cancers including mesothelioma, ovarian, and breast cancers [47]. However, up-regulation of *GPC3* was identified in hepatocellular carcinoma (HCC) by the differential display and SAGE [44, 48]. More recent studies have demonstrated up-regulation of Glypican 3 in majority of HCC at the protein level and suggested its utility for early detection of HCC [49-51]. Although the usefulness of biomarkers identified by gene expression profiling (e.g. *GPC3* in liver cancer) still remains elusive, further evaluation may reveal the utility of identified markers for early detection of specific type of cancers [44, 52, 53].

3.2. Elucidation of Signaling Pathway Targets in Cancer

The study of colon carcinogenesis in familial adenomatous polyposis patients suggested the stepwise accumulation of genetic alterations involved in various signaling pathways at distinct histopathological stages, implicating that about 4-6 genetic events are required for malignant transformation of human epithelial cells [54, 55]. Recently, Bert Vogelstein and his group have explored genetic changes accumulated in cancers by sequencing approximately 13,000 protein coding genes distributed on whole genome [6, 7]. Interestingly, the analyses suggested that there are ~80 DNA mutations that alter amino acids in cancer cells. They performed statistical analyses to distinguish gene mutations that affect the tumorigenicity (driver mutations) from those do not (passenger mutations), and suggested that <15 DNA mutations were likely to be responsible for driving the initiation, progression, or maintenance of the tumor. Although not all mutations are likely to drive the initiation and promotion of the tumor, the number of DNA mutations in developed tumors seems to be higher than previously expected. The authors have discovered a number of candidate cancer genes mutated in breast and colorectal cancers (referred as candidate cancer genes; *CAN*-genes). Strikingly, the analysis of 11 colorectal cancer genomes has ascertained the mutations of oncogenes and tumor suppressor genes that are known to play a pivotal role in the pathogenesis of colorectal cancer, including *APC*, *KRAS*, *EPHA3*, *TGFBR2*, *TP53*, and *SMAD4*. Furthermore, they have demonstrated that many genetic alterations detected by cancer genome sequencing could be integral to smaller number of cell signaling pathways, suggesting the potential involvement of these novel mutated genes in the known characterized pathways.

SAGE is an ideal method to detect all alteration of transcriptome induced by the activation of certain signaling pathways, and has been utilized to identify key target genes in many signaling molecules [56]. The first example of SAGE application for target identification was reported in 1997, using colorectal cancer cell lines to discover p53 targets [57]. Remarkably, the analysis had provided 7 genes previously unrelated to the p53 signaling such as *PIG3*, supporting the strength of the method to provide novel insights into p53 signaling. SAGE has also been utilized to identify the targets of other signaling pathways such as adenomatous polyposis coli (APC), c-Myc, transforming growth factor-β (TGF-β), Tek/Tie2, and Von Hippel-Lindau (VHL)/hypoxia [14-16, 58-62]. A part of these studies are summarized in recent reviews [56], and most of these genes are now recognized as representative targets of each signaling pathway, including peroxisome proliferator-activated receptor (PPAR)-δ and c-Myc in Wnt/β-catenin signaling pathway [14, 15]. Thus, SAGE is a potentially powerful tool to identify key targets of signaling pathways in cancer. The application of SAGE may provide important information to build a connection map of signaling pathways activated by *CAN*-genes.

3.3. Identification of Biomarkers Associated with Prognosis

Investigating Oncogenes/Tumor Suppressor Genes

In general, prognosis of a cancer patient is determined by the nature of tumor cells in terms of proliferation and metastasis. Consistently, analyses of microarray datasets of various tumors revealed the correlation of genes involved in cell growth and proliferation with poor outcomes [63]. Initiation and promotion of tumors are supposed to be dictated by the activation of oncogenes as well as inactivation of tumor suppressor genes, and most of these genes are known to regulate

cell cycle and proliferation [1, 2]. Therefore, investigation of oncogenes/tumor suppressor genes expression may enable prognostic stratification of cancer patients, such as reported in p53 mutations [64, 65]. Since the majority of SAGE data are publicly available, it is feasible to have a glimpse of the expression patterns of any genes of interest in various tumors at once. A recent study utilized the public SAGE dataset for identification of oncogenes associated with poor outcomes in ovarian cancer [66]. The authors investigated the expression of the gene families BTB/POZ (broad-complex, tramtrack, and bric-a-brac/poxvirus and zinc finger) known to be mutated in B cell lymphoma using SAGE libraries, and identified the activation of *NAC1* in ovarian tumors. They further confirmed the overexpression of *NAC1* in other tumors such as pancreatic, liver, prostate, kidney, and breast cancers using SAGE libraries publicly available. Notably, overexpression of *NAC1* correlates with recurrent ovarian cancers, suggesting the utility of *NAC1* for stratification of ovarian cancers with poor prognosis.

Exploring Signaling Pathways

Considering the fact that most of gene products seem to work in a form of complex as a transducer of signaling pathways, one would expect that the identification of specific signal transduction from gene expression data may allow isolating pathways that are correlated with cellular proliferation and prognosis. Indeed, recent advancement of bioinformatics has enabled to identify the lists of pathways activated or inactivated in cancer [9]. We recently combined the SAGE data with bioinformatics tools to identify the lipogenic pathway activated in a subset of HCC (Yamashita, T.; Honda, M.; Takatori, H.; Nishino, R. and Kaneko, S., manuscript in submission). Deactivation of lipogenic pathway resulted in the inhibition of cell proliferation and caused apoptosis *in vitro*. Furthermore, we identified that the overexpression of transcription factor regulating the lipogenic pathway (Sterol Regulatory Element Binding Protein 1; SREBP-1) correlated with poor prognosis, indicating the usefulness of this approach for detection of biomarker pathways. Similarly, a recent study has explored SAGE data to identify the subgroups of estrogen receptor-positive breast cancers with different metabolic activities that may be a determinant of prognosis [67]. Thus, pathway analysis of gene expression in cancer appears a promising method to decipher the mechanism of cell proliferation that is linked to poor prognosis.

Discovering Metastasis-Related Genes

As tumors progress to increase malignancy, they obtain the ability to invade into surrounding tissues to form new tumors at sites distinct from the primary tumors called metastasis. The process of cancer metastasis is considered as the sequential steps of the invasion-metastasis cascades, including local invasiveness, intravasation into and transport through the circulation, extravasation, formation of a micrometastasis, and colonization [68, 69]. The molecular mechanisms involved in this process are complex and influenced by various host and tumor factors, including cell-cell and cell-matrix adhesion, degradation of extracellular matrix, hypoxia, and various growth factors and chemoattractants. Since metastasis accounts for about 90% of cancer death in solid tumors [70], investigation of gene signatures of metastatic and non-metastatic tumors will provide molecular clues for diagnosis of cancer with poor prognosis. Microarray analysis has been extensively used to examine the global gene expression changes in metastatic and non-metastatic cancers [71]. For example, analyses of metastatic and non-metastatic breast cancers have provided the metastatic gene signature that could predict the survival of breast cancer patients [12, 72]. Similarly, analysis of metastatic and non-metastatic HCC patients has provided the gene signature that could distinguish metastatic HCC from non-metastatic even using primary tumor tissues [13]. These data suggested that the transcriptional alterations associated with metastasis are highly represented in their primary tumors, suggesting the utility of this approach for stratification of patients' prognoses. Furthermore, three groups recently utilized SAGE approach to uncover the signatures of breast cancer, thyroid cancer, and melanomas that are metastatic to the lymph nodes [73-75]. Although these studies have shown striking similarities of gene expression patterns between primary and metastatic tumors, they have successfully provided candidate genes associated with metastasis such as *HOXC10*, *LIMD2*, and *UBC9* in breast cancer, papillary thyroid cancer, and melanoma, respectively. Interestingly, all these studies suggested the activation of ATPase transporters or ATP-binding cassette (ABC) multidrug transporters in metastatic cancers, suggesting the involvement of these transporters on metastasis. Investigation of the functional role of these genes may provide profound insight into molecular portraits of metastatic cancer cells within primary tumors.

Predicting the Drug Response

Although cancer patients diagnosed at advanced stages may receive chemo/radiation therapies, it is hard to predict the responses in each cancer patient prior to treatment, mainly due to the lack of specific biomarkers. Estrogen receptors and HER-2 are good examples that have predictive information on outcome and sensitivity to a particular treatment in breast cancer. Identification of drug sensitivity markers using gene expression profiling may stratify patients who do (or do not) respond the particular treatment with distinct prognosis. A cytogenetic study of anaplastic oligodendrogliomas indicated the strong association of allelic losses of chromosome 1p and 19q with longer overall survival and good chemo sensitivity, suggesting the utility of genetic analysis for therapeutic decisions [76]. Similarly, microarray studies have identified the 70-gene signature that could predict high risk groups of breast cancer recurrence who may have the survival benefits by adjuvant chemotherapy [77, 78]. Its prospective validation is currently ongoing through the MINDACT (Microarray in Node-Negative Disease May Avoid Chemotherapy) trial [79]. Gene expression profiling has also provided the mechanistic insights into cellular chemo sensitivity. A recent SAGE study has clarified the overexpression of extracellular matrix genes including *COL6A3* in cisplatin-resistant ovarian cancer cells [80]. Furthermore, an independent SAGE study to compare the gene expression profiles of various tumors with different chemo resistance has shown the differential expression of the extracellular matrix proteins [81]. Indeed, a role for tumor-matrix interactions in the acquisition of drug resistance has recently emerged [82]. SAGE studies have provided new

insights into the function of extracellular matrix proteins on tumor intractability to chemotherapy, and may provide the potential targets for elimination of tumors by inhibiting these targets in combination with chemotherapy.

As described above, enormous amounts of gene expression data generated by SAGE can be exploited to facilitate the identification of useful biomarkers/targets for treatment of cancer. The identified biomarkers may help subgrouping cancer patients with poor prognosis, although rigorous validation studies are clearly required using large-size patients' cohorts.

4. FUTURE GOALS

One of the fascinating advantages of using SAGE in cancer research is to identify the completely novel genes previously not detected that are altered in cancer. Certainly, we often experience the SAGE tags that can not match to the human genome even after the HGP and Celera genomics had provided the reliable human genome sequence data. Some of these tags have been mapped to the virus genome such as human papilloma virus and hepatitis B virus (Yamashita, T.; Honda, M.; Takatori, H.; Nishino, R. and Kaneko, S., manuscript in submission) [83], but the origin of many unknown tags is still uncertain. Recently, all publicly available human LongSAGE tags were investigated to map on human genomes [84]. Strikingly, the results have shown that about 36% of transcripts can not map to the human genome, and 58% of these can not be explained by polymorphisms, exon-intron boundaries, poly-A tail, and contamination of mouse transcripts (derived from SAGE libraries of human ES cell lines co-cultured with mouse embryonic fibroblasts). They have also identified that 31% of tags mapped to the genome are located outside of annotated exons, suggesting that these genes are still not annotated and could be candidates of novel genes. Similar results were obtained by the analysis of LongSAGE libraries of the human brain tissues [85]. Thus, a greater part of the transcripts unable to map to the human genome remains obscure, and about one-third of the transcripts mapped to the human genome is still not annotated. Future studies are clearly required to investigate whether these unmapped tags are derived from simple sequence errors, sequences correlated with unknown diversities of human genomes, or unknown peculiar transcriptional programs to generate unspecified transcripts.

Majority of gene expression profiling studies have been performed using bulk tumor tissues that are composed of various cell types such as epithelial, endothelial, mesenchymal, and inflammatory cells, as depicted in Fig. (**2**). Therefore, identified gene signatures may reflect the altered cellular heterogeneity rather than gene expression changes in tumor epithelial cells. Through SAGE analysis of purified various types of normal and tumor cells, Kornelia Polyak and her group have provided the strong evidence of tumor-microenvironment interaction in invasive breast cancer [86]. Although they found genetic alterations only in tumor epithelial cells, they identified dramatic gene expression changes in stromal cells that affect the metastatic ability of transformed tumor epithelial cells. Similarly, recent studies have successfully shown the differential gene expression between normal and tumor vascular endothelium that can be

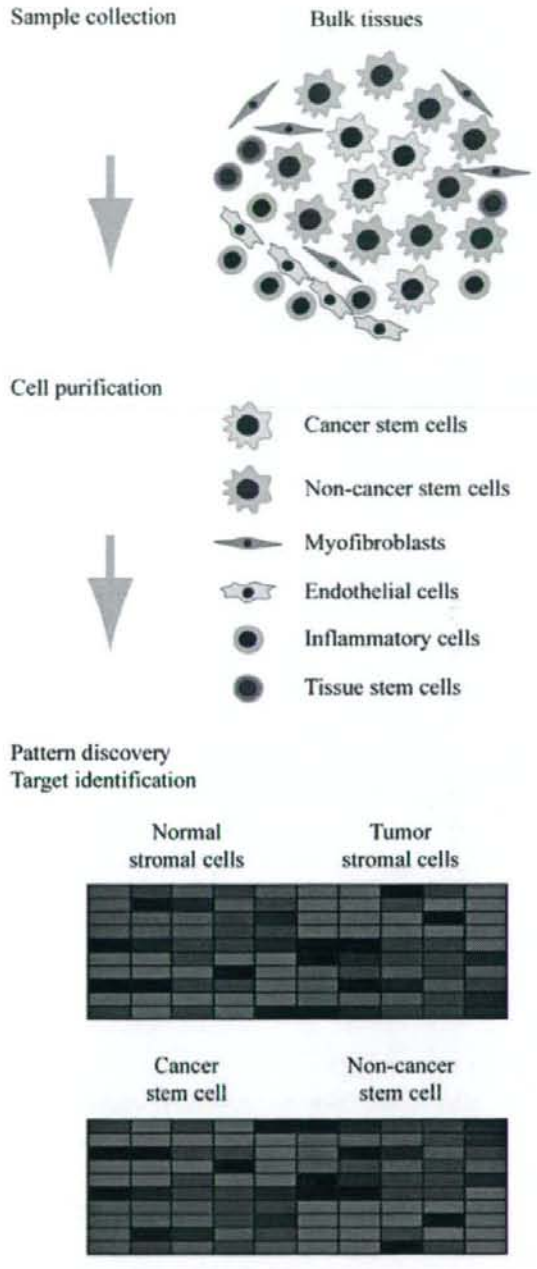

Fig. (2). Deciphering the cancer-microenvironment interaction in bulk tumor tissues by cell purification.

utilized for targeting tumor-specific vascular endothelial cells [87, 88]. These studies have accentuated the concept that cancers are heterogeneous cellular entities whose growth is dependent on reciprocal interactions between tumor cells and the dynamic microenvironment in which they reside [82]. Transcriptome studies of tumor stromal cells as well as

epithelial cells will lead to the discovery of underlying molecular targets in cancer-microenvironment interaction for eradication of tumors.

Cellular heterogeneity exists not only in stromal cells but also in purified tumor epithelial cells. Cancer stem cell concept, a subset of cells with stem cell features is indispensable for the development of tumor, has recently revived by the advancement of stem cell biology [89]. Accumulating evidence suggests the role of cancer stem cells on tumorigenesis, metastasis, and resistance to chemo/radiation, implicating the importance of targeting cancer stem cell population for elimination of tumor [90]. Although critics question the validity of the cancer stem cell hypothesis, comprehensive transcriptome studies of normal and putative cancer stem cells as well as non-cancer stem cells may clarify the key signaling pathways responsible for the stemness of these cells [91], and will provide clues for elimination of tumors in future.

ABBREVIATIONS

SAGE	=	Serial analysis of gene expression
HGP	=	Human genome project
RT-PCR	=	Reverse transcription-polymerase chain reaction
EST	=	Expressed sequence tag
HCC	=	Hepatocellular carcinoma
APC	=	Adenomatous polyposis coli
TGF-β	=	Transforming growth factor-β
VHL	=	Von Hippel-Lindau
PPAR	=	Peroxisome proliferator-activated receptor
BTB/POZ	=	Broad-complex, tramtrack, and bric-a-brac/poxvirus and zinc finger
SREBP-1	=	Sterol regulatory element binding protein 1
ABC	=	ATP-binding cassette
MINDACT	=	Microarray in node-negative disease may avoid chemotherapy

REFERENCES

[1] Vogelstein, B. and Kinzler, K. W. (2004) *Nat. Med.*, **10**(8), 789-799.

[2] Hanahan, D. and Weinberg, R. A. (2000) *Cell*, **100**(1), 57-70.

[3] Liang, P. and Pardee, A. B. (2003) *Nat. Rev. Cancer*, **3**(11), 869-876.

[4] Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A. and Zhu, X. (2001) *Science*, **291**(5507), 1304-1351.

[5] Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge,

C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; de Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S. and Chen, Y. J. (2001) *Nature*, **409**(6822), 860-921.

[6] Sjoblom, T.; Jones, S.; Wood, L. D.; Parsons, D. W.; Lin, J.; Barber, T. D.; Mandelker, D.; Leary, R. J.; Ptak, J.; Silliman, N.; Szabo, S.; Buckhaults, P.; Farrell, C.; Meeh, P.; Markowitz, S. D.; Willis, J.; Dawson, D.; Willson, J. K.; Gazdar, A. F.; Hartigan, J.; Wu, L.; Liu, C.; Parmigiani, G.; Park, B. H.; Bachman, K. E.; Papadopoulos, N.; Vogelstein, B.; Kinzler, K. W. and Velculescu, V. E. (2006) *Science*, **314**(5797), 268-274.

[7] Wood, L. D.; Parsons, D. W.; Jones, S.; Lin, J.; Sjoblom, T.; Leary, R. J.; Shen, D.; Boca, S. M.; Barber, T.; Ptak, J.; Silliman, N.; Szabo, S.; Dezso, Z.; Ustyanksky, V.; Nikolskaya, T.; Nikolsky, Y.; Karchin, R.; Wilson, P. A.; Kaminker, J. S.; Zhang, Z.; Croshaw, R.; Willis, J.; Dawson, D.; Shipitsin, M.; Willson, J. K.; Sukumar, S.; Polyak, K.; Park, B. H.; Pethiyagoda, C. L.; Pant, P. V.; Ballinger, D. G.; Sparks, A. B.; Hartigan, J.; Smith, D. R.; Suh, E.; Papadopoulos, N.; Buckhaults, P.; Markowitz, S. D.; Parmigiani, G.; Kinzler, K. W.; Velculescu, V. E. and Vogelstein, B. (2007) *Science*, **318**(5853) 1108-1113.

[8] Mohr, S.; Leikauf, G. D.; Keith, G. and Rihn, B. H. (2002) *J. Clin. Oncol.*, **20**(14), 3165-3175.

[9] Weeraratna, A. T. (2005) *Ageing Res. Rev.*, **4**(4), 548-563.

[10] Alizadeh, A. A.; Eisen, M. B.; Davis, R. E.; Ma, C.; Lossos, I. S.; Rosenwald, A.; Boldrick, J. C.; Sabet, H.; Tran, T.; Yu, X.; Powell, J. I.; Yang, L.; Marti, G. E.; Moore, T.; Hudson, J., Jr.; Lu, L.; Lewis, D. B.; Tibshirani, R.; Sherlock, G.; Chan, W. C.; Greiner, T. C.; Weisenburger, D. D.; Armitage, J. O.; Warnke, R.; Levy, R.; Wilson, W.; Grever, M. R.; Byrd, J. C.; Botstein, D.; Brown, P. O. and Staudt, L. M. (2000) *Nature*, **403**(6769), 503-511.

[11] Perou, C. M.; Sorlie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Rees, C. A.; Pollack, J. R.; Ross, D. T.; Johnsen, H.; Akslen, L. A.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, S. X.; Lonning, P. E.; Borresen-Dale, A. L.; Brown, P. O. and Botstein, D. (2000) *Nature*, **406**(6797), 747-752.

[12] van de Vijver, M. J.; He, Y. D.; van't Veer, L. J.; Dai, H.; Hart, A. A.; Voskuil, D. W.; Schreiber, G. J.; Peterse, J. L.; Roberts, C.; Marton, M. J.; Parrish, M.; Atsma, D.; Witteveen, A.; Glas, A.; Delahaye, L.; van der Velde, T.; Bartelink, H.; Rodenhuis, S.; Rutgers, E. T.; Friend, S. H. and Bernards, R. (2002) *N. Engl. J. Med.*, **347**(25), 1999-2009.

[13] Ye, Q. H.; Qin, L. X.; Forgues, M.; He, P.; Kim, J. W.; Peng, A. C.; Simon, R.; Li, Y.; Robles, A. I.; Chen, Y.; Ma, Z. C.; Wu, Z. Q.; Ye, S. L.; Liu, Y. K.; Tang, Z. Y. and Wang, X. W. (2003) *Nat. Med.*, **9**(4), 416-423.

[14] He, T. C.; Chan, T. A.; Vogelstein, B. and Kinzler, K. W. (1999) *Cell*, **99**(3), 335-345.

[15] He, T. C.; Sparks, A. B.; Rago, C.; Hermeking, H.; Zawel, L.; da Costa, L. T.; Morin, P. J.; Vogelstein, B. and Kinzler, K. W. (1998) *Science*, **281**(5382), 1509-1512.

[16] Hermeking, H.; Lengauer, C.; Polyak, K.; He, T. C.; Zhang, L.; Thiagalingam, S.; Kinzler, K. W. and Vogelstein, B. (1997) *Mol. Cell*, **1**(1), 3-11.

[17] Adams, M. D.; Dubnick, M.; Kerlavage, A. R.; Moreno, R.; Kelley, J. M.; Utterback, T. R.; Nagle, J. W.; Fields, C. and Venter, J. C. (1992) *Nature*, **355**(6361), 632-634.

[18] Adams, M. D.; Kerlavage, A. R.; Fields, C. and Venter, J. C. (1993) *Nat. Genet.*, **4**(3), 256-267.

[19] Adams, M. D.; Soares, M. B.; Kerlavage, A. R.; Fields, C. and Venter, J. C. (1993) *Nat. Genet.*, **4**(4) 373-380.

[20] Velculescu, V. E.; Zhang, L.; Vogelstein, B. and Kinzler, K. W. (1995) *Science*, **270**(5235) 484-487.

[21] Schena, M.; Shalon, D.; Davis, R. W. and Brown, P. O. (1995) *Science*, **270**(5235), 467-470.

[22] Chee, M.; Yang, R.; Hubbell, E.; Berno, A.; Huang, X. C.; Stern, D.; Winkler, J.; Lockhart, D. J.; Morris, M. S. and Fodor, S. P. (1996) *Science*, **274**(5287), 610-614.

[23] Lockhart, D. J.; Dong, H.; Byrne, M. C.; Follettie, M. T.; Gallo, M. V.; Chee, M. S.; Mittmann, M.; Wang, C.; Kobayashi, M.; Horton, H. and Brown, E. L. (1996) *Nat. Biotechnol.*, **14**(13), 1675-1680.

[24] Brown, P. O. and Botstein, D. (1999) *Nat. Genet.*, **21**(1 Suppl), 33-37.

[25] Lipshutz, R. J.; Fodor, S. P.; Gingeras, T. R. and Lockhart, D. J. (1999) *Nat. Genet.*, **21**(1 Suppl), 20-24.

[26] Husson, H.; Manavalan, P.; Akmaev, V. R.; Russo, R. J.; Cook, B.; Richards, B.; Barberio, D.; Liu, D.; Cao, X.; Landes, G. M.; Wang, C. J.; Roberts, B. L.; Klinger, K. W.; Grubman, S. A.; Jefferson, D. M. and Ibraghimov-Beskrovnaya, O. (2004) *Genomics*, **84**(3), 497-510.

[27] Honda, M.; Yamashita, T.; Ueda, T.; Takatori, H.; Nishino, R. and Kaneko, S. (2006) *Hepatology*, **44**(5), 1122-1138.

[28] Yao, J.; Weremowicz, S.; Feng, B.; Gentleman, R. C.; Marks, J. R.; Gelman, R.; Brennan, C. and Polyak, K. (2006) *Cancer Res.*, **66**(8), 4065-4078.

[29] Zhang, Y.; Huang, J.; Jia, S.; Liu, W.; Li, M.; Wang, S.; Miao, X.; Xiao, H. and Huang, Y. (2007) *Genomics*, **90**(3) 372-379.

[30] Sawiris, G. P.; Sherman-Baust, C. A.; Becker, K. G.; Cheadle, C.; Teichberg, D. and Morin, P. J. (2002) *Cancer Res.*, **62**(10), 2923-2928.

[31] Toyoda, N.; Nagai, S.; Terashima, Y.; Motomura, K.; Haino, M.; Hashimoto, S.; Takizawa, H. and Matsushima, K. (2003) *Genome Res.*, **13**(7), 1728-1736.

[32] Porter, D. and Polyak, K. (2003) *Expert Opin. Ther. Targets*, **7**(6), 759-769.

[33] Simon, R. (2005) *J. Clin. Oncol.*, **23**(29), 7332-7341.

[34] Boon, K.; Osorio, E. C.; Greenhut, S. F.; Schaefer, C. F.; Shoemaker, J.; Polyak, K.; Morin, P. J.; Buetow, K. H.; Strausberg, R. L.; De Souza, S. J. and Riggins, G. J. (2002) *Proc. Natl. Acad. Sci. USA*, **99**(17), 11287-11292.

[35] Lal, A.; Lash, A. E.; Altschul, S. F.; Velculescu, V.; Zhang, L.; McLendon, R. E.; Marra, M. A.; Prange, C.; Morin, P. J.; Polyak, K.; Papadopoulos, N.; Vogelstein, B.; Kinzler, K. W.; Strausberg, R. L. and Riggins, G. J. (1999) *Cancer Res.*, **59**(21), 5403-5407.

[36] Lash, A. E.; Tolstoshev, C. M.; Wagner, L.; Schuler, G. D.; Strausberg, R. L.; Riggins, G. J. and Altschul, S. F. (2000) *Genome Res.*, **10**(7), 1051-1060.

[37] Zhang, L.; Zhou, W.; Velculescu, V. E.; Kern, S. E.; Hruban, R. H.; Hamilton, S. R.; Vogelstein, B. and Kinzler, K. W. (1997) *Science*, **276**(5316), 1268-1272.

[38] Hibi, K.; Liu, Q.; Beaudry, G. A.; Madden, S. L.; Westra, W. H.; Wehage, S. L.; Yang, S. C.; Heitmiller, R. F.; Bertelsen, A. H.; Sidransky, D. and Jen, J. (1998) *Cancer Res.*, **58**(24), 5690-5694.

[39] Hough, C. D.; Sherman-Baust, C. A.; Pizer, E. S.; Montz, F. J.; Im, D. D.; Rosenshein, N. B.; Cho, K. R.; Riggins, G. J. and Morin, P. J. (2000) *Cancer Res.*, **60**(22), 6281-6287.

[40] Sasaki, M.; Nakahira, K.; Kawano, Y.; Katakura, H.; Yoshimine, T.; Shimizu, K.; Kim, S. U. and Ikenaka, K. (2001) *Cancer Res.*, **61**(12), 4809-4814.

[41] Porter, D. A.; Krop, I. E.; Nasser, S.; Sgroi, D.; Kaelin, C. M.; Marks, J. R.; Riggins, G. and Polyak, K. (2001) *Cancer Res.*, **61**(15), 5697-5702.

[42] Cuevas, I. C.; Slocum, A. L.; Jun, P.; Costello, J. F.; Bollen, A. W.; Riggins, G. J.; McDermott, M. W. and Lal, A. (2005) *Cancer Res.*, **65**(12), 5070-5075.

[43] Oue, N.; Hamai, Y.; Mitani, Y.; Matsumura, S.; Oshimo, Y.; Aung, P. P.; Kuraoka, K.; Nakayama, H. and Yasui, W. (2004) *Cancer Res.*, **64**(7), 2397-2405.

[44] Yamashita, T.; Kaneko, S.; Hashimoto, S.; Sato, T.; Nagai, S.; Toyoda, N.; Suzuki, T.; Kobayashi, K. and Matsushima, K. (2001) *Biochem. Biophys. Res. Commun.*, **282**(2), 647-654.

[45] Polyak, K. and Riggins, G. J. (2001) *J. Clin. Oncol.*, **19**(11), 2948-2958.

[46] Riggins, G. J. (2001) *Dis. Markers*, **17**(2), 41-48.

[47] Jakubovic, B. D. and Jothy, S. (2007) *Exp. Mol. Pathol.*, **82**(2), 184-189.

[48] Hsu, H. C.; Cheng, W. and Lai, P. L. (1997) *Cancer Res.*, **57**(22), 5179-5184.

[49] Capurro, M.; Wanless, I. R.; Sherman, M.; Deboer, G.; Shi, W.; Miyoshi, E. and Filmus, J. (2003) *Gastroenterology*, **125**(1), 89-97.

[50] Hippo, Y.; Watanabe, K.; Watanabe, A.; Midorikawa, Y.; Yamamoto, S.; Ihara, S.; Tokita, S.; Iwanari, H.; Ito, Y.; Nakano, K.; Nezu, J.; Tsunoda, H.; Yoshino, T.; Ohizumi, I.; Tsuchiya, M.; Ohnishi, S.; Makuuchi, M.; Hamakubo, T.; Kodama, T. and Aburatani, H. (2004) *Cancer Res.*, **64**(7), 2418-2423.

[51] Libbrecht, L.; Severi, T.; Cassiman, D.; Vander Borght, S.; Pirenne, J.; Nevens, F.; Verslype, C.; van Pelt, J. and Roskams, T. (2006) *Am. J. Surg. Pathol.*, **30**(11) 1405-1411.

[52] Jia, H. L.; Ye, Q. H.; Qin, L. X.; Budhu, A.; Forgues, M.; Chen, Y.; Liu, Y. K.; Sun, H. C.; Wang, L.; Lu, H. Z.; Shen, F.; Tang, Z. Y. and Wang, X. W. (2007) *Clin. Cancer Res.*, **13**(4), 1133-1139.

[53] Llovet, J. M.; Chen, Y.; Wurmbach, E.; Roayaie, S.; Fiel, M. I.; Schwartz, M.; Thung, S. N.; Khitrov, G.; Zhang, W.; Villanueva, A.; Battiston, C.; Mazzaferro, V.; Bruix, J.; Waxman, S. and Friedman, S. L. (2006) *Gastroenterology*, **131**(6) 1758-1767.

[54] Kinzler, K. W. and Vogelstein, B. (1996) *Cell*, **87**(2), 159-170.

[55] Hahn, W. C. and Weinberg, R. A. (2002) *Nat. Rev. Cancer*, **2**(5), 331-341.

[56] Hermeking, H. (2003) *Curr. Opin. Oncol.*, **15**(1) 44-49.

[57] Polyak, K.; Xia, Y.; Zweier, J. L.; Kinzler, K. W. and Vogelstein, B. (1997) *Nature*, **389**(6648), 300-305.

[58] Zawel, L.; Yu, J.; Torrance, C. J.; Markowitz, S.; Kinzler, K. W.; Vogelstein, B. and Zhou, S. (2002) *Proc. Natl. Acad. Sci. USA*, **99**(5), 2848-2853.

[59] Menssen, A. and Hermeking, H. (2002) *Proc. Natl. Acad. Sci. USA*, **99**(9), 6274-6279.

[60] Chen, S. H.; Babichev, Y.; Rodrigues, N.; Voskas, D.; Ling, L.; Nguyen, V. P. and Dumont, D. J. (2005) *Physiol. Genomics*, **22**(2), 257-267.

[61] Jiang, Y.; Zhang, W.; Kondo, K.; Klco, J. M.; St Martin, T. B.; Dufault, M. R.; Madden, S. L.; Kaelin, W. G., Jr. and Nacht, M. (2003) *Mol. Cancer Res.*, **1**(6), 453-462.

[62] Potapova, O.; Anisimov, S. V.; Gorospe, M.; Dougherty, R. H.; Gaarde, W. A.; Boheler, K. R. and Holbrook, N. J. (2002) *Cancer Res.*, **62**(11), 3257-3263.

[63] Whitfield, M. L.; George, L. K.; Grant, G. D. and Perou, C. M. (2006) *Nat. Rev. Cancer*, **6**(2), 99-106.

[64] Greenblatt, M. S.; Bennett, W. P.; Hollstein, M. and Harris, C. C. (1994) *Cancer Res.*, **54**(18), 4855-4878.

[65] Poeta, M. L.; Manola, J.; Goldwasser, M. A.; Forastiere, A.; Benoit, N.; Califano, J. A.; Ridge, J. A.; Goodwin, J.; Kenady, D.; Saunders, J.; Westra, W.; Sidransky, D. and Koch, W. M. (2007) *N. Engl. J. Med.*, **357**(25) 2552-2561.

[66] Nakayama, K.; Nakayama, N.; Davidson, B.; Sheu, J. J.; Jinawath, N.; Santillan, A.; Salani, R.; Bristow, R. E.; Morin, P. J.; Kurman, R. J.; Wang, T. L. and Shih Ie, M. (2006) *Proc. Natl. Acad. Sci. USA*, **103**(49), 18739-18744.

[67] Mandal, S. and Davie, J. R. (2007) *BMC Cancer*, **7**, 181.

[68] Stracke, M. L. and Liotta, L. A. (1992) *In vivo*, **6**(4) 309-316.

[69] Cairns, R. A.; Khokha, R. and Hill, R. P. (2003) *Curr. Mol. Med.*, **3**(7), 659-671.

[70] Nguyen, D. X. and Massague, J. (2007) *Nat. Rev. Genet.*, **8**(5), 341-352.

[71] Yang, J. (2006) *Assay Drug. Dev. Technol.*, **4**(4) 483-488.

[72] Ramaswamy, S.; Ross, K. N.; Lander, E. S. and Golub, T. R. (2003) *Nat. Genet.*, **33**(1), 49-54.

[73] Moschos, S. J.; Smith, A. P.; Mandic, M.; Athanassiou, C.; Watson-Hurst, K.; Jukic, D. M.; Edington, H. D.; Kirkwood, J. M. and Becker, D. (2007) *Oncogene*, **26**(29), 4216-4225.

[74] Abba, M. C.; Sun, H.; Hawkins, K. A.; Drake, J. A.; Hu, Y.; Nunez, M. I.; Gaddis, S.; Shi, T.; Horvath, S.; Sahin, A. and Aldaz, C. M. (2007) *Mol. Cancer Res.*, **5**(9), 881-890.

[75] Cerutti, J. M.; Oler, G.; Michaluart, P., Jr.; Delcelo, R.; Beaty, R. M.; Shoemaker, J. and Riggins, G. J. (2007) *Cancer Res.*, **67**(16), 7885-7892.

[76] Cairncross, J. G.; Ueki, K.; Zlatescu, M. C.; Lisle, D. K.; Finkelstein, D. M.; Hammond, R. R.; Silver, J. S.; Stark, P. C.; Macdonald, D. R.; Ino, Y.; Ramsay, D. A. and Louis, D. N. (1998) *J. Natl. Cancer Inst.*, **90**(19), 1473-1479.

[77] van 't Veer, L. J.; Dai, H.; van de Vijver, M. J.; He, Y. D.; Hart, A. A.; Mao, M.; Peterse, H. L.; van der Kooy, K.; Marton, M. J.; Witteveen, A. T.; Schreiber, G. J.; Kerkhoven, R. M.; Roberts, C.; Linsley, P. S.; Bernards, R. and Friend, S. H. (2002) *Nature*, **415**(6871), 530-536.

[78] Naderi, A.; Teschendorff, A. E.; Barbosa-Morais, N. L.; Pinder, S. E.; Green, A. R.; Powe, D. G.; Robertson, J. F.; Aparicio, S.; Ellis, I. O.; Brenton, J. D. and Caldas, C. (2007) *Oncogene*, **26**(10), 1507-1516.

[79] Cardoso, F.; Van't Veer, L.; Rutgers, E.; Loi, S.; Mook, S. and Piccart-Gebhart, M. J. (2008) *J. Clin. Oncol.*, **26**(5) 729-735.

[80] Sherman-Baust, C. A.; Weeraratna, A. T.; Rangel, L. B.; Pizer, E. S.; Cho, K. R.; Schwartz, D. R.; Shock, T. and Morin, P. J. (2003), *Cancer Cell*, **3**(4) 377-386.

[81] Stein, W. D.; Litman, T.; Fojo, T. and Bates, S. E. (2004) *Cancer Res.*, **64**(8), 2805-2816.

[82] Tlsty, T. D. and Coussens, L. M. (2006) *Annu. Rev. Pathol.*, **1**, 119-150.

[83] Shadeo, A.; Chari, R.; Vatcher, G.; Campbell, J.; Lonergan, K. M.; Matisic, J.; van Niekerk, D.; Ehlen, T.; Miller, D.; Follen, M.; Lam, W. L. and MacAulay, C. (2007) *BMC Genomics*, **8**, 142.

[84] Keime, C.; Semon, M.; Mouchiroud, D.; Duret, L. and Gandrillon, O. (2007) *BMC Bioinformatics*, **8**, 154.

[85] Peters, B. A.; St Croix, B.; Sjoblom, T.; Cummins, J. M.; Silliman, N.; Ptak, J.; Saha, S.; Kinzler, K. W.; Hatzis, C. and Velculescu, V. E. (2007) *Genome Res.*, **17**(3), 287-292.

[86] Allinen, M.; Beroukhim, R.; Cai, L.; Brennan, C.; Lahti-Domenici, J.; Huang, H.; Porter, D.; Hu, M.; Chin, L.; Richardson, A.; Schnitt, S.; Sellers, W. R. and Polyak, K. (2004) *Cancer Cell*, **6**(1), 17-32.

[87] St Croix, B.; Rago, C.; Velculescu, V.; Traverso, G.; Romans, K. E.; Montgomery, E.; Lal, A.; Riggins, G. J.; Lengauer, C.; Vogelstein, B. and Kinzler, K. W. (2000) *Science*, **289**(5482), 1197-1202.

[88] Seaman, S.; Stevens, J.; Yang, M. Y.; Logsdon, D.; Graff-Cherry, C. and St Croix, B. (2007) *Cancer Cell*, **11**(6), 539-554.

[89] Clarke, M. F.; Dick, J. E.; Dirks, P. B.; Eaves, C. J.; Jamieson, C. H.; Jones, D. L.; Visvader, J.; Weissman, I. L. and Wahl, G. M. (2006) *Cancer Res.*, **66**(19), 9339-9344.

[90] Hill, R. P. and Perris, R. (2007) *J. Natl. Cancer Inst.*, **99**(19), 1435-1440.

[91] Shipitsin, M.; Campbell, L. L.; Argani, P.; Weremowicz, S.; Bloushtain-Qimron, N.; Yao, J.; Nikolskaya, T.; Serebryiskaya, T.; Beroukhim, R.; Hu, M.; Halushka, M. K.; Sukumar, S.; Parker, L. M.; Anderson, K. S.; Harris, L. N.; Garber, J. E.; Richardson, A. L.; Schnitt, S. J.; Nikolsky, Y.; Gelman, R. S. and Polyak, K. (2007) *Cancer Cell*, **11**(3), 259-273.

Received: February 13, 2008 **Accepted: March 12, 2008**

ELSEVIER

Journal of Hepatology 49 (2008) 946–954

Journal of Hepatology

www.elsevier.com/locate/jhep

Expression of multidrug resistance-associated protein 3 and cytotoxic T cell responses in patients with hepatocellular carcinoma☆

Eishiro Mizukoshi, Masao Honda, Kuniaki Arai, Tatsuya Yamashita, Yasunari Nakamoto, Shuichi Kaneko*

Department of Gastroenterology, Graduate School of Medicine, Kanazawa University, Kanazawa, Ishikawa 920-8641, Japan

***Background/Aims*: Multidrug resistance-associated protein 3 (MRP3) is a carrier-type transport protein belonging to the ABC transporters. It is expressed in normal tissues, and enhanced expression in many cancers has been reported. In this study, we investigated the usefulness of MRP3 as a target antigen in immunotherapy for hepatocellular carcinoma (HCC).**

***Methods*: The MRP3 expression level in HCC tissue was measured by quantitative PCR. MRP3-specific T cell responses were investigated by several immunological techniques using peripheral blood mononuclear cells or tumor-infiltrating lymphocytes.**

***Results*: The MRP3 expression level in HCC tissue was significantly higher than that in non-cancerous tissue ($P < 0.05$). MRP3-specific cytotoxic T cells (CTLs) could be induced regardless of liver function, the presence or absence of HCV infection, the blood AFP level, and the stage of HCC. The CTLs showed cytotoxicity against HCC cells overexpressing MRP3. A negative correlation was present between the MRP3 expression level in HCC tissue and the frequency of MRP3-specific CTLs. The frequency of MRP3-specific CTLs increased after HCC treatment, such as transcatheter arterial embolization and radiofrequency ablation.**

***Conclusions*: Our study demonstrates that MRP3 is a potential candidate for tumor antigen with strong immunogenicity in HCC immunotherapy.**

Keywords: **Immune response; CD8; HLA-A24; Hepatitis; Cancer**

1. Introduction

Hepatocellular carcinoma (HCC) is treatable by hepatectomy or percutaneous ablation when the lesion is localized to some extent, and radical therapeutic effects can be obtained when the resection or cauterization with a safety margin can be performed [1,2]. However, active hepatitis and cirrhosis in the surrounding non-tumor liver tissues exhibit high carcinogenetic potentials to develop *de novo* HCC, and therefore, the recurrence rate of HCC after treatment is very high [3,4].

To protect against recurrence, tumor antigen-specific immunotherapy is an attractive strategy. For the development of HCC-specific immunotherapy and analysis of immune responses to the treatment, the identification of HCC-specific tumor antigens or their antigenic epitopes is necessary. However, only a few HCC-specific tumor antigens and their antigenic epitopes have been identified [5–10].

MRP3 is a carrier-type transport protein belonging to the ABC transporters that transport substances against

Received 11 January 2008; received in revised form 23 April 2008; accepted 7 May 2008; available online 5 June 2008

Associate Editor: V. Barnaba

☆ The authors declare that they do not have anything to disclose regarding funding from industries or conflict of interest with respect to this manuscript.

* Corresponding author. Tel.: +81 76 265 2230; fax: +81 76 234 4250.

E-mail address: skaneko@m-kanazawa.jp (S. Kaneko).

Abbreviations: HLA, human leukocyte antigen; IFN, interferon; PBMC, peripheral blood mononuclear cell; TIL, tumor-infiltrating lymphocytes; HCV, hepatitis C virus; ELISPOT, enzyme linked immunospot.

0168-8278/$34.00 © 2008 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jhep.2008.05.012

the concentration gradient in an ATP energy-dependent manner [11]. It is expressed at a high level in the small and large intestine, pancreas, placenta, and adrenal cortex [12], and recent studies have reported that its expression is enhanced in various cancer cells [13–15]. Yamada et al. demonstrated that MRP3 was a tumor rejection antigen recognized by cytotoxic T cells (CTLs), using lymphocytes infiltrating into lung adenocarcinoma, and identified its CTL epitope [13]. These reports suggest that MRP3 may be useful as a target antigen in HCC immunotherapy. However, the MRP3 expression level in HCC tissue has been controversial [16,17], and the association between the expression level and the degree of the immune response to MRP3 in HCC patients has not been clarified.

In this study, we measured the MRP3 expression level in various hematoma cell lines and HCC tissues in HCC patients, and analyzed immune responses to MRP3 using peripheral blood mononuclear cells (PBMCs) and tumor-infiltrating lymphocytes (TILs) to investigate the usefulness of MRP3 in HCC immunotherapy.

2. Materials and methods

2.1. Patients

This study examined 103 HLA-A24-positive patients with HCC (Table 1). All subjects were negative for Abs to human immunodeficiency virus (HIV), and gave written informed consent to participate in this study in accordance with the Helsinki declaration. The diagnosis of HCC was histologically confirmed by taking US-guided needle biopsy specimens in 35 cases, surgical resection in 13 cases, and autopsy in 4 cases. For the remaining 51 patients, the diagnosis was based on typical hypervascular tumor staining on angiography in addition to typical findings, which showed hyperattenuated areas in the early phase and hypoattenuation in the late phase on dynamic CT [18]. Eleven healthy blood donors with HLA-A24, who did not have a history of cancer and were negative for HBsAg and anti-HCVAb, served as controls.

2.2. Laboratory and virologic testing

Blood samples were tested for HBsAg and HCVAb by commercial immunoassays (Fuji Rebio, Tokyo, Japan). HLA-based typing of PBMC from patients and normal donors was performed as previously described [10].

The serum AFP level was measured by enzyme immunoassay (AxSYM AFP, Abbott Japan, Tokyo, Japan) and the pathological grading of tumor cell differentiation was assessed according to the general rules for the clinical and pathologic study of primary liver cancer [19]. The severity of liver disease (stage of fibrosis) was evaluated according to the criteria of Desmet et al. [20] using biopsy specimens of liver tissue, where F4 was defined as cirrhosis.

2.3. Cell lines

Eight human hepatoma cell lines: HepG2, Alex, Huh6, HLE, HLF, Hep3B, SKHep1, and Huh7, were cultured in DMEM (Gibco, Grand Island, NY, USA) with 10% fetal calf serum (FCS) (Gibco, Grand Island, NY, USA). The HLA-A*2402 gene-transfected C1R cell line (C1R-A24) [21] was cultured in RPMI 1640 medium containing 10% FCS and 500 μg/ml of hygromycin B (Sigma, St. Louis, MO, USA), and K562 was cultured in RPMI 1640 medium containing 10% FCS.

2.4. Quantitative real time detection (RTD)-PCR

We performed quantitative RTD-PCR using TaqMan Universal Master Mix (PE Applied Biosystems, Foster City, CA, USA). Primer pairs and probes for MRP3 and β-actin were obtained from TaqMan assay reagents library. Total RNA was isolated from cell lines and liver tissue samples using an RNA extraction kit (Micro RNA Extraction Kit, Stratagene, La Jolla, CA, USA). We reverse-transcribed 1 μg of isolated RNA to cDNA using SuperScript® II RT (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, and the resultant cDNA was amplified with appropriate TaqMan assay reagents as previously described [22].

2.5. Preparation of PBMCs and TILs

PBMCs and TILs were isolated as previously described [9]. Fresh PBMCs were used for the CTL induction, and the remaining PBMCs and TILs were resuspended in RPMI 1640 medium containing 80% FCS and 10% dimethyl sulfoxide and cryopreserved until used. In patients with treatment, PBMCs were obtained before and 2–4 weeks after the treatment.

2.6. CTL induction and cytotoxicity assay

Peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$, which were identified to contain a HLA-A24 restricted CTL epitope [13], were used for the induction of MRP3-specific T cells (Table 2). Peptides were synthesized at Mimotope (Melbourne, Australia) and Sumitomo Pharmaceuticals (Osaka, Japan). They were identified using mass spectrometry, and their purities were determined to be >80% by analytical HPLC. CTLs were expanded from PBMCs as previously described [23]. Briefly, four hundred thousand cells per well were stimulated with synthetic peptides at 10 μg/ml, 10 ng/ml rIL-7 and 100 pg/ml rIL-12 (Sigma, St. Louis, Mo) in RPMI 1640 supplemented with 10% heat inactivated human AB serum, 100 U/ml penicillin and 100 μg/ml streptomycin. The cultures were re-stimulated with 10 μg/ml peptide, 20 U/ml rIL-2 (Sigma, St. Louis, MO) and 10^5 mytomicin C treated autologous PBMCs on days 7 and 14. On days 3, 10 and 17, 100 μl of RPMI with 10% human AB serum and 10 U/ml rIL-2 (final concentration) was added to each well.

C1R-A24 cells and human hepatoma cell lines were used as target cells. Cytotoxicity assays were performed in at least 10 HCC patients for each peptide as previously described [10]. Spontaneous release

Table 1
Characteristics of the patients studied

Clinical diagnosis	No. of patients	Sex M/F	Age (yr) Mean ± SD	ALT (IU/L) Mean ± SD	AFP (ng/ml) mean ± SD	Etiology (B/C/B + C/Others)	Child Pugh (A/B/C)	Diff. degree[a] (Wel/Mod/Por/ND)	Tumor size[b] (Large/Small)	Tumor multiplicity (Multiple/Solitary)	Vascular Invasion (+/−)	TNM stage (I/II/IIIA/IIIB/IIIC/IV)
HCC patients	103	79/24	63 ± 10	65 ± 37	3155 ± 15946	19/73/2/9	61/39/3	17/31/4/51	79/24	73/30	30/73	24/51/16/1/3/8
Normal donors	11	8/3	35 ± 2	ND	ND	ND	ND	ND	ND	ND	ND	ND

[a] Histological degree of HCC; wel: well-differentiated, mod: moderately differentiated, por: poorly differentiated, ND: not determined.
[b] Tumor size was divided into either 'small' (≤2 cm) or 'large' (>2 cm).

Table 2
Peptides

Peptide	Source	Start position	Amino acid sequence	HLA restriction
$MRP3_{503}$	MRP3	503	LYAWEPSFL	HLA-A24
$MRP3_{692}$	MRP3	692	AYVPQQAWI	HLA-A24
$MRP3_{765}$	MRP3	765	VYSDADIFL	HLA-A24
HIV env_{584}	HIV envelope	584	RYLRDQQLL	HLA-A24
CMV $pp65_{328}$	CMV pp65	328	QYDPVAALF	HLA-A24
AFP_{403}	AFP	403	KYIQESQAL	HLA-A24

was <15% of the maximum release for all experiments. For the assay using hepatoma cell lines, the cytotoxic activity was considered positive when it was higher than that of CTL against K562 which shows non-specific lysis. The assay was performed at least three times for each peptide.

2.7. ELISPOT assay

ELISPOT assays were performed as previously described with the following modifications [9,10]. Peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ were used for the detection of MRP3-specific T cells. Negative controls consisted of a HIV envelope-derived peptide ($HIVenv_{584}$) [24]. Positive controls consisted of 10 ng/ml phorbol 12-myristate 13-acetate (PMA, Sigma) or a CMV pp65-derived peptide ($CMVpp65_{328}$) [25]. The colored spots were counted with a KS ELISpot Reader (Zeiss, Tokyo, Japan). The number of specific spots was determined by subtracting the number of spots in the absence of antigen from the number of spots in its presence. Responses for peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ in HCC patients were considered positive if more than the mean + 3SD specific spots in healthy normal donors were detected and if the number of spots in the presence of antigen was at least twofold greater than the number of spots in its absence. Responses for peptides $HIVenv_{584}$ and $CMVpp65_{328}$ were considered positive if more than 10 specific spots were detected and if the number of spots in the presence of antigen was at least twofold greater than the number of spots in the absence of antigen.

2.8. Tetramer staining and flow cytometry

Peptide $MRP3_{765}$ specific tetramer was purchased from Medical Biological Laboratories Co., Ltd (Nagoya, Japan). Tetramer staining was performed according to a previously reported method with several modifications [10]. In brief, PBMCs were stained with CD8-PerCP (BD PharMingen, San Diego, CA, USA) and tetramer-PE (10 μl) for 30 min at room temperature. Cells were washed, fixed with 0.5% paraformaldehyde/PBS, and analyzed on a FACSCalibur™ flow cytometer. Data analysis was undertaken with CELLQuest™ software (Becton–Dickinson, San Jose, CA, USA).

2.9. Statistical analysis

Data are expressed as means ± SD. The χ^2 test with Yates' correction, Fisher's exact probability test, and the unpaired t-test were used for statistical analyses where appropriate. Linear regression lines for the relationship between expression of MRP3 mRNA and MRP3-specific immune responses were calculated using Pearson's correlation coefficient. A level of $P < 0.05$ was considered significant.

3. Results

3.1. Patient profile

The clinical profiles of the patients are shown in Table 1. In 52 patients, HCC was histologically classified as well-, moderately, and poorly differentiated HCC in 17, 31, and 4, respectively. In the other patients, HCC was diagnosed based on typical CT findings and AFP elevation. On tumor classification based on the size and number, the tumor was large (>2 cm) in 79, small (≤2 cm) in 24, multiple in 73, and solitary in 30. Vascular invasion was noted in 30 patients. On tumor classification using the TNM staging of the Union Internationale Contre Le Cancer (UICC) classification system (6th version), 24, 51, 16, 1, 3, and 8 patients were staged I, II, IIIA, IIIB, IIIC, and IV, respectively.

3.2. Expression of MRP3 in hepatoma cell lines and HCC tissues

To investigate the MRP3 expression level in HCC, we measured MRP3 mRNA in 8 hepatoma cell lines by real-time PCR. The expression ratio of MRP3 to β-actin, measured as an internal control, is shown in Fig. 1A. All hepatoma cell lines except HLF expressed MRP3, but the expression level varied among the cell lines. HepG2, Hep3B and Huh7 showed high expression levels, but Alex, HLE, SKHep1 and Huh6 showed low expression levels.

The MRP3 expression level in HCC tissues was compared with non-cancerous tissues in specimens obtained from 20 HCC patients by US-guided needle tumor biopsy or surgical resection. The MRP3 expression level was significantly higher in HCC tissue than in the non-cancerous tissue ($P < 0.05$) (Fig. 1B). In the analysis of the individual MRP3 expression levels, 11 of 20 (55%) HCC tissues showed higher expression level than that of Huh 7 whose average of expression level is 1.0 (Fig. 1C).

3.3. Cytotoxic activity of MRP3 peptide-specific CTL against hepatoma cell lines

Whether the MRP3-derived peptides used were capable of inducing peptide-specific CTL from PBMCs was investigated in at least 10 HCC patients. The CTLs specific for $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ were induced in 3, 3 and 2 patients, respectively. As shown in Fig. 2A, all CTL induced with $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ showed high-level cytotoxicity against C1RA24 cells pulsed with the corresponding peptides.

These CTLs exhibited cytotoxicity against hepatoma cell lines with the HLA-A24 molecule and high expression of MRP3, HepG2 and HLE, but not against MRP3-hypoexpressing Huh6 and MRP3-overexpressing Huh7 without HLA-A24 molecule (Fig. 2B).

3.4. T Cell responses to MRP3-derived peptides assessed by IFN-γ ELISPOT analysis

To determine a significant number of T cells that specifically reacted with $MRP3_{503}$, $MRP3_{692}$ and $MRP3_{765}$

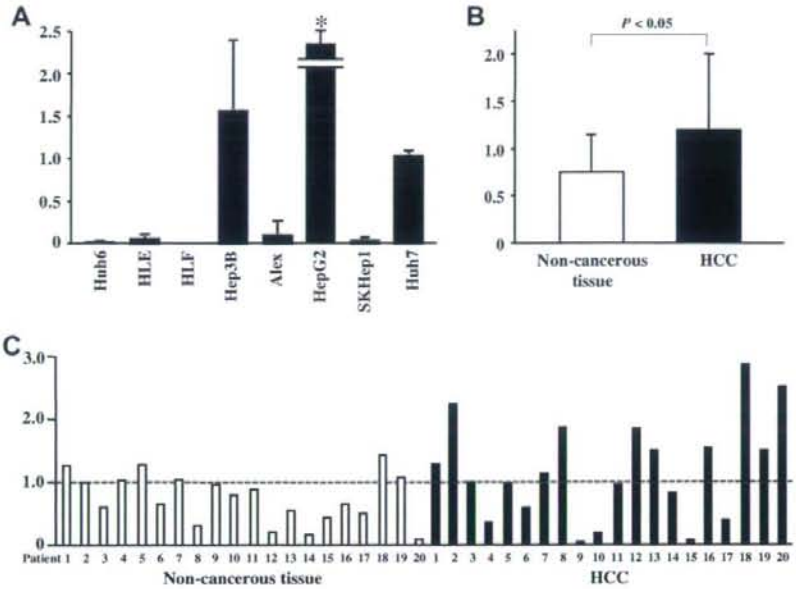

Fig. 1. Expression levels of MRP3 mRNA. (A) Expression of MRP3 mRNA was measured by real-time PCR in hepatoma cell lines. (B) Comparison of MRP3 mRNA expression levels between non-tumor (white bar) and tumor (solid bar) tissues. The data are expressed as means + SD. The unpaired *t*-test was used for a statistical analysis. (C) Comparison of MRP3 mRNA expression levels between non-tumor (white bar) and tumor (solid bar) tissues in individual HCC patients. * denotes 6.15 + 4.21.

peptides in HCC patients, ELISPOT assays were performed using PBMCs from 11 healthy donors. The number of specific spots was 0.2 ± 0.5, 1.5 ± 2.1, 0.9 ± 1.0, 1.3 ± 2.0, and 13.3 ± 15.7 cells/3×10^5 PBMCs, respectively (Fig. 3). Similarly, cells that specifically reacted with the peptides were counted in HCC patient-derived PBMCs. Regarding a number of T cells that specifically reacted with the peptide of larger than the mean + 3SD of that in healthy donor-derived PBMCs as a significant response, 20.0, 14.1, and 21.4% of the patients showed significant responses to $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$, respectively (Fig. 4A). A significant response specific for $CMVpp65_{328}$ was detected in 51.0% and 36.4% of the HCC patients and healthy donors, respectively, showing no significant difference between the 2 groups. On the other hand, no significant response for $HIVenv_{584}$ was observed in both groups.

On similar analysis of TIL, 75.0, 75.0, and 37.5% of the patients showed significant responses to $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$, respectively, revealing that the frequencies were higher than those in PBMCs (Fig. 4B).

3.5. Detection of $MRP3_{765}$ tetramer$^+$ and CD8$^+$ T lymphocytes in PBMCs

The frequency of MRP3-specific T cells was also investigated using $MRP3_{765}$ tetramer in 20 HCC patients. To confirm the specificity of $MRP3_{765}$ tetramer, we tried to detect the tetramer$^+$ cells in a CTL line induced by stimulation with $MRP3_{765}$ peptide. The frequency of $MRP3_{765}$ tetramer$^+$ cells in CD8$^+$ cells was increased from 0.03% before to 9.15% after stimulation (Fig. 5A). When PBMC was stimulated with irrelevant peptide (AFP_{403}), the frequency of $MRP3_{765}$ tetramer$^+$ cells was only 0.08%.

To count the frequency of tetramer$^+$ cells in peripheral blood, we used freshly isolated non-stimulated PBMCs for the assay. The tetramer$^+$ and CD8$^+$ T cells accounted for 0.00–0.23% in PBMCs of HCC patients (Fig. 5B). Next, the results were compared with those of ELISPOT assay. In patients 1 to 7, both $MRP3_{765}$ tetramer$^+$ and IFN-γ producing cells responding to the peptide in ELISPOT assay were detected. In contrast, in patients 8 to 13, the frequency of tetramer$^+$ cells was high, but no significant increase in the $MRP3_{765}$ peptide-specific T cell count was detected by the ELISPOT assay.

3.6. MRP3-specific T cell responses and clinical features of HCC patients

To clarify the clinical characteristics of MRP3-specific T cell responses in HCC patients, the clinical background was compared between patients who showed positive responses to MRP3-derived peptides on ELISPOT assay and those who did not. No significant differences were noted between the 2 groups (Table 3).

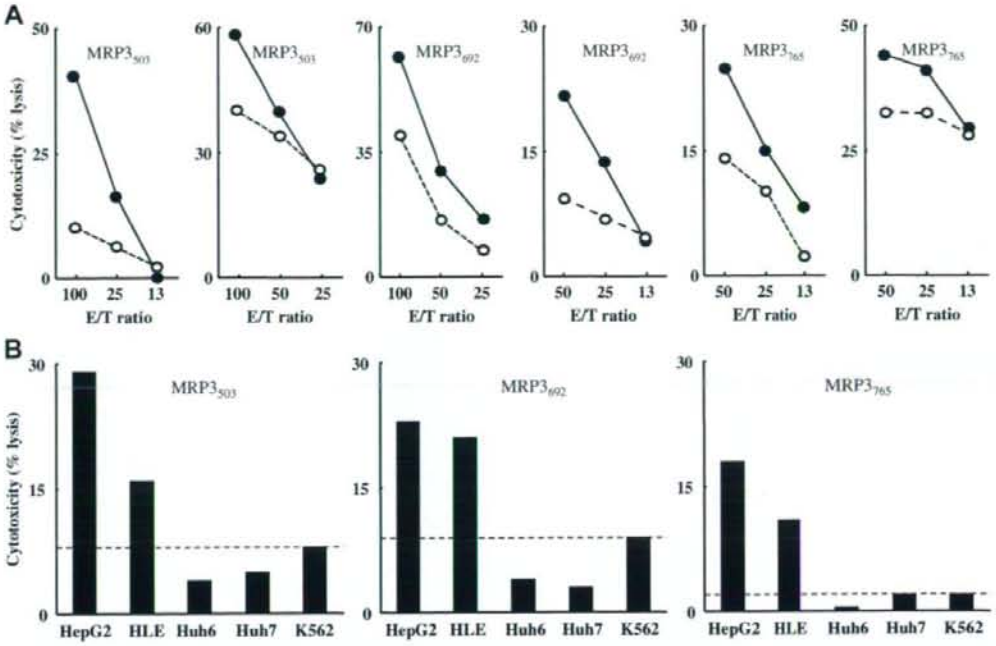

Fig. 2. Cytotoxicity of MRP3-specific T-cell lines derived with peptide in patients with HCC. (A) The cytotoxicity of T-cell lines was determined by a standard 6-h cytotoxicity assay at various effector to target (E/T) ratios against C1R-A*2402 cells pulsed with or without one of the MRP3-derived peptides listed in Table 2. The open circle shows the cytotoxicity against C1R-A*2402 cells pulsed without a peptide. The closed circle shows the cytotoxicity against C1R-A*2402 cells pulsed with a peptide. (B) Cytotoxicity of MRP3-specific T-cell lines derived with peptide was also measured against hepatoma cell lines. The cytotoxicity was considered positive when it was higher than that against K562 which shows non-specific lysis. HepG2 expresses MRP3 and has HLA-A*2402. Huh 6 and HLE show a low expression of MRP3 and have HLA-A*2402. Huh 7 shows MRP3 expression, but does not have HLA-A*2402. Cytotoxicity was determined by a standard 6-h cytotoxic assay (E/T ratio of 50:1).

In 20 HCC patients in whom the MRP3 expression level in HCC tissue could be measured, the relationship between the expression level and frequency of MRP3-specific T cells was investigated. A significant negative correlation was present between the MRP3 expression level in HCC tissue and MRP3-specific T cell frequency ($r = -0.54$, $P < 0.05$) (Fig. 6A). When the relationship between the MRP3 expression level in HCC tissue and CMVpp65-specific T-cell frequency was similarly analyzed, no significant correlation was present. Furthermore, when the patients were divided into groups with high and low HCC tissue MRP3 expression levels, setting the border to the mean MRP3 expression level in the normal liver tissues, 0.743, the peripheral blood MRP3-specific T cell frequency was significantly higher in the low- than in the high-level group ($p < 0.05$) (Fig. 6B). The CMVpp65-specific T cell frequency was not significantly different between the 2 groups.

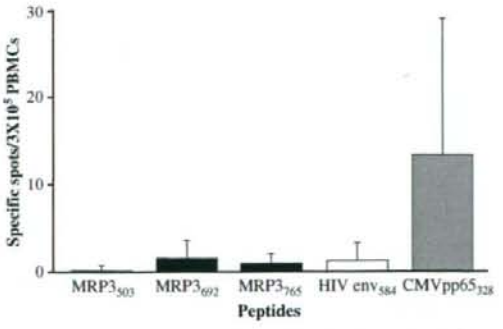

Fig. 3. Direct ex-vivo analysis (IFN-γ ELISPOT assay) of peripheral blood T cell responses to MRP3-derived peptides (peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$: solid bars) or control peptides (peptides $HIVenv_{584}$ and $CMVpp65_{328}$: open and grey bars, respectively) in healthy normal donors. The data are expressed as means + SD.

3.7. Enhancement of MRP3-specific T cell responses after anti-cancer treatment

Several studies including our report have clarified that HCC treatment enhanced HCC-specific immune responses [9,26,27]. We investigated whether MRP3-specific T cell responses observed in HCC patients were enhanced by HCC treatment. In 12 patients who underwent TAE or radiofrequency ablation (RFA) or both

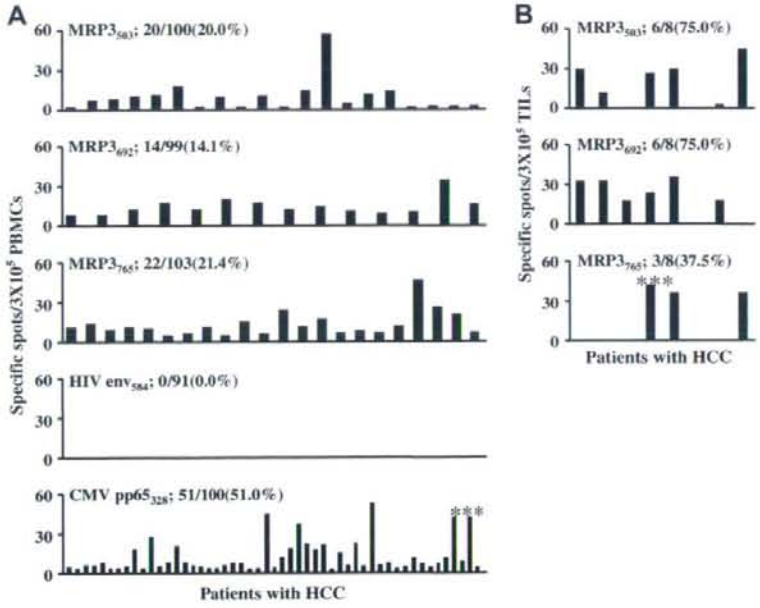

Fig. 4. Direct ex-vivo analysis (IFN-γ ELISPOT assay) of PBMCs (A) and TILs (B) response to MRP3-derived peptides (peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$) or control peptides (peptides $HIVenv_{584}$ and $CMVpp65_{328}$) in HCC patients. Only significant IFN-γ responses are included. Responses to peptides $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ were considered positive if more than the mean + 3SD specific spots in healthy normal donors were detected and if the number of spots in the presence of antigen was at least twofold greater than that in its absence. Responses to peptides $HIVenv_{584}$ and $CMVpp65_{328}$ were considered positive if more than 10 specific spots were detected and if the number of spots in the presence of antigen was at least twofold greater than that in its absence. The peptide sequences are described in Table 2. * denotes 770 specific spots. ** denotes 210 specific spots. * denotes 72 specific spots.**

without MRP3-specific T-cell responses before treatment, changes in the MRP3-specific T cell frequency were investigated by measuring the frequency by ELISPOT assay before and after treatment. The $MRP3_{503}$, $MRP3_{692}$, or $MRP3_{765}$ peptide-specific T cell frequency was increased after treatment in 8 of the 12 patients (Table 4). In contrast, the immune response to $HIVenv_{584}$ peptide was not enhanced in any patient, and that to $CMVpp65_{328}$ peptide was enhanced in one patient.

4. Discussion

The expression of MRP3 has been reported in several normal tissues and cancer cells [14,15,28,29]. Although MRP3 expression in HCC tissue was confirmed by immunohistochemical staining [16], the expression level varied among patients, and a conclusion has not been reached as to whether the expression is increased compared to that in normal liver tissue [16,17]. In this study, MRP3 expression in HCC tissue was detected in all 20 HCC patients, and the expression level was significantly higher than that in non-cancerous tissue.

The presence of MRP3-recognizing CTL has been reported in lung, colon, bladder, and renal cancer patients [13,30]. However, to our knowledge, there is no report showing the presence of MRP3-specific CTL in HCC patients. In this study, we showed that MRP3-specific CTL could be induced by stimulating PBMCs with MRP3-derived peptides, and the induced CTL showed cytotoxicity against hepatoma cell lines overexpressing MRP3. Based on these findings, we confirmed that MRP3-specific CTLs exist in HCC patients and MRP3 serves as an immunogenic antigen in HCC.

The frequency of peripheral blood CTL specific to each MRP3 epitope was similar to the reported frequencies of CTL against other tumor antigen epitopes [9,10,31–33]. The CTLs were induced even in an early stage of HCC and regardless of HCV infection. In TILs, MRP3-specific CTL were more frequently detected, compared to that in peripheral blood, suggesting that MRP3-specific CTL are not only present in peripheral blood but also infiltrate into the tumor.

The presence and frequency of MRP3-specific CTL were also confirmed using $MRP3_{765}$ tetramer. However, MRP3-specific CTL could not be detected by ELISPOT assay in 6 patients despite a high frequency being

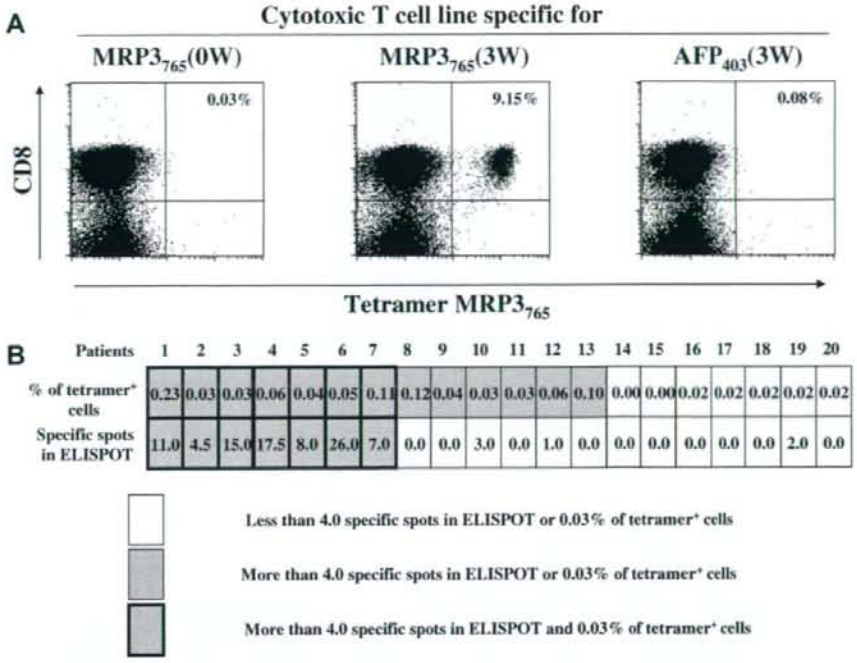

Fig. 5. Detection of MRP3-specific, HLA-A*2402-tetramer$^+$ and CD8$^+$ T lymphocytes in the peripheral blood. (A) Specificity of the MRP3$_{765}$ tetramer was confirmed by staining peptide-specific and non-specific in vitro-expanded T-cell lines. (B) Analysis of the association between the frequency of tetramer$^+$ cells and IFN-γ-producing cells detected on ELISPOT assay was performed in 20 patients.

detected by the tetramer. These findings were similar to those of hTERT-specific CTL in our previous study [10], suggesting the presence of MRP3-specific non-functional T cells in HCC patients.

In the analysis of association between the HCC tissue MRP3 expression level and MRP3-specific T-cell frequency in peripheral blood, a negative correlation was detected, suggesting that MRP3-specific immune responses exert an immune pressure on MRP3-expressing HCC cells. Recent studies have shown the involvement of MRP3 in the resistance to anti-tumor drugs and poor prognosis in several cancer patients [14,15,28,29,34,35]. Taken together with these reports, our results suggest the possibility that MRP3-targeting immunotherapy not only simply eliminates cancers but also improves drug resistance and the prognosis by inhibiting MRP3 expression in cancer cells.

Further to evaluate the usefulness of MRP3 in HCC immunotherapy, we also investigated the association between HCC treatment and the MRP3-specific CTL frequency. As we and other groups previously reported [9,26,27], the MRP3-specific CTL frequency was increased after treatment in 8 of the 12 patients in whom no immune response to MRP3 was detected before treatment, whereas the HIV env$_{584}$- and CMVpp65$_{328}$-specific CTL frequencies were not increased, excluding one patient, suggesting that this phenomenon represents the enhancement of MRP3-specific immune responses. These findings also confirmed that MRP3 is an antigen expressed in HCC tis-

Table 3
Univariate analysis of the effect of variables on the T cell response against MRP3

	Patients with positive T cell response	Patients without positive T cell response	*p*-value[a]
No. of patients	38	65	
Age (years)[b]	61.4 ± 10.0	64.3 ± 9.7	NS
Sex (M/F)	30/8	49/16	NS
AFP level (≤20/>20)	17/21	21/44	NS
Diff. degree of HCC (well/moderate or poor/ND)[c]	5/14/19	12/21/32	NS
Tumor multiplicity (multiple/solitary)	30/8	43/22	NS
Vascular invasion (+/−)	12/26	18/47	NS
TNM factor			
(T1/T2–4)	6/32	18/47	NS
(N0/N1)	36/2	64/1	NS
(M0/M1)	36/2	57/8	NS
TNM stage (I/II–IV)	6/32	18/47	NS
Histology of non-tumor liver (LC/chronic hepatitis)	30/8	55/10	NS
Liver function (Child A/B/C)	23/14/1	38/25/2	NS
Etiology (HCV/HBV/others)	26/7/5	49/12/4	NS

[a] NS, not significant.
[b] Data are expressed as means ± SD.
[c] ND, not determined.

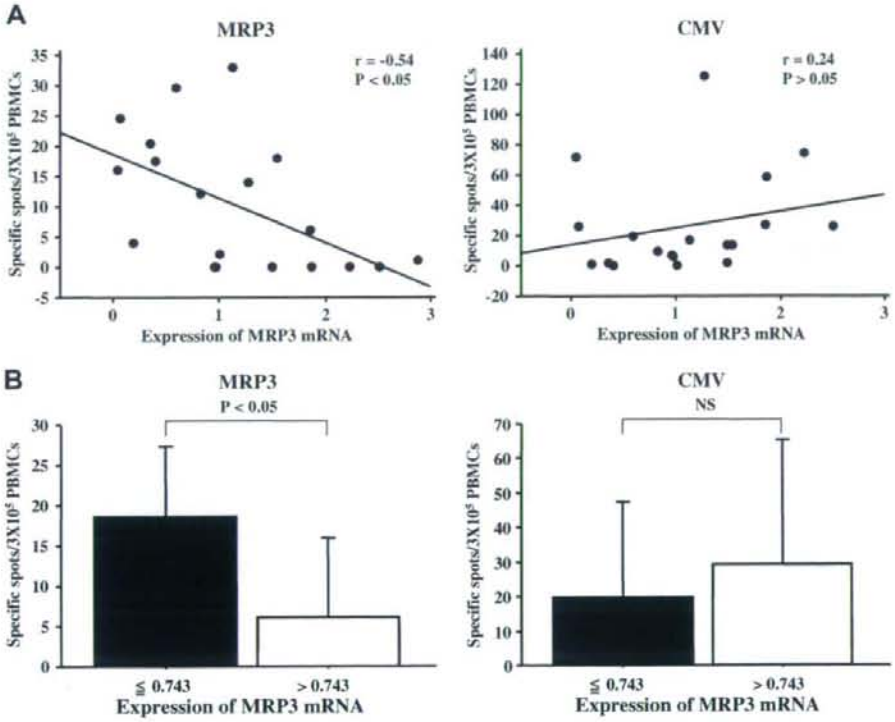

Fig. 6. Analysis of the association between the frequencies of MRP3- specific T cells detected on ELSPOT assay and the expression levels of MRP3 mRNA in HCC tissues. The frequency of MRP3-specific T cells was calculated by the sum of specific spots against $MRP3_{503}$, $MRP3_{692}$, and $MRP3_{765}$ peptides. (A) Linear regression lines for the relationship between the expression of MRP3 mRNA and the frequency of MRP3- or CMVpp65-specific T cells were calculated using Pearson's correlation coefficient. (B) Analysis of the frequency of MRP3- or CMVpp65-specific T cells in patients with low and high expression levels of MRP3 mRNA in HCC tissues.

sue, and has strong immunogenicity that readily induces CTL in vivo.

In conclusion, our study demonstrates that MRP3 is a potential candidate for tumor antigen with strong immunogenicity in HCC immunotherapy.

Acknowledgements

The authors thank Maki Kawamura, Kazumi Fushimi, Nami Nishiyama and Mikiko Nakamura for technical assistance.

Table 4
T cell response to MRP3-derived peptides by ELISPOT assay before and after treatment

	Treatment[a]	Before treatment					After treatment				
		$MRP3_{503}$	$MRP3_{692}$	$MRP3_{765}$	$HIVenv_{584}$	$CMVpp65_{328}$	$MRP3_{503}$	$MRP3_{692}$	$MRP3_{765}$	$HIVenv_{584}$	$CMVpp65_{328}$
Patient 1	TAE + RF	1	0	0	0	9	**5**	5	3	2	6
Patient 2	TAE + RF	0	0	0	0	0	**2**	2	0	0	2
Patient 3	TAE + RF	0	2	0	4	5	**8**	1	**6**	2	**13**
Patient 4	TAE + RF	0	0	0	0	**16**	0	0	0	0	**46**
Patient 5	TAE + RF	0	4	0	1	3	**5**	0	**37**	0	ND[b]
Patient 6	TAE + RF	0	0	0	0	**61**	0	0	4	1	**59**
Patient 7	TAE	0	0	0	0	**92**	0	0	**6**	0	**129**
Patient 8	TAE	0	1	0	1	9	0	1	0	0	0
Patient 9	RF	0	0	0	0	**74**	**6**	7	2.5	2.5	**65**
Patient 10	RF	0	0	0	0.5	0	**3**	0	1	0.5	9.5
Patient 11	RF	1.5	0	1	1	9.5	0	3	**9.5**	1	5.5
Patient 12	RF	0	0	0	0	**13**	0	0	0	0	6

Bold and underlined letters indicate a significant increase as described in materials and methods.

[a] TAE, transcatheter arterial embolization; RF, radiofrequency ablation.

[b] ND: not determined.

References

[1] Llovet JM, Burroughs A, Bruix J. Hepatocellular carcinoma. Lancet 2003;362:1907–1917.

[2] Lin SM, Lin CJ, Lin CC, Hsu CW, Chen YC. Radiofrequency ablation improves prognosis compared with ethanol injection for hepatocellular carcinoma < or =4 cm. Gastroenterology 2004;127:1714–1723.

[3] Ercolani G, Grazi GL, Ravaioli M, Del Gaudio M, Gardini A, Cescon M, et al. Liver resection for hepatocellular carcinoma on cirrhosis: univariate and multivariate analysis of risk factors for intrahepatic recurrence. Ann Surg 2003;237:536–543.

[4] Omata M, Tateishi R, Yoshida H, Shiina S. Treatment of hepatocellular carcinoma by percutaneous tumor ablation methods: ethanol injection therapy and radiofrequency ablation. Gastroenterology 2004;127:S159–S166.

[5] Butterfield LH, Ribas A, Meng WS, Dissette VB, Amarnani S, Vu HT, et al. T-cell responses to HLA-A*0201 immunodominant peptides derived from alpha-fetoprotein in patients with hepatocellular cancer. Clin Cancer Res 2003;9:5902–5908.

[6] Shang XY, Chen HS, Zhang HG, Pang XW, Qiao H, Peng JR, et al. The spontaneous CD8+ T-cell response to HLA-A2-restricted NY-ESO-1b peptide in hepatocellular carcinoma patients. Clin Cancer Res 2004;10:6946–6955.

[7] Zerbini A, Pilli M, Soliani P, Ziegler S, Pelosi G, Orlandini A, et al. Ex vivo characterization of tumor-derived melanoma antigen encoding gene-specific CD8+cells in patients with hepatocellular carcinoma. J Hepatol 2004;40:102–109.

[8] Komori H, Nakatsura T, Senju S, Yoshitake Y, Motomura Y, Ikuta Y, et al. Identification of HLA-A2- or HLA-A24-restricted CTL epitopes possibly useful for glypican-3-specific immunotherapy of hepatocellular carcinoma. Clin Cancer Res 2006;12:2689–2697.

[9] Mizukoshi E, Nakamoto Y, Tsuji H, Yamashita T, Kaneko S. Identification of alpha-fetoprotein-derived peptides recognized by cytotoxic T lymphocytes in HLA-A24+ patients with hepatocellular carcinoma. Int J Cancer 2006;118:1194–1204.

[10] Mizukoshi E, Nakamoto Y, Marukawa Y, Arai K, Yamashita T, Tsuji H, et al. Cytotoxic T cell responses to human telomerase reverse transcriptase in patients with hepatocellular carcinoma. Hepatology 2006;43:1284–1294.

[11] Borst P, Elferink RO. Mammalian ABC transporters in health and disease. Annu Rev Biochem 2002;71:537–592.

[12] Kiuchi Y, Suzuki H, Hirohashi T, Tyson CA, Sugiyama Y. cDNA cloning and inducible expression of human multidrug resistance associated protein 3 (MRP3). FEBS Lett 1998;433:149–152.

[13] Yamada A, Kawano K, Koga M, Matsumoto T, Itoh K. Multidrug resistance-associated protein 3 is a tumor rejection antigen recognized by HLA-A2402-restricted cytotoxic T lymphocytes. Cancer Res 2001;61:6459–6466.

[14] Tada Y, Wada M, Migita T, Nagayama J, Hinoshita E, Mochida Y, et al. Increased expression of multidrug resistance-associated proteins in bladder cancer during clinical course and drug resistance to doxorubicin. Int J Cancer 2002;98:630–635.

[15] Young LC, Campling BG, Cole SP, Deeley RG, Gerlach JH. Multidrug resistance proteins MRP3, MRP1, and MRP2 in lung cancer: correlation of protein levels with drug response and messenger RNA levels. Clin Cancer Res 2001;7:1798–1804.

[16] Nies AT, Konig J, Pfannschmidt M, Klar E, Hofmann WJ, Keppler D. Expression of the multidrug resistance proteins MRP2 and MRP3 in human hepatocellular carcinoma. Int J Cancer 2001;94:492–499.

[17] Zollner G, Wagner M, Fickert P, Silbert D, Fuchsbichler A, Zatloukal K, et al. Hepatobiliary transporter expression in human hepatocellular carcinoma. Liver Int 2005;25:367–379.

[18] Araki T, Itai Y, Furui S, Tasaka A. Dynamic CT densitometry of hepatic tumors. AJR Am J Roentgenol 1980;135:1037–1043.

[19] Liver cancer study group of Japan. General rules for the clinical and pathological study of primary liver cancer. Second English Edition. Kanehara & Co., Ltd., Tokyo. 2003.

[20] Desmet VJ, Gerber M, Hoofnagle JH, Manns M, Scheuer PJ. Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology 1994;19:1513–1520.

[21] Oiso M, Eura M, Katsura F, Takiguchi M, Sobao Y, Masuyama K, et al. A newly identified MAGE-3-derived epitope recognized by HLA-A24-restricted cytotoxic T lymphocytes. Int J Cancer 1999;81:387–394.

[22] Honda M, Yamashita T, Ueda T, Takatori H, Nishino R, Kaneko S. Different signaling pathways in the livers of patients with chronic hepatitis B or chronic hepatitis C. Hepatology 2006;44:1122–1138.

[23] Mizukoshi E, Nascimbeni M, Blaustein JB, Mihalik K, Rice CM, Liang TJ, et al. Molecular and immunological significance of chimpanzee major histocompatibility complex haplotypes for hepatitis C virus immune response and vaccination studies. J Virol 2002;76:6093–6103.

[24] Ikeda-Moore Y, Tomiyama H, Miwa K, Oka S, Iwamoto A, Kaneko Y, et al. Identification and characterization of multiple HLA-A24-restricted HIV-1 CTL epitopes: strong epitopes are derived from V regions of HIV-1. J Immunol 1997;159:6242–6252.

[25] Kuzushima K, Hayashi N, Kimura H, Tsurumi T. Efficient identification of HLA-A*2402-restricted cytomegalovirus-specific CD8(+) T-cell epitopes by a computer algorithm and an enzyme-linked immunospot assay. Blood 2001;98:1872–1881.

[26] Zerbini A, Pilli M, Penna A, Pelosi G, Schianchi C, Molinari A, et al. Radiofrequency thermal ablation of hepatocellular carcinoma liver nodules can activate and enhance tumor-specific T-cell responses. Cancer Res 2006;66:1139–1146.

[27] Ayaru L, Pereira SP, Alisa A, Pathan AA, Williams R, Davidson B, et al. Unmasking of alpha-fetoprotein-specific CD4(+) T cell responses in hepatocellular carcinoma patients undergoing embolization. J Immunol 2007;178:1914–1922.

[28] Konig J, Hartel M, Nies AT, Martignoni ME, Guo J, Buchler MW, et al. Expression and localization of human multidrug resistance protein (ABCC) family members in pancreatic carcinoma. Int J Cancer 2005;115:359–367.

[29] Steinbach D, Wittig S, Cario G, Viehmann S, Mueller A, Gruhn B, et al. The multidrug resistance-associated protein 3 (MRP3) is associated with a poor outcome in childhood ALL and may account for the worse prognosis in male patients and T-cell immunophenotype. Blood 2003;102:4493–4498.

[30] Komohara Y, Harada M, Arima Y, Suekane S, Noguchi M, Yamada A, et al. Anti-cancer vaccine candidates in specific immunotherapy for bladder carcinoma. Int J Oncol 2006;29:1555–1560.

[31] Nagorsen D, Keilholz U, Rivoltini L, Schmittel A, Letsch A, Asemissen AM, et al. Natural T-cell response against MHC class I epitopes of epithelial cell adhesion molecule, her-2/neu, and carcinoembryonic antigen in patients with colorectal cancer. Cancer Res 2000;60:4850–4854.

[32] Griffioen M, Borghi M, Schrier PI, Osanto S. Detection and quantification of CD8(+) T cells specific for HLA-A*0201-binding melanoma and viral peptides by the IFN-gamma-ELISPOT assay. Int J Cancer 2001;93:549–555.

[33] Rentzsch C, Kayser S, Stumm S, Watermann I, Walter S, Stevanovic S, et al. Evaluation of pre-existent immunity in patients with primary breast cancer: molecular and cellular assays to quantify antigen-specific T lymphocytes in peripheral blood mononuclear cells. Clin Cancer Res 2003;9:4376–4386.

[34] Oguri T, Isobe T, Fujitaka K, Ishikawa N, Kohno N. Association between expression of the MRP3 gene and exposure to platinum drugs in lung cancer. Int J Cancer 2001;93:584–589.

[35] Steinbach D, Lengemann J, Voigt A, Hermann J, Zintl F, Sauerbrey A. Response to chemotherapy and expression of the genes encoding the multidrug resistance-associated proteins MRP2, MRP3, MRP4, MRP5, and SMRP in childhood acute myeloid leukemia. Clin Cancer Res 2003;9:1083–1086.

Biochimica et Biophysica Acta 1784 (2008) 764–772

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Protein expression profile characteristic to hepatocellular carcinoma revealed by 2D-DIGE with supervised learning

Reiji Teramoto [a,*], Hirotaka Minagawa [b], Masao Honda [c], Kenji Miyazaki [a], Yo Tabuse [a], Ken'ichi Kamijo [b], Teruyuki Ueda [c], Shuichi Kaneko [c]

[a] Bio-IT Center, NEC Corporation, 34, Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan
[b] Nano Electronics Research Laboratories, NEC Corporation, 34, Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan
[c] Department of Gastroenterology, Graduate School of Medicine, Kanazawa University, 13-1 Takara-machi, Kanazawa, 920-8641, Japan

ARTICLE INFO

Article history:
Received 28 September 2007
Received in revised form 7 January 2008
Accepted 20 February 2008
Available online 5 March 2008

Keywords:
Hepatocelluar carcinoma
Two-dimensional difference gel electrophoresis
Supervised learning
Protein profiling
Feature selection

ABSTRACT

Hepatocellular carcinoma (HCC) is one of the most common and aggressive human malignancies. Although several major risks related to HCC, e.g., hepatitis B and/or hepatitis C virus infection, aflatoxin B1 exposure, alcohol drinking and genetic defects have been revealed, the molecular mechanisms leading to the initiation and progression of HCC have not been clarified. To reduce the mortality and improve the effectiveness of therapy, it is important to detect the proteins which are associated with tumor progression and may be useful as potential therapeutic or diagnosis targets. However, previous studies have not yet revealed the associations among HCC cells, histological grade and AFP. Here, we performed two-dimensional difference gel electrophoresis (2D-DIGE) combined with MS for 18 HCC patients. To focus not on individual proteins but on multiple proteins associated with pathogenesis, we introduce the supervised feature selection based on stochastic gradient boosting (SGB) for identifying protein spots that discriminate HCC/non HCC, histological grade of moderate/well and high α-fetoprotein (AFP)/low AFP level without arbitrariness. We detected 18, 25 and 27 protein spots associated with HCC, histological grade and AFP level, respectively. We confirmed that SGB is able to identify the known HCC-related proteins, e.g., heat shock proteins, carbonic anhydrase 2. Moreover, we identified the differentially expressed proteins associated with histological grade of HCC and AFP level and found that aldo-keto reductase 1B10 (AKR1B10) is related to well differentiated HCC, keratin 8 (KRT8) is related to both histological grade and AFP level and protein disulfide isomerase-associated 3 (PDIA3) is associated with both HCC and AFP level. Our pilot study provides new insights on understanding the pathogenesis of HCC, histological grade and AFP level.

1. Introduction

Hepatocellular carcinoma (HCC) is a common and aggressive malignant tumor with especially high prevalence in Asia and relatively low prevalence in Europe and North America [1,2]. Recently, the incidence of HCC in the US and UK has increased substantially over the last two decades [3,4]. Although several major risks related to HCC, e.g., hepatitis B and/or hepatitis C virus infection, aflatoxin B1 exposure, alcohol drinking and genetic defects have been revealed [5], the molecular mechanisms leading to the initiation and progression of HCC are not well-known, because most HCC patients are diagnosed at an advanced stage. To improve survival, further investigations of HCC progression markers and the mechanisms of HCC are mandatory.

Over the past decade, there has been significant progress in the development of systematic approaches to the study of HCC at the genomic and proteomic level. Several studies have employed a cDNA microarray to identify unique gene expression signatures that are associated with HCC [6–9], and it has yielded some potential HCC markers, i.e., osteopontin [10]. Some earlier proteomic HCC studies used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)/post-staining and mass spectrometry to study differentially expressed proteins in cell lines [11,12] and liver tumor tissues [13–15]. Among them, aldehydrogenase isozymes are suggested to be closely correlated to HCC and lamin B1 was identified as a marker for cirrhosis [13,14] However, previous studies have not yet investigated the associations of protein expression profiles with HCC/non HCC, histological grade and AFP level.

Two-dimensional difference gel electrophoresis (2D-DIGE), developed by Unlu et al. [16], is a modification of traditional 2D technology in which multiple protein samples are pre-labeled with different fluorescence dyes, mixed together and run on a 2D gel [17–19] and becomes one of the major advances in quantitative proteomics. In 2D-DIGE, different samples pre-labeled with mass- and charge-matched fluorescent cyanine dyes, Cy3 and Cy5, are separated on the same 2D gel together with the internal standard prepared by mixing equal

* Corresponding author.
E-mail address: r-teramoto@bq.jp.nec.com (R. Teramoto).

1570-9639/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2008.02.011

Table 1
Characteristics of patients used in this study

Patient number	Sex[a]	Age	Histology[b]	AFP(ng/ml)
1	M	51	Moderate	576
2	M	77	Moderate	27
3	M	64	Moderate	135
4	M	65	Well	416
5	M	48	Moderate	89
6	F	69	Moderate	93
7	F	66	Well	NA
8	M	45	Well	<10
9	F	75	Well	<10
10	M	46	Moderate	287
11	M	66	Well	<10
12	M	75	Moderate	<10
13	F	67	Well	<10
14	M	64	Moderate	78
15	M	68	Well	<10
16	M	74	Moderate	<10
17	M	75	Well	<10
18	M	57	NA	NA

[a] M, Male; F, Female.
[b] Moderate, moderate-differentiated HCC tissue; well, well-differentiated HCC tissue.

amounts of all samples and labeled with a third cyanine dye, Cy2. Coseparation of different samples on the same gel suppresses subtle changes in experimental conditions, thus enabling accurate spot detection and matching. The internal standard run on the background of all gels also facilitates gel-to-gel spot matching and allows the derivation of statistically reliable comparisons of protein abundances [20].

Previous studies have been largely focused on individual candidate genes and/or proteins using the statistical tests for univariate, e.g., *t*-test and Mann–Whitney test. Moreover, such approach may be insufficient to precisely define the genetic basis of molecular pathogenesis or identify proteins associated with clinical states, because events of molecular pathogenesis involve multiple genes and proteins. The statistical tests require null hypothesis and significant level and it is difficult to appropriately determine them a priori. Supervised feature selection has been used so far to derive a set of genes and/or proteins which discriminate two sample groups [10]. However, even if one uses it, the problem which one must determine the number of discriminative spots with arbitrary threshold remains.

To address these problems, we introduce the supervised feature selection based on stochastic gradient boosting (SGB) [21] for identifying protein spots that well discriminate HCC from non HCC, moderate grade from well grade and high AFP group from low AFP level group. SGB is a state of the art boosting machine and can perform feature selection in the process of learning without any arbitrary threshold. We apply SGB to protein expression of paired HCC and non HCC samples derived from the eighteen patients. We demonstrate that SGB is useful to identify informative proteins and provide further insights into the pathogenesis of HCC and the difference of histological grade and AFP level.

2. Materials and methods

2.1. HCC samples

Eighteen patients were admitted to the Kanazawa University Hospital and received surgical resection for treatment of a solitary HCC. We listed the characteristics of patients in Table 1. The HCC sample and adjacent non-tumor liver sample were snap frozen in liquid nitrogen, and used for 2-DE analysis. The HCC and non-tumor samples were histologically diagnosed. Serological tests of hepatitis C virus RNA by Amplicore analysis (Roche Diagnostic Systems) for 17 out of 18 samples were positive, and one sample was negative. All HCC samples were belonged to clinical stage I according to the TNM classification and most of the tumor size was less than 2 cm in diameter without any apparent vascular invasion. Histological typing of HCC was assessed accoding to Ishak et al. [22]. All strategies used for protein expression analysis were approved by the ethical committee of Kanazawa University Hospital.

2.2. Protein labeling with cyanine dye

Samples for 2-DE were homogenized with lysis buffer (7 M urea, 2 M thiourea, 4% w/v CHAPS, 0.8 μM aprotinin, 15 μM pepstatin, 0.1 mM PMSF, 0.5 mM EDTA, 30 mM Tris–HCl, pH 8.5) and then centrifuged at 13000 rpm 20 min at 4 °C. The supernatants were used as protein samples. The protein concentrations were determined with a protein assay reagent (Bio-Rad) according to the manufacture. The protein samples of tumor tissue were designated as HCC sample and the protein sample of noncancerous liver was designated as non HCC sample.

The experimental strategy is shown in Fig. 1. Since there may be bias-preferential labeling, paired samples (50 μg), HCC and non HCC regions, were reciprocally labeled with Cy3 and Cy5. The labeled samples were combined and separated on 2-DE gels together with the internal standard (IS), which was prepared by mixing equal amount of all non HCC and HCC samples, and labeled with Cy2. Labeling reactions were carried

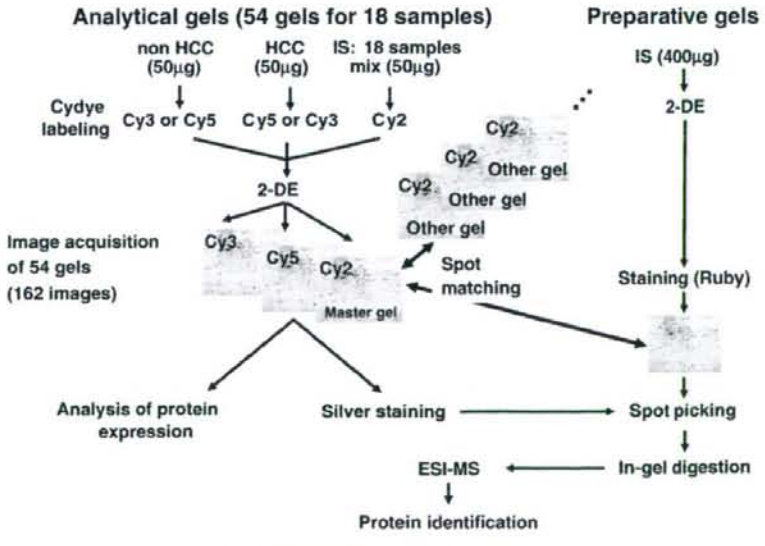

Fig. 1. The experimental strategy for 2D-DIGE.

out according to the manufacture. Each sample was labeled with 400 pmol of CyDye (GE Healthcare) on ice for 30 min in the dark and reactions were stopped by adding 1 μl of 10 mM lysine. The CyDye-labeled samples (non HCC, HCC, and IS) were mixed and left for 10 min on ice in the dark. The mixtures were added to an equal volume of the sample buffer (7 M urea, 2 M thiourea, 1% v/v IPG buffer (GE Healthcare), 2.4% v/v Destreak Reagent (GE Healthcare), 4% w/v CHAPS). Mixed samples were then adjusted to 450 μl with the rehydration buffer (7 M urea, 2 M thiourea, 0.5% v/v IPG buffer (GE Healthcare), 1.2% v/v Destreak Reagent (GE Healthcare), 4% w/v CHAPS, trace of bromophenol blue).

2.3. Analytical and preparative 2-DE for 2D-DIGE analyses

Analytical 2-DE was performed as follows: The CyDye labeled protein samples were electrophoresed in the first dimension on IPG gels (Immobiline DryStrip, GE Healthcare, pH 3-10, 24 cm) using the IPGphor system (GE Healthcare). After rehydration at 20 °C for 12 h, IEF was carried out at 500 V for 500 Vh, at 1000 V for 1000 Vh, and at 8000 V for 70,000 Vh in the dark. The gel strips were equilibrated in 12 ml of the equilibration buffer A (50 mM Tris–HCl, pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, 1% w/v DTT, trace of bromophenol blue) for 15 min with gently shaking, and then were equilibrated in 12 ml of the equilibration buffer B (50 mM Tris–HCl, pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, 2.5% w/v iodoacetamide, trace of bromophenol blue) for 15 min with gently shaking. The equilibrated strips were loaded on the top of 12.5% SDS-polyacrylamide gels (24×20 cm) and sealed with 0.5% w/v agarose. 2D separation was performed overnight at 1 W/gel using Ettan DALTII (GE Healthcare). After 2-DE, gels were scanned with a Typhoon 9410 scanner (GE Healthcare) using filters conformable to each dye's excitation and emission wavelength, and then the scanned gels were stained with Silver Staining Reagent (GE Healthcare) without glutaraldehyde for protein identification. Samples were run in triplicate to obtain statistically reasonable results (Fig. 1).

To obtain an adequate amount of the proteins from the individual spots for protein identification, 400 μg of the IS samples was separately run on 2-DE gel as described above. After 2-DE, preparative gel was fixed in 10% v/v methanol and 7% v/v acetic acid aqueous solution for 30 min and stained with a fluorescent dye, SYPRO Ruby (Invitrogen) overnight at room temperature according to the manufacturer. The gel was washed twice with 10% v/v methanol and 7% v/v acetic acid for 30 min and then washed three times with distilled water for 5 min. The washed gel was scanned with the Typhoon scanner.

2.4. Analysis of gel images

For CyDye-labeled analytical gels, a 488 nm laser and 520 nm emission filters were used for Cy2 images, a 532 nm laser and 580 nm emission filters for Cy3 images, and a 633 nm laser and 670 nm emission filters for Cy5 images, respectively. Preparative gel stained with SYPRO Ruby was scanned with a 457 nm laser and 610 nm emission filters. All gel images were obtained at 100 μm resolution and processed using ImageQuant (GE Healthcare) prior to image analysis. The analysis of gel images was performed with the DeCyder software (GE Healthcare). The Cy2, Cy3, and Cy5 images of the same gel were electrically merged and the proteins were detected as Cy3/Cy2 and Cy5/Cy2 spot pairs.

2.5. Protein identification

The protein spots were excised from the SYPRO Ruby stained gel using the Ettan spot picker (GE Healthcare) or manually excised from the silver stained analytical gels. The excised protein spots collected in 96-well plates were digested in gel with porcine trypsin (Promega) using an automatic digestion robot, ProGest (Genomic Solutions), according to the manufacture. For LC-ESI-IT MS/MS analysis using LCQ Deca XP (Thermo Electron), the digested and dried peptide samples were dissolved in 10 μl of 0.1% formic acid in 2% acetonitrile (ACN). The dissolved samples were loaded on a C18 silica gel capillary column (Magic C18, 50×0.2 mm), and the elution from the column was directly connected through a sprayer to an ESI-IT MS. The mobile phase A of LC was 2% ACN containing 0.1% formic acid, and the mobile phase B was 90% ACN containing 0.1% formic acid, respectively. A linear gradient from 5–65% of concentration B was applied to elute peptides. The ESI-IT MS was operated in positive ion mode over the range of 350–2000 (m/z) and the database search was carried out against the IPI Human using MASCOT (Matrixscience). The following search parameters were used: trypsin was used as the cutting enzyme, one missed cleavage was allowed, mass tolerance for average peptide window was set to ±1 Da, the MS/MS tolerance window was set to ±0.8 Da, and carbamidomethyl cystein and oxidized methionine were chosen as fixed and variable modifications, respectively.

2.6. Quantitative Real time detection (RTD)-PCR

Total RNA was isolated from liver samples using an RNA extraction kit(Stratagene). We reverse-transcribed 1 μg of isolated RNA to cDNA using SuperScript® II RT (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, and the resultant cDNA was amplified with appropriate TaqMan assay reagents(PE Applied Biosystems, CA). Primer pairs and probes for aldo-keto reductase family 1, member B10 (aldose reductase), argininosuccinate synthetase, fructose-1,6-bisphosphatase 1 and enoyl Coenzyme A hydratase, short chain, 1, mitochondrial were obtained from TaqMan assay reagents library.

2.7. cDNA microarray

Aliquots of total RNA (5 μg) were subjected to amplification with antisense RNA (aRNA) using a Message AmpTM aRNA kit (Ambion) as recommended by the manufacturer. The fluorescently labeled aRNA samples prepared from reference (Cy3) and test samples (Cy5) were used for microarray hybridization as described previously [23,24].

2.8. Missing spot volume estimation

Since the DeCyder software cannot find spot pairs among gels automatically and completely or spots exist under detection limit, there are many spots with missing spot volume. Since missing spot volumes are difficult to handle, we estimated the spot volumes based on k-nearest neighbor. The procedures for missing value estimation are illustrated in Fig. 2 and as follows.

Calculate Euclid distance $d_{nm} = \sqrt{\sum_{i=1}^{M} (x_{ni} - x_{mi})^2}$ between spot n and m without M gels, which are not found in both spots n and m.

Estimate the missing spot volume x_r by weighted average using spot volumes and Euclid distance d_{ri} between spot with missing spot volume and k-nearest spots. x_r is calculated by

$$x_r = \frac{\sum_{i=1}^{k} \frac{x_i}{d_{ri}}}{\sum_{i=1}^{k} \frac{1}{d_{ri}}}$$

This technique is originally developed for missing values of gene expression data obtained by DNA microarray [25]. Since existence of missing value estimation is a common problem in both DNA microarray and 2D-DIGE, we applied it to spot volume data in this study. In DNA microarray, Troyanskaya et al.reported that normalized root mean squared error between actual values and estimated values is less than 0.18 when 20 % missing values exist [25]. Hence, errors between actual values and estimated values are admissible for further analysis like DNA micorarray. The R package used for missing value estimation, impute, is available at http://cran.r-project.org. After that, for each spot pair, the normalized ratio of spot volume relative to IS (Cy3:Cy2 or Cy5:Cy2) was calculated for further analysis.

2.9. Stochastic gradient boosting (SGB)

We introduced variable importance for detection of discriminative spots using SGB [21,26]. In the following, we describe the learning algorithm for SGB and the calculation procedure for variable importance.

Assume that $X=\{x_1,\ldots, x_p\}$ is a p-dimensional vector of spot volumes associated with a sample label, which is represented $Y(X)$ and $\{T_1,\ldots, T_M\}$ is a set of M classification and regression trees and $F_M(T_1,\ldots, T_M)$ is an addictive function which combines these trees' output. In classification, we assume a two-class case, $Y(X)=\{-1, +1\}$. The discriminative function $F_M(T_1,\ldots, T_M)$ is the estimate of half of the log-odds ratio.

$$\frac{1}{2}\log\frac{\Pr\ (Y=+1|X)}{\Pr\ (Y=-1|X)}$$

where $\Pr(Y=+1|X)$ and $\Pr(Y=-1|X)$ are the probabilities of $Y(X)=+1, -1$, given X, respectively. The estimate $\hat{Y}(X)$ of the class label $Y(X)$ is calculated as

$$\hat{Y}(X) = \text{sign}(F_M(X))$$

and where the estimates of $\Pr(Y=+1|X)$ and $\Pr(Y=-1|X)$ are

$$\hat{p}r(Y=+1|X) = \frac{1}{1+\exp(-2F_M(X))}\hat{P}r(Y=-1|X) = 1-\hat{P}r(Y=+1|X)$$
$$= \frac{1}{1+\exp(2F_M(X))}$$

respectively.

Given data on a set of spot volumes (X) and tissue labels (Y) from N gels for training, $D=\{(X_1, Y_1),\ldots,(X_N, Y_N)\}$, the algorithm of SGB is as follows to build a sequence of M trees.

(1) start: $F_0=0$
(2) for $m=1$ to M do:
(3) for $n=1$ to N do:
(4) $r_n = \frac{2y_n}{1+\exp[-2y_nF_{m-1}(X_n)]}$
(5) end n
(6) $\{n_1,\ldots, n_s\}$ = *resampling_from*$\{1,\ldots,N\}$
(7) learn a regression tree T_m from $\{(r_{n_1}, X_{n_1}), \ldots, (r_{n_s}, X_{n_s})\}$
(8) $F_m(T_1(X),\ldots,T_m(x)) = F_{m-1}(T_1(X),\ldots,T_{m-1}(x)) + vT_m(x)$
(9) end m

where s and v are constant parameters, which should be determined empirically. The parameter s, $1<<s\leq N$, is the size of the random subset of values of $\{r_n\}$ used to build

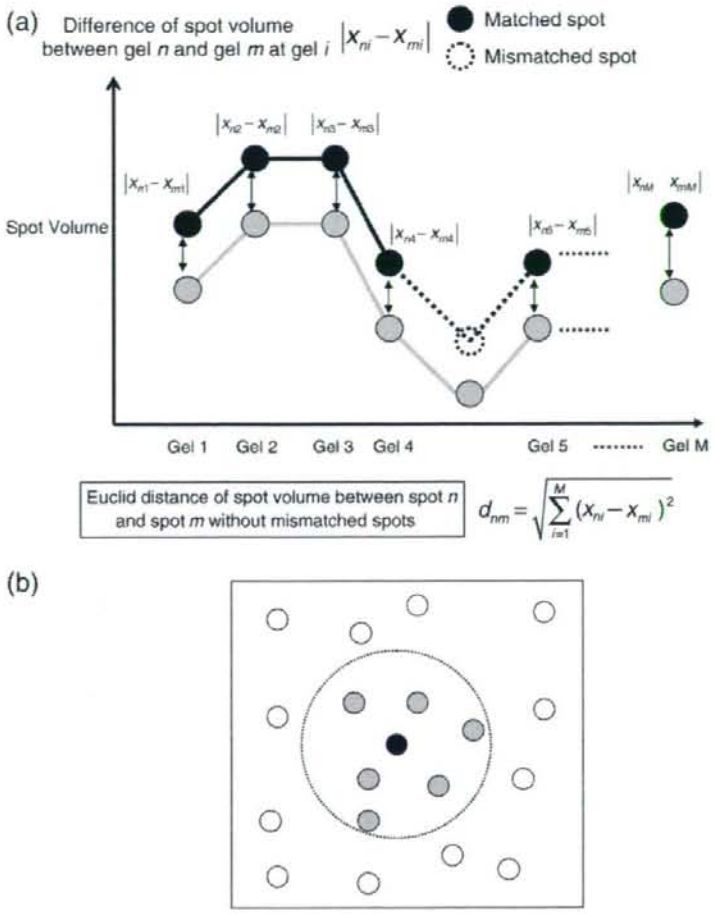

Fig. 2. Illustration of missing value estimation based on k-nearest neighbors. (a) Calculation of Euclid distance between spot n and m without mismatched spots. (b) Longitudinal axis and abscissa axis represent a vector space spanned by spot volumes. Black spot, gray spots and white spots represent missing value spot, k-nearest spots and other spots, respectively.

each tree. One will refer instead to the parameter $f=\frac{1}{2}$, the fraction of training data used by each tree. The parameter v, is the so-called shrinkage parameter introduced to avoid overfittng. The values r_n in the fourth lines of the algorithms are the gradients of the loss function $L(Y,F(X))$:

$$r_n=-\frac{\partial}{\partial F(X_n)}L\left(Y_n,F(X_n)|_{F=F_{m-1}(X_n)}\right)$$

The regression algorithm corresponds to the least squares loss function $L=(Y-F(X))^2$, and the classification algorithm employs the loss function $L=\log(1+\exp(-2YF(X)))$, which is called the binomial log-likelihood function. The loss function $L=\exp(-YF(X))$ corresponds to the loss function of Adaboost which is the most popular boosting algorithm.

2.10. Supervised feature selection with SGB

A classification and regression tree is known for its ability to select important spots among many spots. Moreover, it gives a readily interpretable model, describing the relationship between spots and predictions. Although SGB loses such interpretability, a measure of how each spots contributes to the prediction accuracy can be calculated during the course of training. According to Friedman [26], the importance of variable V_p^2 of x_p is defined as follows.

$$V_p^2=\frac{1}{M}\sum_{m=1}^{M}V_p^2(T_m)$$

where $V_p^2(T_m)$ is the importance of x_p in the construction of m-th tree. $V_p^2(T_m)$ is defined as follows.

$$V_p^2(T_m)=\sum_{t=1}^{J_m-1}\hat{\sigma}_t^2\cdot I(\text{split\_variable}(t)=p)$$

Where J_m is the number of nonterminal nodes in the m-th tree, split_variable(t) is the variable which is split on at node t and $\hat{\sigma}_t^2$ is the improvement in the mean squared error if node t is split, compared to if it were left as a terminal node. Hence, $V_p^2(T_m)$ is the sum of these improvements over all internal nodes of tree m for which variable p is the splitting variable, and V_p^2 is its mean over all trees. Once the squared importance V_p^2 of all variables is calculated, importance is taken the squared root of them and then is normalized by average of them. According to the definition of importance, V_p^2 is equal to zero when a variable is not used for splitting. This means that there is no contribution of

Table 2
The number of labeled gels

(A)			
	HCC	Non HCC	Total
Number	54	54	108
(B)			
	Well	Moderate	Total
Number	24	27	51
(C)			
	AFP < 10(ng/ml)	AFP < 10(ng/ml)	Total
Number	24	24	48

Three gels are obtained from one tissue. (A) HCC, (B) histological grade, (C) AFP level.

Table 3
The number of training sets and test sets

(A)		
	Training	Test
Non HCC	43	11
HCC	41	13
(B)		
	Training	Test
Well	16	8
Moderate	20	7
(C)		
	Training	Test
AFP<10(ng/ml)	20	4
AFP>10(ng/ml)	14	10

(A) HCC, (B) histological grade, (C) AFP level.

a variable to class discrimination. Thus, we can eliminate irrelevant variables without any threshold. The R package used for stochastic gradient boosting, gbm, is available at http://cran.r-project.org.

3. Results and discussions

3.1. Proteomic profiling of non HCC and HCC tissues using 2D-DIGE

Labeled protein samples were separated on 2-DE, and fluorescence images were obtained. In parallel to the analytical gels, samples prepared from non HCC and HCC tissues were separately run on the preparative gels for protein identification. The proteins were visualized with SYPRO Ruby, and images were directly matched to CyDye images of analytical gels. Since 184 spots (125 proteins) consistently expressed in at least 70% of the all gels, we identified them as the proteins with consolidated expression intensities. Note that histological grade of moderate/well and high/low AFP level gels are derived from HCC samples only. Although we detected thousands spots on 2D gels, most of them are remained as unidentified. To enable us to interpret biologically, we analyzed identified spots only.

As mentioned in section 2.2, since the non HCC sample was labeled with Cy3 (or Cy5) dye, and the HCC sample was labeled with Cy5 (or Cy3) dye, we obtained three gels from one tissue sample. The numbers of HCC and non HCC gels, histological grade of moderate/well ones and high/low AFP ones are shown in Table 2.

3.2. Detection of proteins associated with HCC, histological grade and AFP level

To detect of the proteins associated with HCC, histological grade and AFP level, we performed supervised feature selection as described in section 2.10. To validate discriminative ability of selected spots, we randomly divided all gels to training sets and test sets. The number of training sets and test sets are summarized in Table 3. Since the number of HCC and non HCC gels was large, the rate of training sets and test sets was fixed about 4:1. On the other hand, since the number of images used to histological grade and AFP level was relatively small, the rate of training sets and test sets was set about 3:1 in cell differentiation and AFP level. We employed training sets for feature selection and constructing the predictive model. To detect spots associated with histological grade and AFP level, we used only HCC samples. The overall procedure is illustrated in Fig. 3. According to the previous study [21], we employed the parameters f, and M of SGB as 0.9, 0.01 and 1000, respectively. Table 4 show the selected spots associated with HCC, histological grade of moderate/well and AFP level. Other annotation data for these spots in Table S1 (Supplementary material). The number of selected spots associated with HCC, histological grade and AFP level is 18, 25 and 27, respectively. Although there are spots whose ratios are close to one, they are informative spots, because SGB is able to detect them as discriminative spots unlike statistical test for univariate, e.g., t-test. The positions of selected spots in the preparative 2D gel are shown in Fig. 4.

3.3. Evaluation of classification performance

We evaluated the classification performance by the selected spots on the basis of test sets according to the procedure illustrated in Fig. 3. This evaluation procedure is reasonable in terms of supervised prediction as discussed by Dupuy and Simon [27]. The performance was evaluated based on accuracy. Accuracy is defined as follows.

$$\text{accuracy} = \frac{\text{number of correctly predicted samples}}{\text{number of all samples}}$$

We constructed the predictive models using selected spots and training sets and applied to test sets as shown in Fig. 3. Table 5 shows the classification performance of SGB for HCC/non HCC, histological grade of moderate/well and high AFP (AFP> 10(ng/ml))/low AFP (AFP< 10(ng/ml)) level. From Table 5, SGB is able to classify both training sets

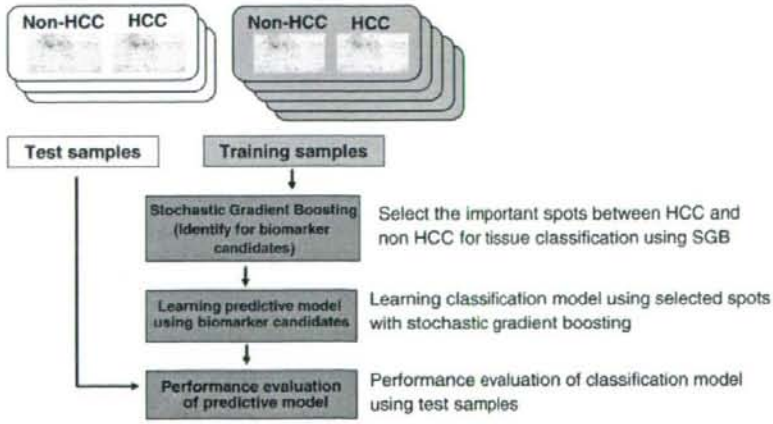

Fig. 3 Illustration of overall procedure for performance evaluation for HCC/non HCC classification. In histological grade of moderate/well grade and high/low AFP level, the same procedure is employed.