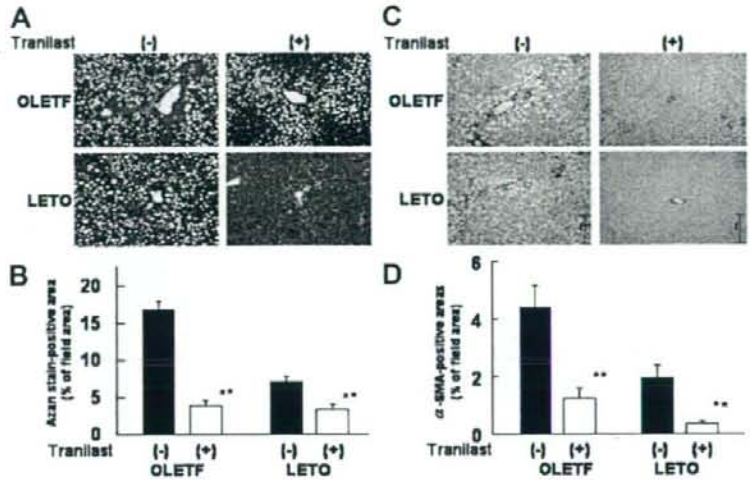

Fig. 2. Tranilast inhibited activation of stellate cells and development of hepatic fibrosis. (A) Paraffin-embedded sections of the liver were stained with Azan stain (original magnification, ×100). (B) Percentages of fibrous area (blue area in Azan stain) were calculated under a microscope using an image analysis system. (C) Immunohistochemical staining using a monoclonal mouse antihuman α-SMA antibody (original magnification, ×100). (D) The α-SMA–positive area was quantified morphometrically in the liver sections. Compared with the MCD diet alone, the tranilast-mixed diet significantly improved liver fibrosis and activation of stellate cells. Values are means ± SEM. $^{**}P < 0.01$ versus the MCD diet.

fibrosis in rats, the expression of TGF-β, a master regulator of fibrogenesis in the liver, was estimated by quantitative real-time PCR (Fig. 4A). Treatment with tranilast down-regulated mRNA for TGF-β in the liver of OLETF-MCD rats, and also had a tendency to down-regulate expression in the liver of LETO-MCD rats. Moreover, tranilast suppressed the expression of TGF-β target genes involved in fibrogenesis, collagen I, and PAI-1 (Fig. 4B, C). Consistent with the findings for mRNA, immunohistochemical analysis showed the suppression of TGF-β protein expression in the liver of tranilast-treated rats (Fig. 4D).

Tranilast Up-Regulated mRNA Expression of Genes Involved in Fatty Acid β-Oxidation. To clarify the mechanism by which tranilast improves hepatic steatosis in the rats, the expression of genes involved in fatty acid β-oxidation was measured by quantitative real-time PCR (Fig. 5A, B). Messenger RNA expression for PPAR-α, a transcriptional factor central to the regulation of β-oxidation, was significantly up-regulated by tranilast in the liver of OLETF-MCD and LETO-MCD rats. In addition, tranilast also induced mRNA expression for CPT-1, a rate-limiting enzyme for β-oxidation that controls fatty acid entry into mitochondria.

Tranilast Reduced the Number of Kupffer Cells and TNF-α Production. To clarify the mechanism by which tranilast improved hepatic inflammation in the rats, we evaluated the effect of tranilast on Kupffer cells, the resident liver macrophages. We performed an immunohistochemical analysis using a monoclonal antibody specific for the ED1 antigen, a marker of rat monocytes/macrophages and Kupffer cells. Although ED1-positive cells significantly increased in the liver of OLETF-MCD rats compared with LETO-MCD rats, tranilast reduced the number of these cells in both LETO-MCD and OLETF-MCD rats (Fig. 6A, B). Because Kupffer cells are the primary source of hepatic TNF-α,[30] we measured hepatic TNF-α mRNA levels. Real-time PCR analysis showed that tranilast dramatically down-regulated TNF-α mRNA (Fig. 6C). Serum levels of TNF-α also

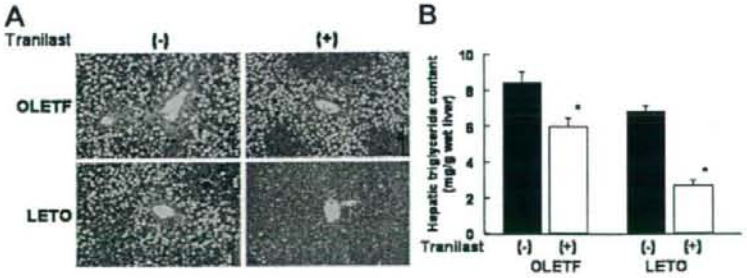

Fig. 3. Tranilast prevented development of hepatic steatosis. (A) Paraffin-embedded sections of the liver were stained with hematoxylin-eosin (original magnification, ×100). (B) Tranilast significantly reduced hepatic triglyceride content in these two groups of rats. Values are means ± SEM. $^{*}P < 0.05$ versus the MCD diet.

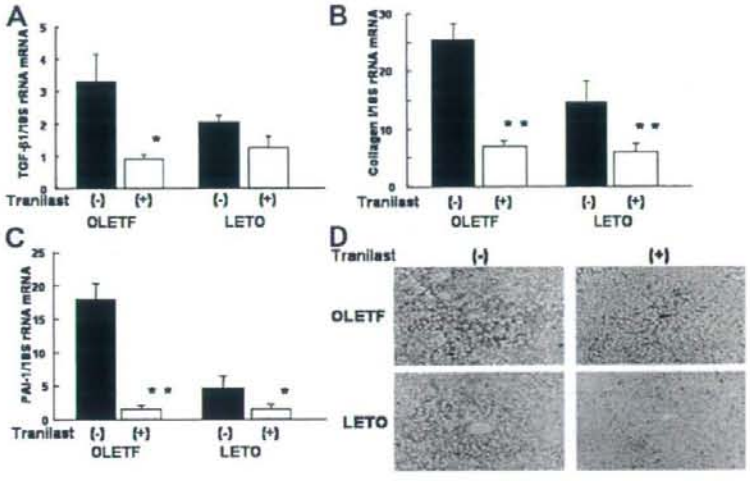

Fig. 4. Tranilast suppressed hepatic expression of TGF-β. (A) Treatment with tranilast down-regulated mRNA for TGF-β in the liver of OLETF-MCD rats. (B, C) Tranilast suppressed expression of TGF-β target genes, genes for procollagen, and PAI-1. Values are means ± SEM. $*P < 0.05$ versus MCD diet; $**P < 0.01$ versus MCD diet. (D) Paraffin-embedded sections of the liver were immunostained with a TGF-β antibody. Immunohistochemical analysis showed that TGF-β expression was down-regulated in the liver of tranilast-treated rats (original magnification ×100).

decreased after treatment with tranilast (Table 1). In the livers of OLETF-MCD rats, tranilast down-regulated the expression of VCAM-1 mRNA, an adhesion molecule that facilitates leukocyte recruitment; a similar trend was also observed in the livers of LETO-MCD rats (Fig. 6D).

Tranilast Inhibited LPS-Induced TNF-α and IL-6 Production in RAW 264.7 Macrophages. To determine whether tranilast has a direct effect on activated macrophages, we performed several experiments using RAW 264.7 cells. The incubation of RAW 264.7 cells with LPS for 6 hours resulted in a marked increase in TNF-α, IL-6, and iNOS mRNA expression (Fig. 7A-C). The concomitant administration of tranilast significantly suppressed LPS-induced expression of these genes. In addition, tranilast treatment strongly inhibited the release of TNF-α and IL-6 from RAW 264.7 cells into the culture medium (Fig. 7D, E).

Tranilast Attenuated Intrahepatic Oxidative Stress. To clarify the antioxidative effect of tranilast, we used Western blot analysis to examine the expression of 4-hydroxynonenal (4-HNE)–modified proteins, which are indicators of oxidative stress, in liver lysates. Tranilast administration significantly decreased 4-HNE–modified proteins in the livers of OLETF-MCD and LETO-MCD rats (Fig. 8A-D). Similarly, Western blot analysis showed that transilast treatment significantly reduced the pres-

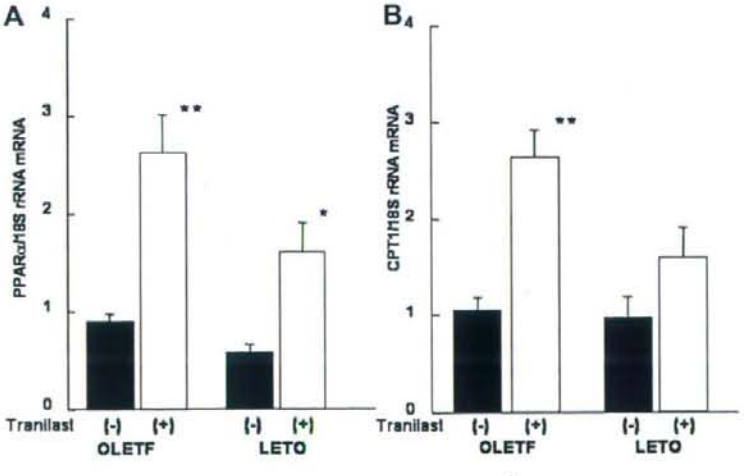

Fig. 5. Tranilast up-regulated the genes involved in fatty acid β-oxidation. (A) Messenger RNA expression for PPAR-α, a transcriptional factor central to the regulation of β-oxidation, was significantly up-regulated by tranilast in the liver of OLETF-MCD and LETO-MCD rats. (B) Tranilast induced mRNA expression for CPT-1, a rate-limiting enzyme for β-oxidation that controls fatty acid entry into mitochondria. Values are means ± SEM. $*P < 0.05$, $**P < 0.01$ versus the MCD diet.

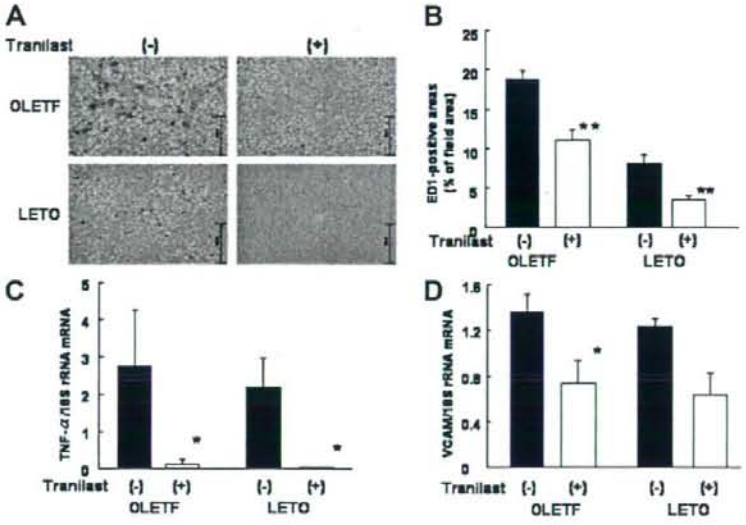

Fig. 6. Tranilast inhibited the intrahepatic recruitment of Kupffer cells and the production of TNF-α in a dietary animal model of NASH. (A) Immunohistochemical staining using an ED1 antibody (original magnification ×100). (B) The ED1-positive area was quantified morphometrically in liver sections. Compared with the MCD diet alone, the tranilast-mixed diet significantly inhibited the permeation of macrophages in the liver. Values are means ± SEM. $^{**}P < 0.01$ versus the MCD diet. (C) Tranilast suppressed TNF-α mRNA expression. Values are means ± SEM. $^{*}P < 0.05$ versus the MCD diet. (D) Tranilast significantly down-regulated the expression of vascular cell adhesion molecule mRNA, an adhesion molecule (VCAM), in the livers of OLETF-MCD rats.

ence of thiobarbituric acid-reactive substances, another marker of oxidative stress (data not shown).

Discussion

Hepatic fibrosis is a well-known histological and biochemical hallmark of NASH. Progressive fibrosis causes liver insufficiency and portal hypertension and is a risk factor for developing hepatocellular carcinoma. Thus, hepatic fibrosis is an important therapeutic target for NASH, and various inhibitors of fibrosis are now being investigated as potential candidate drugs.[31] In the current study, we demonstrated that tranilast, an antifibrogenic agent, improved hepatic fibrosis in an animal model of NASH. Administration of tranilast inhibited the activation of hepatic stellate cells and reduced the expression of genes for TGF-β and its target molecules, which probably led to the prevention of fibrosis. The current study provides *in vivo* evidence that antifibrogenic treatment by tranilast is a beneficial and promising treatment approach for NASH.

Among the numerous profibrogenic mediators, a functional hierarchy might exist, which points to TGF-β as being the most effective.[32] Concomitant with increased TGF-β activity during hepatic fibrosis, hepatic stellate cells increase the production and deposition of collagen. It is well known that the blockade of TGF-β synthesis or signaling using different experimental strategies prevents liver fibrosis in various animal models.[15-18] To date, however, it is unclear whether an oral inhibitor of TGF-β activity is effective in preventing liver fibrosis in animal models. We demonstrated that the administration of tranilast reduced TGF-β gene expression and α-SMA–positive cells in the liver of MCD diet–fed rats, suggesting that treatment with tranilast ameliorates hepatic fibrosis and liver damage through the direct suppression of TGF-β expression. These findings are supported by a previous *in vitro* study demonstrating that tranilast inhibits the activation of cultured rat stellate cells through the suppression of TGF-β mRNA.[10] In addition, tranilast also down-regulated the hepatic expression of PAI-1, one of the TGF-β target genes. PAI-1 has been reported to be a therapeutic target of metformin for hepatic injury caused by ethanol,[33] which suggests that tranilast might also be effective in treating this disease.

This study is the first to demonstrate that tranilast exerts broad and profound effects on Kupffer cells, the resident liver macrophages. Previous reports showed that tranilast inhibited cytokine release from human monocytes/macrophages[8,34]; however, it was unclear whether tranilast acted on Kupffer cells. Our experiments indicate that tranilast suppresses recruitment of Kupffer cells into the liver and inhibits the release of proinflammatory cytokines such as TNF-α from RAW 264.7 macrophages. Therefore, tranilast may attenuate MCD diet–induced hepatic inflammation through its inhibitory effects on Kupffer cells. In addition, Kupffer cells are also a source of TGF-β,[35] and tranilast is reported to inhibit the release of TGF-β from human monocytes/macrophages.[34] Thus, tranilast may suppress TGF-β production by inactivating both Kupffer cells and hepatic stellate cells.

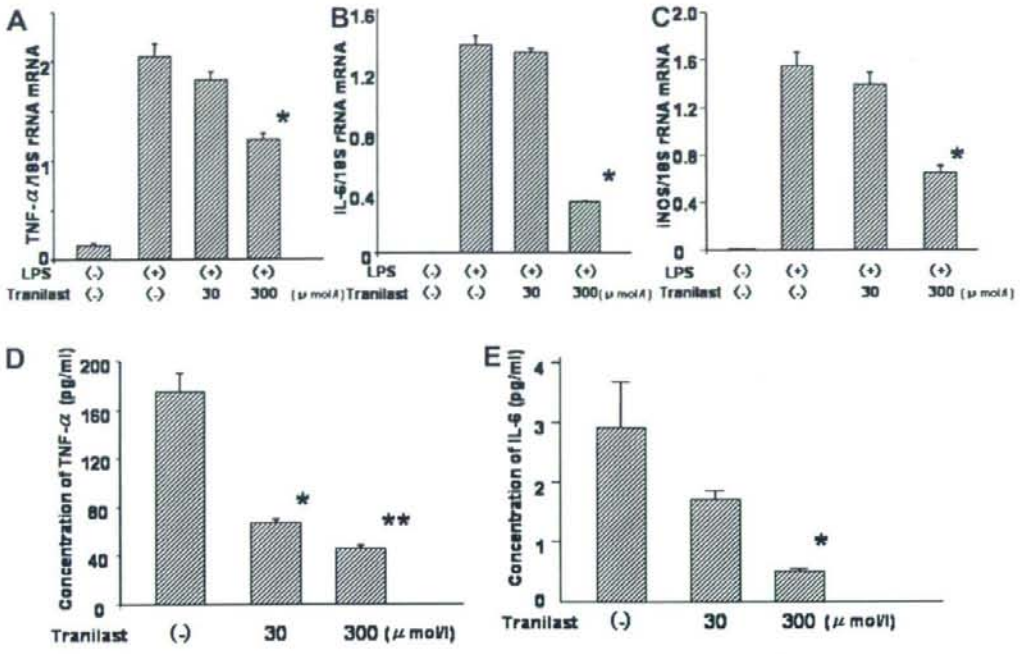

Fig. 7. Tranilast inhibited TNF-α, IL-6, and iNOS production in RAW 264.7 macrophages. (A) Tranilast significantly down-regulated TNF-α mRNA expression in RAW 264.7 cells stimulated by LPS. (B) Tranilast significantly down-regulated IL-6 expression in RAW cells. (C) Tranilast significantly down-regulated iNOS expression in RAW cells. Values are means ± SEM. $^*P < 0.05$ versus LPS(+), tranilast(−). (D) Tranilast treatment decreased the concentration of TNF-α released into the culture medium. (E) Tranilast treatment decreased the concentration of IL-6 released into the culture medium. Values are means ± SEM. $^*P < 0.05$, $^{**}P < 0.01$ versus tranilast(−).

Obesity, type 2 diabetes, and dyslipidemia frequently coexist with NAFLD. Insulin resistance is an underlying factor in these metabolic disorders and may play a critical role in the pathogenesis of NAFLD. Several lines of evidence have indicated that insulin-sensitizing agents, such as pioglitazone, improve histological findings and liver injury in both animals and patients with NASH, the progressive form of NAFLD.[6,36] In the current study, insulin sensitivity in rats was unaffected by treatment with tranilast. In addition, unlike the action of pioglitazone,[6,36] tranilast was beneficial in preventing hepatic fibrosis in the LETO-MCD rats, which mimics animal models of NASH, without insulin resistance. These results suggest that the inhibitory action of tranilast is mediated through an insulin-resistance–independent pathway. Therefore, nonresponders to insulin-sensitizing agents might benefit from treatment with tranilast for preventing hepatic fibrosis in NASH.

Surprisingly, we found that tranilast prevented the development of not only hepatic fibrosis but also liver steatosis in MCD-induced NASH. This finding was reinforced by the reduction of liver triglyceride content. These findings were unexpected because little information is available regarding the effect of tranilast on intracellular lipid accumulation. Real-time PCR revealed that the expression of genes involved in fatty acid β-oxidation, such as CPT-I and PPAR-α, were up-regulated, which suggests that tranilast-induced enhancement of β-oxidation may have attenuated hepatic steatosis in rats. Our results were consistent with a previous study showing that TGF-β signaling pathways suppress PPAR-α activity, a transcriptional factor central to the regulation of β-oxidation, and reduce fatty acid β-oxidation in cardiomyocytes.[37] Therefore, tranilast-mediated inhibition of TGF-β action may enhance β-oxidation through the up-regulation of PPAR-α. However, one limitation of our study is that the analysis was performed only at the gene transcriptional levels. Further studies evaluating changes in β-oxidation capacity are needed to clarify the mechanism by which tranilast attenuates hepatic steatosis.

In the current study, OLETF-MCD rats treated with tranilast showed significant body weight loss. Be-

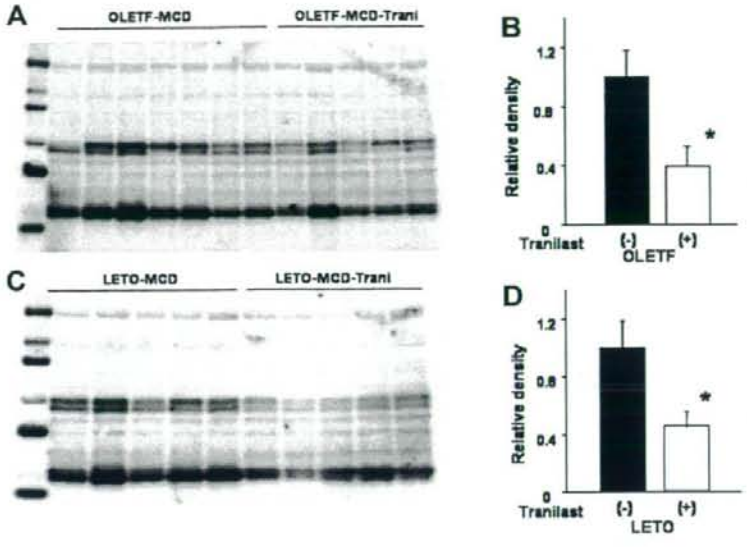

Fig. 8. Tranilast suppressed oxidative stress. (A,C) Western blot analysis of 4-HNE-modified proteins and (B,D) subsequent quantification showed that tranilast treatment significantly decreased 4-HNE in the livers of OLETF-MCD and LETO-MCD rats. Values are means ± SEM. *$P < 0.05$ versus the MCD diet.

cause tranilast significantly reduced body weight in MCD-OLETF rats but not in MCD-LETO rats, it is unlikely that this action reflects a toxic effect. Certainly, we cannot rule out the possibility that body weight reduction may at least partly contribute to the tranilast-induced attenuation of hepatic steatosis and the up-regulation of gene expression involved in fatty acid β-oxidation. However, our data demonstrated direct inhibitory effects of tranilast on activated macrophages both *in vivo* and *in vitro*, suggesting that tranilast ameliorated liver pathology independently of its weight-reducing effects. Critical parameters contributing to body weight maintenance include caloric intake, physical activity, and adaptive thermogenesis.[38,39] Tranilast treatment showed no apparent affects on food intake or physical appearance; future studies are required to examine the effect of tranilast on adaptive thermogenesis and energy expenditure. Tranilast may inhibit activated macrophages infiltrating the adipose tissues, leading to the inhibition of adipose tissue hyperplasia. These potential effects of tranilast are under investigation in MCD diet–untreated animal models of obesity and type 2 diabetes without cachexia.

Serological analysis revealed that all the animals treated with tranilast showed improvement of serum ALT levels compared with the untreated control. These findings indicate that treatment with tranilast for 8 weeks did not induce hepatic toxicity, as evidenced by the histomorphology of the liver. Tranilast has long been used safely for the treatment of patients with bronchial asthma and atopic dermatitis. However, because TGF-β–mediated fibrogenesis is involved in an active wound-healing process that is essential for tissue remodeling, long-term treatment with a TGF-β inhibitor may produce adverse effects.[31] To determine the safety of administering tranilast for patients with NASH, further investigations of long-term outcomes are required.

In conclusion, our experiments demonstrated that tranilast prevented the development of NASH in a dietary rat model, possibly through an insulin-resistance–independent pathway. Our findings suggest that targeting TGF-β with tranilast is a novel approach to the prevention and treatment of NASH.

Acknowledgment: The authors thank A. Nakano, C. Minami, Y. Hasebe, A. Katayama, H. Takayama, Y. Hashimoto, and Y. Fujita for providing technical assistance.

References

1. Reid AE. Nonalcoholic steatohepatitis. Gastroenterology 2001;121:710-723.
2. Angulo P. Nonalcoholic fatty liver disease. N Engl J Med 2002;346:1221-1231.
3. Marchesini G, Bugianesi E, Forlani G, Cerrelli F, Lenzi M, Manini R, et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. HEPATOLOGY 2003;37:917-923.
4. Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. Diabetes 2001;50:1844-1850.
5. Marceau P, Biron S, Hould FS, Marceau S, Simard S, Thung SN, et al. Liver pathology and the metabolic syndrome X in severe obesity. J Clin Endocrinol Metab 1999;84:1513-1517.

6. Ota T, Takamura T, Kurita S, Matsuzawa N, Kita Y, Uno M, et al. Insulin resistance accelerates a dietary rat model of nonalcoholic steatohepatitis. Gastroenterology 2007;132:282-293.
7. Ota T, Takamura T, Kaneko S. Pioglitazone in nonalcoholic steatohepatitis. N Engl J Med 2007;356:1068; author reply 1068-1069.
8. Suzawa H, Kikuchi S, Ichikawa K, Koda A. Inhibitory action of tranilast, an anti-allergic drug, on the release of cytokines and PGE2 from human monocytes-macrophages. Jpn J Pharmacol 1992;60:85-90.
9. Bonnet F, Cao Z, Cooper ME, Cox AJ, Kelly DJ, Gilbert RE. Tranilast attenuates vascular hypertrophy, matrix accumulation and growth factor overexpression in experimental diabetes. Diabetes Metab 2003;29:386-392.
10. Ikeda H, Inao M, Fujiwara K. Inhibitory effect of tranilast on activation and transforming growth factor beta 1 expression in cultured rat stellate cells. Biochem Biophys Res Commun 1996;227:322-327.
11. Miyazawa K, Kikuchi S, Fukuyama J, Hamano S, Ujiie A. Inhibition of PDGF- and TGF-beta 1-induced collagen synthesis, migration and proliferation by tranilast in vascular smooth muscle cells from spontaneously hypertensive rats. Atherosclerosis 1995;118:213-221.
12. Liu X, Hu H, Yin JQ. Therapeutic strategies against TGF-beta signaling pathway in hepatic fibrosis. Liver Int 2006;26:8-22.
13. Chapman HA. Disorders of lung matrix remodeling. J Clin Invest 2004; 113:148-157.
14. Casini A, Pinzani M, Milani S, Grappone C, Galli G, Jezequel AM, et al. Regulation of extracellular matrix synthesis by transforming growth factor beta 1 in human fat-storing cells. Gastroenterology 1993;105:245-253.
15. George J, Roulot D, Koteliansky VE, Bissell DM. *In vivo* inhibition of rat stellate cell activation by soluble transforming growth factor beta type II receptor: a potential new therapy for hepatic fibrosis. Proc Natl Acad Sci U S A 1999;96:12719-12724.
16. Arias M, Sauer-Lehnen S, Treptau J, Janoschek N, Theuerkauf I, Buettner R, et al. Adenoviral expression of a transforming growth factor-beta1 antisense mRNA is effective in preventing liver fibrosis in bile-duct ligated rats. BMC Gastroenterol 2003;3:29.
17. Qi Z, Atsuchi N, Ooshima A, Takeshita A, Ueno H. Blockade of type beta transforming growth factor signaling prevents liver fibrosis and dysfunction in the rat. Proc Natl Acad Sci U S A 1999;96:2345-2349.
18. Nakamura T, Sakata R, Ueno T, Sata M, Ueno H. Inhibition of transforming growth factor beta prevents progression of liver fibrosis and enhances hepatocyte regeneration in dimethylnitrosamine-treated rats. HEPATOLOGY 2000;32:247-255.
19. Williams EJ, Gaca MD, Brigstock DR, Arthur MJ, Benyon RC. Increased expression of connective tissue growth factor in fibrotic human liver and in activated hepatic stellate cells. J Hepatol 2000;32:754-761.
20. Shigeki S, Murakami T, Yata N, Ikuta Y. Treatment of keloid and hypertrophic scars by iontophoretic transdermal delivery of tranilast. Scand J Plast Reconstr Surg Hand Surg 1997;31:151-158.
21. Yamada H, Tajima S, Nishikawa T, Murad S, Pinnell SR. Tranilast, a selective inhibitor of collagen synthesis in human skin fibroblasts. J Biochem (Tokyo) 1994;116:892-897.
22. Akahori H, Ota T, Torita M, Ando H, Kaneko S, Takamura T. Tranilast prevents the progression of experimental diabetic nephropathy through suppression of enhanced extracellular matrix gene expression. J Pharmacol Exp Ther 2005;314:514-521.
23. Kawano K, Hirashima T, Mori S, Saitoh Y, Kurosumi M, Natori T. Spontaneous long-term hyperglycemic rat with diabetic complications. Otsuka Long-Evans Tokushima Fatty (OLETF) strain. Diabetes 1992;41: 1422-1428.
24. Mathews CE, Leiter EH. Rodent models for the study of diabetes. In: Joslin's diabetes mellitus, 14th ed. Boston: Lippincott Williams & Wilkins; 2005:291-328.
25. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94:2467-2474.
26. Takamura T, Sakurai M, Ota T, Ando H, Honda M, Kaneko S. Genes for systemic vascular complications are differentially expressed in the livers of type 2 diabetic patients. Diabetologia 2004;47:638-647.
27. Matsuzawa N, Takamura T, Kurita S, Misu H, Ota T, Ando H, et al. Lipid-induced oxidative stress causes steatohepatitis in mice fed an atherogenic diet. HEPATOLOGY 2007;46:1392-1403.
28. Saiura A, Sata M, Hirata Y, Nagai R, Makuuchi M. Tranilast inhibits transplant-associated coronary arteriosclerosis in a murine model of cardiac transplantation. Eur J Pharmacol 2001;433:163-168.
29. Izawa A, Suzuki J, Takahashi W, Amano J, Isobe M. Tranilast inhibits cardiac allograft vasculopathy in association with p21(Waf1/Cip1) expression on neointimal cells in murine cardiac transplantation model. Arterioscler Thromb Vasc Biol 2001;21:1172-1178.
30. Su GL. Lipopolysaccharides in liver injury: molecular mechanisms of Kupffer cell activation. Am J Physiol Gastrointest Liver Physiol 2002;283: G256-G265.
31. Gressner AM, Weiskirchen R. Modern pathogenetic concepts of liver fibrosis suggest stellate cells and TGF-beta as major players and therapeutic targets. J Cell Mol Med 2006;10:76-99.
32. Gressner AM, Weiskirchen R, Breitkopf K, Dooley S. Roles of TGF-beta in hepatic fibrosis. Front Biosci 2002;7:d793-d807.
33. Bergheim I, Guo L, Davis MA, Lambert JC, Beier JI, Duveau I, et al. Metformin prevents alcohol-induced liver injury in the mouse: critical role of plasminogen activator inhibitor-1. Gastroenterology 2006;130:2099-2112.
34. Capper EA, Roshak AK, Bolognese BJ, Podolin PL, Smith T, Dewitt DL, et al. Modulation of human monocyte activities by tranilast, SB 252218, a compound demonstrating efficacy in restenosis. J Pharmacol Exp Ther 2000;295:1061-1069.
35. Kolios G, Valatas V, Kouroumalis E. Role of Kupffer cells in the pathogenesis of liver disease. World J Gastroenterol 2006;12:7413-7420.
36. Belfort R, Harrison SA, Brown K, Darland C, Finch J, Hardies J, et al. A placebo-controlled trial of pioglitazone in subjects with nonalcoholic steatohepatitis. N Engl J Med 2006;355:2297-2307.
37. Sekiguchi K, Tian Q, Ishiyama M, Burchfield J, Gao F, Mann DL, et al. Inhibition of PPAR-alpha activity in mice with cardiac-restricted expression of tumor necrosis factor: potential role of TGF-beta/Smad3. Am J Physiol Heart Circ Physiol 2007;292:H1443-H1451.
38. Lowell BB, Spiegelman BM. Towards a molecular understanding of adaptive thermogenesis. Nature 2000;404:652-660.
39. Lin J, Wu PH, Tarr PT, Lindenberg KS, St-Pierre J, Zhang CY, et al. Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1alpha null mice. Cell 2004;119:121-135.

ELSEVIER

Journal of Hepatology 49 (2008) 207–216

Journal of Hepatology

www.elsevier.com/locate/jhep

Identification of novel candidate tumour marker genes for intrahepatic cholangiocarcinoma☆

Ryuhei Nishino[1], Masao Honda[1], Taro Yamashita[1], Hajime Takatori[1], Hiroshi Minato[2], Yoh Zen[2], Motoko Sasaki[3], Hiroyuki Takamura[4], Katsuhisa Horimoto[5], Tetsuo Ohta[4], Yasuni Nakanuma[3], Shuichi Kaneko[1,*]

[1]*Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[2]*Pathology Section, Kanazawa University Hospital, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[3]*Department of Human Pathology, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[4]*Department of Gastroenterologic Surgery, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[5]*Biological Network Team, Computational Biology Research Centre, National Institute of Advanced Industrial Science and Technology, Japan*

See Editorial, pages 160–162

***Background/Aims*: Specific markers are required for early detection and diagnosis of intrahepatic cholangiocarcinoma (ICC); however, the tumour markers currently in use are not specific for ICC.**

***Methods*: We compared an ICC cDNA library with that of hepatocellular carcinoma (HCC) by serial analysis of gene expression (SAGE). The expression patterns in each were confirmed by quantitative real-time reverse transcriptase-polymerase chain reaction (RT-PCR), immunoblotting and immunohistochemical analysis of 74 samples including 16 ICC samples.**

***Results*: A comparison of the two libraries revealed distinct gene expression patterns for each type of liver cancer. In addition to the known tumour markers, we detected nine novel genes associated with ICC. By comparing the mean transcript abundance in the ICC library with those in other libraries, including gastric, colon, prostate and breast cancer, together with our RT-PCR results, we identified three genes as specific markers of ICC: biglycan, insulin-like growth factor-binding protein 5 and claudin-4. Immunoblotting and immunohistochemical analyses showed that claudin-4 was highly expressed in ICC. Moreover, discrimination analysis revealed that a combination of these genes could be used to distinguish ICC from HCC or metastatic adenocarcinoma.**

***Conclusions*: We identified novel marker genes of ICC that are potentially useful for the diagnosis of liver cancer.**

***Keywords*: Differential diagnosis; Hepatocellular carcinoma; Intrahepatic cholangiocarcinoma; SAGE**

Received 5 October 2007; received in revised form 10 March 2008; accepted 24 March 2008; available online 5 May 2008
Associate Editor: J.M. Llovet

☆ 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 2231; fax: +81 76 234 4250.
E-mail address: skaneko@m-kanazawa.jp (S. Kaneko).

Abbreviations: ICC, intrahepatic cholangiocarcinoma; HCC, hepatocellular carcinoma; AFP, alpha-foetoprotein; KRT, keratin; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; SAGE, serial analysis of gene expression; HBV, hepatitis B virus; HCV, hepatitis C virus; NL, normal liver; CLD, chronic liver disease; RT-PCR, reverse transcriptase-polymerase chain reaction; polyA-RNA, polyadenylated RNA; mRNA, messenger RNA; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CITED4, Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4; GSTP1, glutathione-*S*-transferase pi; BGN, biglycan; IGFBP5, insulin-like growth factor-binding protein 5; CLDN4, claudin-4; PFKP, platelet type phosphofructokinase; TM4SF1, transmembrane 4 L six family member 1; CAPN1, calpain 1 (μ/I) large subunit; CLDN10, claudin-10; S100A6, S100 calcium-binding protein A6; Wnt, wingless-type MMTV integration site; TGF-beta, transforming growth factor-beta.

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.2008.03.025

1. Introduction

Intrahepatic cholangiocarcinoma (ICC), which arises from bile duct cells in the liver, currently accounts for approximately 15% of primary liver cancer [1]. Though it is less familiar than hepatocellular carcinoma (HCC), the incidence of ICC is increasing, especially in the United States, the United Kingdom and Australia [2–5].

The differential diagnosis of liver cancer depends on a combination of serological, radiological and histological examinations; however, the process can be difficult, especially in advanced stages. In addition, liver cancer may include metastasis from other organs, making the diagnosis even more difficult. Nevertheless, a correct diagnosis is essential to select the proper therapy and to determine a patient's prognosis.

Tumour markers are routinely used in the differential diagnosis of liver cancer. For example, alpha-foetoprotein (AFP) is a sensitive and specific marker of HCC [6]; keratin (KRT) 7 and KRT19 are amongst the markers of ICC. KRT7 and KRT19, which are expressed in the bile duct epithelium and ICC [7–10], can be used to distinguish HCC from ICC immunohistochemically. These are not specific to ICC, however, and are expressed in other cancers such as non-small-cell lung carcinoma [11]. Other proteins such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) have been used as serum markers for ICC [12,13], but these are also not specific to ICC because they are overexpressed in other malignant tumours. Thus, identifying ICC-specific markers will be valuable for differentially diagnosing liver cancer to characterise ICC at the molecular level and to develop improved therapies for patients with ICC. Recently, Lodi et al. cited claudin-4 (CLDN4) expression as a possible marker in differentiating biliary tract cancers from HCC [14]. However, the gene's ability to distinguish metastatic liver cancer has not been evaluated.

In recent years, serial analysis of gene expression (SAGE) [15] and DNA microarrays have been used to comprehensively analyse gene expression, including comparisons of the expression patterns in normal liver [16] and HCC [17–19]. We used SAGE to construct cDNA libraries of ICC and HCC and performed comprehensive analyses of the expression patterns in these tumours to identify novel markers of ICC.

2. Materials and methods

2.1. Tissue samples

For SAGE analysis, we prepared one ICC sample and three HCC samples (Table 1). The ICC sample was obtained by the surgical resection of a solitary cancerous lesion in the liver, which was diagnosed histopathologically as a moderately differentiated cholangiocellular carcinoma (Fig. 1A). The HCC samples were obtained from three surgically resected cancerous lesions, all of which were diagnosed histopathologically as well-differentiated HCCs (Fig. 1B). The clinical characteristics of the patients are provided in Table 1.

In total, 74 samples (i.e., 16 ICC, seven normal liver (NL), 20 chronic liver disease (CLD), 26 HCC (Tables 1 and 2) and five extrahepatic adenocarcinoma) were used for quantitative real-time reverse transcriptase-polymerase chain reaction detection (RT-PCR) and discrimination analyses. The NL samples were obtained by the surgical resection of colon tumours from patients with metastatic liver cancer, whereas the CLD samples were non-cancerous liver samples obtained from patients with HCC. The ICC samples were taken from nine patients by surgical resection and six cadavers at autopsy, in addition to the sample used for SAGE. The characteristics of the patients are provided in Table 2. As samples of extrahepatic adenocarcinoma, we used BD Premium Total RNA™ (BD Biosciences Clontech, Palo Alto, CA) taken from adenocarcinomas of human breast (Cat. No.636635), colon (636634), stomach (636629), uterus (636628) and lung (636633).

The study protocol conformed to the ethical guidelines of the Declaration of Helsinki (1975). All patients provided written informed consent for the analysis of the biopsy specimens and the hospital ethics committee approved the study.

2.2. SAGE

Each tissue sample was homogenised in liquid nitrogen and total RNA was extracted using a ToTALLY RNA™ kit (Ambion, Austin, TX). The polyadenylated RNA (polyA-RNA) was subsequently purified using a MicroPoly(A)Pure™ kit (Ambion). Aliquots of polyA-RNA (3 μg) from each ICC and HCC sample were used for SAGE, as described previously [15–17]. The NL tissue and HCC cDNA libraries have also been described previously [16,17]. To compare liver cancer with cancers from other organs, we obtained 16 SAGE libraries from the National Center for Biotechnology Information (NCBI) SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/). These libraries originated from the tumours of the stomach (G189 and G234), colon (Tu98 and Tu102), prostate (PrCA-1 and Chen_Tumour_Pr) and breast (95-259, 95-347, DCIS, DCIS-2, DCIS-3, DCIS-4, DCIS-5, IDC-3, IDC-4 and IDC-5).

For the analysis of molecular functions, we used all databases registered in MetaCore™ from GeneGo Inc. (http://portal.genego.com/cgi/index.cgi). Statistical significance in this database was calculated using the basic equation (Supplementary Fig. 1).

2.3. Quantitative real-time RT-PCR detection

Template cDNA was synthesised from 1 μg of total RNA using SuperScript™ II RT (Invitrogen, San Diego, CA). The quantitative real-time RT-PCR detection was performed using a TaqMan® Gene Expression Assay kit (Applied Biosystems, Foster City, CA). The amount of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) (Hs99999905) mRNA in each sample was used to standardise the quantity of each of the following mRNA: Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4 (*CITED4*) (Hs00388363), glutathione-*S*-transferase pi (*GSTP1*) (Hs00168310), biglycan (*BGN*) (Hs00156076), insulin-like growth factor-binding protein 5 (*IGFBP5*) (Hs00181213), claudin-4 (*CLDN4*) (Hs00533616), phosphofructokinase platelet type (*PFKP*) (Hs00242993), transmembrane 4 L six family member 1 (*TM4SF1*) (Hs00371997), calpain 1 (mu/l) large subunit (*CAPN1*) (Hs00559804) and claudin-10 (*CLDN10*) (Hs00199599).

2.4. Immunoblot analysis

Each tissue sample was homogenised in liquid nitrogen and a protein extract was prepared using radioimmunoprecipitation assay (RIPA) buffer. The extracts (7 μg/lane) were subsequently electrophoresed on SDS–10% polyacrylamide gels and transferred onto polyvinylidene fluoride. An extract prepared from KMBC cells, a human extrahepatic bile duct carcinoma cell line [21], was used as a positive control. The blots were then incubated for 1 h with goat polyclonal antibodies against claudin-4 (1:200 dilution; Santa Cruz Biotechnol-

Table 1
Characteristics of the samples used for RT- PCR and SAGE

	NL $N = 7$	CLD $N = 20$	HCC $N = 26^a$	ICC $N = 16^a$
Mean age, years (SE)	64.6 (4.3)	62.7 (2.1)	65.0 (1.9)	67.9 (2.7)
Sex				
Female	3	5	6	5
Male	4	15	20	11
Hepatitis virus infection				
HBsAg-positive	0	8	9	
HCV-Ab-positive	0	12	16	
Mean ALT (SE) (IU/L)	14.2 (2.0)	61.3 (13.9)	52.6 (11.0)	47.0 (13.8)
Mean ALP (SE) (IU/L)	234.4 (36.2)	289.0 (33.1)	321.3 (31.5)	482.0 (87.1)
Mean GGT (SE) (IU/L)	93.3 (72.0)	88.0 (14.4)	82.9 (14.1)	185.5 (61.1)
Mean AFP (SE) (ng/mL)	NA	81.8 (29.7)	80.9 (24.6)	NA
Mean CEA (SE) (ng/mL)	NA	NA	5.3 (0.8)	20.9 (14.0)
Mean CA19-9 (SE) (U/L)	NA	NA	36.7 (3.7)	1171.5 (767.1)

NL, normal liver; CLD, chronic liver disease; HCC, hepatocellular carcinoma; ICC, intrahepatic cholangiocarcinoma; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutamyl transferase; AFP, alpha-foetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; NA, not available.

[a] We used one ICC sample and three HCC samples for SAGE.

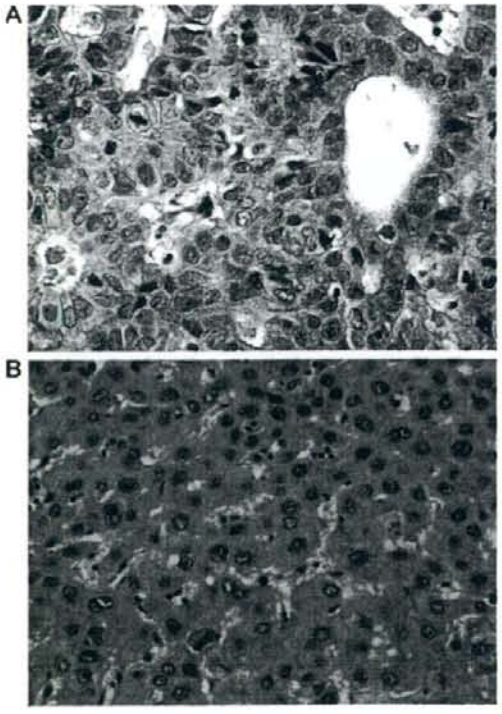

Fig. 1. Histopathological findings of the samples used for the serial analysis of gene expression. (A) Moderately differentiated intrahepatic cholangiocellular carcinoma. (B) Well-differentiated hepatocellular carcinoma. [This figure appears in colour on the web.]

ogy, Santa Cruz, CA). The blots were then washed and exposed to rabbit anti-goat IgG (1:1000 dilution) for 30 min and visualised using the ECL™ kit (GE Healthcare UK Ltd., Little Chalfont, UK).

2.5. Immunohistochemical analysis

For immunohistochemical analysis, we used 16 ICC samples and 16 HCC samples. Deparaffinised sections were treated with primary antibodies against CLDN4 (diluted 1:24, rabbit polyclonal antibodies; Cat. No. sc-17664; Santa Cruz Biotechnology) for 40 min at room temperature after appropriate antigen-retrieval treatment. Primary antibody/antigen binding was detected using a standard streptavidin–biotin–peroxidase technique (LSAB2 Kit Universal/HRP Rabbit/Mouse; DAKO, Glostrup, Denmark) and visualised using a DAB+ (3,3′-diaminobenzidine tetrahydrochloride) Liquid System (DAKO).

2.6. Discrimination analysis

We performed discrimination analysis based on the mRNA expression data using Dr. SPSS II software (SPSS Inc., Chicago, IL) to create an ICC classifier. Fifty-three of the seventy-four samples mentioned above were selected randomly: seven ICC, four NL, 20 CLD, 17 HCC and five extrahepatic adenocarcinoma samples. These 53 samples, which were subjected to RT-PCR analysis, were considered to be the training group. To verify the validity of the analysis, we performed RT-PCR using another 21 samples: nine ICC, three NL and nine HCC samples.

2.7. Statistical analyses

One-way ANOVA and Tukey–Kramer analyses were used for the statistical analysis of RT-PCR data.

3. Results

3.1. SAGE profiles of ICC

We obtained a total of 93,874 SAGE tags: 34,079 from the ICC library and 59,795 from the HCC library. Of these, 30,859 tags were unique: 14,168 and 16,689

Table 2
Clinicopathological features and discrimination analysis of the ICC samples used for RT-PCR

No.	Age (yr)	Gender	Virus	Tumour location	Size (cm)	Stage[a]	Gross appearance	Diff.	Serological marker		Z-Score	Predicted result[d]
									CEA[b] (ng/mL)	CA19-9[c] (IU/mL)		
ICC samples of the training group												
1	58	M	HCV	P	3.0 × 2.5	3	MF	Mod	5.20	23.00	4.02	ICC
2	73	M	HCV	A/P	6.5 × 5.5	4B	MF	Poor	NA	NA	7.09	ICC
3	72	M	No	A/P	7.0 × 6.5	4A	MF	Mod	7.50	NA	1.56	ICC
4	76	M	HBV	L/Med	5.3 × 4.7	4B	MF+PDI	Mod	46.00	8,600.00	2.16	ICC
5	68	F	HCV	A/P	9.0 × 6.0	4B	MF+PDI	Poor	NA	NA	0.56	ICC
6	65	M	HCV	Mu	4.0 × 3.0	4A	MF	Poor	0.00	87.00	1.31	ICC
7	53	M	HCV	Med	4.5 × 3.0	4A	PDI+MF	Poor	5.20	1,161.00	6.31	ICC
ICC samples of test group												
8	35	F	No	A/P	6.8 × 4.5	4A	MF + PDI	Poor	455.43	168.80	1.34	Not ICC
9	81	F	No	L	2.6 × 2.0	3	MF	Mod	164.58	5.10	1.63	ICC
10	73	M	No	P	5.0 × 4.0	3	MF	Mod	191,886.86	1.10	206.40	ICC
11	71	F	No	A/P	10.0 × 9.0	3	MF	Mod	3335.47	NA	13.28	ICC
12	72	F	No	M	5.0 × 4.5	3	PDI	Mod	28,614.09	NA	43.17	ICC
13	70	M	No	M	4.0 × 3.0	3	MF	Well	4533.44	4.10	16.17	ICC
14	61	M	HCV	P	1.8 × 1.5	3	MF	Mod	573.55	3.80	2.05	ICC
15	80	M	HCV	A/P	3.5 × 2.7	3	MF	Mod	2914.25	0.00	9.16	ICC
16	62	M	No	L/Med	5.2 × 3.7	3	MF	Mod	11,024.51	3.40	28.45	ICC

ICC, intrahepatic cholangiocarcinoma; Diff, differentiation; M, male; F, female; A, anterior segment; P, posterior segment; Med, medial segment; L, lateral segment; Mu, multiple segment; MF, mass-forming type; PDI, periductal infiltrating type; well, well differentiated; mod, moderately differentiated; poor, poorly differentiated.

[a] International Hepato–Pancreato–Biliary Association (IHBPA) classification [20].

[b] The normal range of CEA is <5 ng/mL.

[c] The normal range of CA19-9 is <37 IU/mL.

[d] The predicted results were determined using leave-one-out cross-validation.

tags from the ICC and HCC library, respectively. Scatterplots using all tags indicated that the ICC expression profile was different from that of HCC ($R = 0.4629$) and NL ($R = 0.3438$; Supplementary Fig. 2), whereas the HCC expression profiles were correlated with those of NL ($R = 0.8632$; data not shown). We also observed a strong correlation between the expression profile of the HCC used here and that of another HCC ($R = 0.8027$; data not shown) [17]. These findings suggest that the expression profile of ICC differed from those of HCC and NL.

To eliminate as much sequence error as possible, only tags with two or more hits were selected from the ICC library, resulting in 3707 tags. When we searched for these tags in the NCBI SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/), we found that 1348 (36.4%) corresponded to single known genes, 341 (9.2%) corresponded to single expressed sequence tags (EST) and 1740 (46.9%) had multiple matches. The remaining 278 tags (7.5%) did not correspond to any known gene.

3.2. Gene signatures of ICC

We compared the transcript abundance in the ICC and HCC libraries without excluding tags with one hit and found that the levels of expression of 1898 genes were enhanced over fivefold in ICC. As expected, several genes, previously reported to be upregulated in ICC, were amongst the 20 genes that showed the greatest degree of upregulation (Table 3), including *KRT7*, *KRT19* and *S100 calcium-binding protein A6* (*S100A6*) [22]. We also identified several genes not previously well characterised as associated with ICC, such as *BGN* and *IGFBP5*. Some of the ICC-specific genes were not detected in the NL library, suggesting that these gene signatures may be specific to cholangiocytes.

To determine the molecular functions of the 1898 genes that were overexpressed in ICC than in HCC, we analysed all the SAGE tags representing these genes in the MetaCore™ from GeneGo Inc. (Supplementary Table 2). The biological processes of these genes included the regulation of cell adhesion molecules, translation initiation and the regulation of wingless-type MMTV integration site (Wnt) signalling ($P < 0.01$; statistical significance calculated using the basic equation in Supplementary Fig. 1).

We also found that 702 genes were underexpressed over fivefold in ICC relative to HCC without excluding tags with one hit; the 20 genes that showed the greatest degree of downregulation in ICC are listed in Supplementary Table 3. Amongst these were genes encoding apolipoprotein, fibronectin and haptoglobin, which were also underexpressed in ICC relative to NL. The biological processes associated with these genes (Supplementary Table 4) included immune response and the

Table 3
Twenty genes overexpressed in ICC compared to HCC

SAGE Tag	RefSeq ID	Gene name	Tag count			Fold (ICC/HCC)
			ICC	HCC	NL	
GACGCCGAAC	NM_133467	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4	111	1	9	122.00
CCTGGTCCCA	NM_005556	Keratin 7	103	0	0	113.00
AATAGAAATT	NM_000582	Secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T lymphocyte activation 1)	83	1	0	92.00
CCCCCTGGAT	NM_014624	S100A6 S100 calcium-binding-protein A6 (calcyclin)	66	0	0	73.00
CTTCCAGCTA	NM_004039	Annexin A2	55	1	3	61.00
GCAATCCTGT	NM_006998	Secretagogin, EF hand calcium-binding protein	52	0	0	58.00
GACATCAAGT	NM_002276	Keratin 19	50	0	0	55.00
GCAAAGAAAA	NM_006769	LIM domain only 4	44	0	3	49.00
CAGGCCCCAC	NM_005620	S100 calcium-binding protein A11 (calgizzarin)	41	1	0	46.00
GGAGACTTCC	NM_001153	Annexin A4	36	0	3	39.00
TGGCCCCACC	NM_002654	Pyruvate kinase, muscle	36	1	3	39.00
GCATTTGACA	NM_001046	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2	33	0	0	36.00
AACTTGGCCA	NM_002423	Matrix metalloproteinase 7 (matrilysin, uterine)	30	1	0	33.00
AGGTCCTAGC	NM_000852	Glutathione-*S*-transferase pi	30	0	6	33.00
ATCTTTCTGG	NM_003406	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	30	1	0	33.00
TGATGTCTGG	NM_020182	Transmembrane, prostate androgen induced RNA	30	0	0	33.00
GACCGCAGGA	NM_001845	Collagen, type IV, alpha 1	27	1	0	30.00
GCCTGTCCCT	NM_001711	Biglycan	27	1	6	30.00
GGAACAAACA	NM_013230	CD24 molecule	27	0	0	30.00
TATGAATGCT		No reliable match	27	0	0	30.00

SAGE, serial analysis of gene expression; ICC, intrahepatic cholangiocarcinoma; HCC, hepatocellular carcinoma; NL, normal liver.

regulation of inflammation via complement activation ($P < 0.01$). These results suggest that the gene expression pattern of ICC is different from that of HCC.

Amongst the genes that were upregulated more than 20 times in the ICC library compared with the HCC library, we selected nine that were not previously well characterised to be overexpressed in ICC (Table 4) for further analysis. We defined these genes as the candidate markers of ICC and compared the SAGE libraries originating from extrahepatic adenocarcinoma (Supplementary Table 5). By comparing the mean transcript abundance in ICC with those in the other libraries, excluding tags with one hit, we found that eight genes (*CITED4*, *BGN*, *IGFBP5*, *CLDN4*, *PFKP*, *TM4SF1*, *CAPN1* and *CLDN10*) showed a threefold greater expression in ICC than in at least three other organs. These genes appear to be useful not only for the differentiation of ICC and HCC, but also for the differentiation of ICC and metastatic liver cancer.

3.3. RT-PCR and protein expression analyses

The above findings are based on the library analysis of one ICC sample; thus, their reliability must be validated with multiple samples. Moreover, the application of SAGE results to routine diagnostic examination is indispensable. We performed RT-PCR using RNA samples isolated from seven ICC, four NL, 20 CLD, 17 HCC and five extrahepatic adenocarcinomas (breast, colon, stomach, ovary and lung) to validate the expression data for the eight genes mentioned above. Of these eight genes, three (*BGN*, *IGFBP5* and *CLDN4*) were more highly expressed in ICC than in NL, CLD, HCC or extrahepatic adenocarcinomas ($P < 0.05$; Fig. 2; Supplementary Table 1), suggesting that these three genes are the specific gene signatures for ICC.

Of *BGN*, *IGFBP5* and *CLDN4*, the expression of CLDN4 was confirmed by immunoblotting (Fig. 3A) and immunohistochemical analyses (Fig. 3B and C). Immunoblotting revealed strong CLDN4 expression in six of seven patients with ICC and KMBC. In contrast, slight or no expression was detected in HCC and NL (Fig. 3A). Immunohistochemical analysis showed intense membrane staining in ICC, whereas no CLDN4 expression was observed in HCC (Fig. 3B and C).

3.4. Discrimination analysis

Based on the expression patterns of *BGN*, *IGFBP5* and *CLDN4* in the 53 samples of the training group, we performed discrimination analysis to obtain coefficients for these three genes, which were then used to

Table 4
Nine genes not previously well characterised and showing >20-fold overexpression from the ICC library compared to the HCC library

RefSeq ID	Gene name	Tag count		Ratio of RT-PCR					P value[a]
		ICC	HCC	ICC	NL	CLD	HCC	Other	
NM_133467	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4	122	1	27.56	1.37	1.81	1.13	0.13	0.0305
NM_000852	Glutathione-S-transferase pi	33	0	10.61	2.33	2.54	1.8	1.4	–
NM_001711	Biglycan	30	1	5.4	1.06	1.19	0.68	0.35	<0.0001
NM_000599	Insulin-like growth factor-binding protein 5	24	0	11.13	1.86	0.68	0.83	0.21	<0.0001
NM_001305	Claudin-4	24	1	344.98	1.71	7.92	8.56	0.44	0.0036
NM_002627	Phosphofructokinase, platelet	24	0	160.68	0.5	0.13	0.08	6.16	0.0089
NM_014220	Transmembrane 4 superfamily member 1	24	0	4,687.43	25.12	37.64	498.02	0.44	0.1882
NM_005186	Calpain 1, (μ/I) large subunit	21	1	4.32	1.33	0.92	0.85	5.76	<0.0001
NM_006984	Claudin-10	21	0	6.25	0.56	2.69	0.11	21.26	0.0638

ICC, intrahepatic cholangiocarcinoma; NL, normal liver; CLD, chronic liver disease; HCC, hepatocellular carcinoma; Other, extrahepatic adenocarcinoma; ICC, intrahepatic cholangiocarcinoma; HCC, hepatocellular carcinoma.
[a] One-way ANOVA.

define the Z-score as positive in ICC (Fig. 4A). The prediction performance of the discrimination analysis showed an efficient ROC curve (AUC = 0.987) (Fig. 4B). To verify the validity of this classifier, we calculated the Z-score for the 21 samples in the test group and found that the score was positive for all the ICC samples (Fig. 4C). Thus, we could distinguish ICC from the other cancers using this classifier.

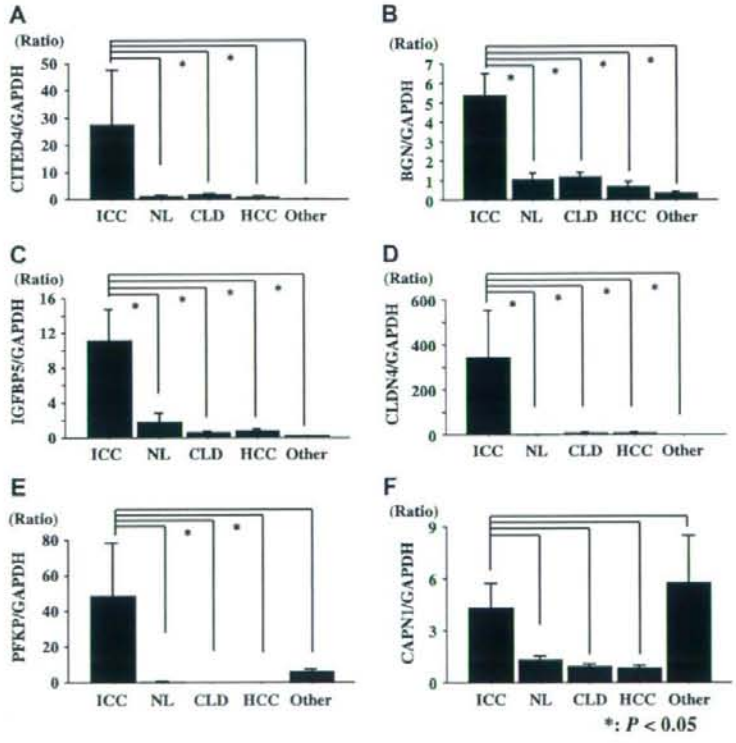

Fig. 2. Results of RT-PCR for intrahepatic cholangiocarcinoma (ICC), normal liver (NL), chronic liver disease (CLD), hepatocellular carcinoma (HCC) and other adenocarcinomas. GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CTTED4, Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4; BGN, biglycan; IGFBP5, insulin-like growth factor-binding protein 5; CLDN4, claudin-4; PFKP, platelet-type phosphofructokinase; CAPN1, calpain 1 (μ/I) large subunit.

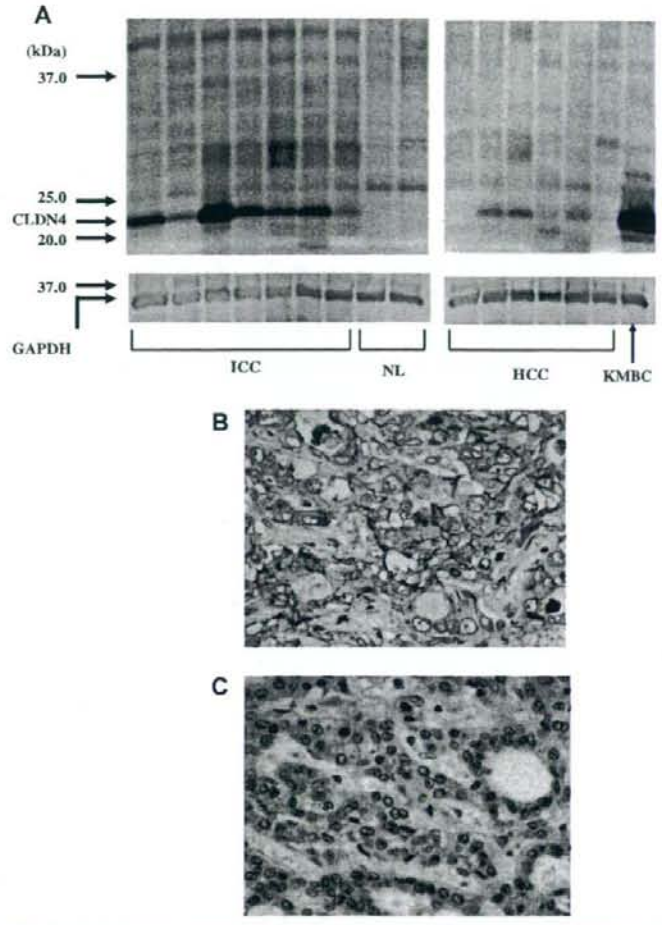

Fig. 3. Representative CLDN4 Western blots and immunohistochemical data. (A) CLDN4 (25 kDa) expression was strong in the intrahepatic cholangiocarcinoma (ICC) samples. The molecular mass marker (kDa) is indicated. (B) Intense membrane staining in intrahepatic cholangiocarcinoma. (C) Hepatocellular carcinoma showing no CLDN4 expression. GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CLDN4, claudin-4; NL, normal liver; HCC, hepatocellular carcinoma; KMBC, a human extrahepatic bile duct carcinoma cell line [21]. [This figure appears in colour on the web.]

4. Discussion

Using gene expression profiling, we identified a distinct gene signature for ICC. We chose the SAGE technique, which is used to comprehensively identify and quantify the expression of known and unknown genes to analyse the gene expression in ICC. In SAGE methods, gene expressions of different organs or tissues are comparable because the tag sequence directly reflects RNA copy number, and the global standardised methods of SAGE enable the analysis and comparison of the data obtained from different laboratories through SAGE maps (http://www.ncbi.nlm.nih.gov/SAGE/). Another advantage of SAGE is that it allows the identification of genes not previously reported, including expressed sequence tags (ESTs). In fact, we detected some ESTs that were overexpressed in ICC compared to HCC (data not shown) and these novel sequences may represent new tumour markers for ICC.

When we compared the ICC and HCC libraries, *KRT7* and *KRT19* were amongst the genes showing overexpression in ICC. Both are expressed in bile duct epithelium, indicating that ICC arises from bile duct epithelial cells. In addition, we observed highly abundant

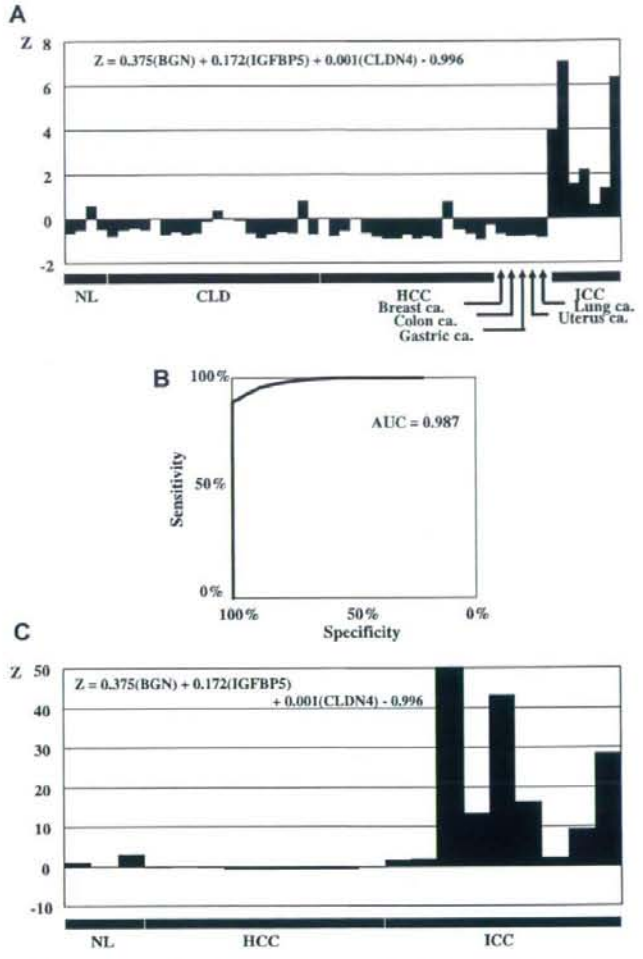

Fig. 4. Discrimination analysis of gene expression data. From the coefficient obtained for each gene, the *Z*-score was calculated as $Z = 0.375$ (BGN) + 0.172 (IGFBP5) + 0.001 (CLDN4) − 0.996. This score was positive for the intrahepatic cholangiocarcinoma (ICC) sample but negative for the other samples. (A) Results from the samples of training group. (B) Prediction performance of *Z*-score using the ROC curve. (C) Results from samples of the test group. BGN, biglycan; IGFBP5, insulin-like growth factor-binding protein 5; CLDN4, claudin-4; NL, normal liver; CLD, chronic liver disease; HCC, hepatocellular carcinoma; AUC, area under the curve.

expression in ICC of the genes encoding S100A6 and matrix metalloproteinase 7 [22–24], both previously reported in ICC. Therefore, our findings confirm the expression patterns reported previously for ICC.

We also analysed the molecular functions of the genes showing an over fivefold higher expression in ICC than in HCC and found that many of them were associated with the regulation of cell adhesion molecules, translation initiation and the regulation of Wnt signalling. Conversely, amongst the genes with an over fivefold overexpression in HCC than in ICC, the majority were associated with immunity and inflammation. These findings indicate that the characteristic gene expression patterns in ICC differ from those in HCC.

In addition to tissue-specific genes, cancer-associated genes, such as those encoding annexin A2 [25–27], S100 calcium-binding protein A11 [28], glutathione-*S*-transferase pi [29], transmembrane, prostate androgen-induced RNA [30] and CD24 antigen [31] (Table 3), were also overexpressed in ICC compared to HCC.

Thus, genes that were differentially expressed between ICC and HCC do not merely reflect differences in the cell types or tissue types from which the tumours arise, but differences in the ICC and HCC tumours themselves.

Analyses using SAGE libraries showed that nine genes, which were not previously well characterised as expressed in ICC, were overexpressed in ICC than in HCC. By comparing the mean transcript abundance in ICC with those in other libraries, including gastric, colon, prostate and breast cancer, we found that eight genes were overexpressed in ICC than in the other organs: *CITED4*, *BGN*, *IGFBP5*, *CLDN4*, *PFKP*, *TM4SF1*, *CAPN1* and *CLDN10*. As these findings were based on one ICC sample, we confirmed their validity using RT-PCR and found that the expression of *BGN*, *IGFBP5* and *CLDN4* was higher in ICC than in HCC, NL, CLD or adenocarcinomas originating from other organs.

BGN encodes a matrix proteoglycan involved in the metabolism of connective tissue by binding to collagen and TGF-beta. It is expressed in the lung, spleen and liver of mice [32]. Immunohistochemically, the BGN protein has been shown to be expressed in Disse's space, cholangioles and the vascular wall, but not in normal liver [33].

IGFBP5 encodes a protein that binds to IGF-I and -II, but its precise function is not yet known. This protein has been reported to promote osteoblast mitosis [34] and to cause the involution of mammary gland cells by inducing apoptosis via IGF-dependent and -independent pathways [35]. However, its function in liver and bile duct cells is unknown.

CLDN4 encodes an essential membrane protein of the claudin family. This protein is a component of tight junctions, a membrane receptor for *Clostridium perfringens* enterotoxin, and is thought to be involved in organogenesis. In addition, CLDN4 is highly expressed in the small intestine, but only weakly expressed in mouse liver [36]. Whilst our study was under way, Lodi et al. [14] reported that the level of CLDN4 expression was higher in biliary tract cancers than in HCC.

Other candidate genes more suitable for the differentiation of ICC from HCC and metastatic liver cancers may exist, but our three genes – BGN, IGFBP5 and CLDN4 – were not previously reported to be overexpressed in ICC, although CLDN4 was cited whilst our study was under way.

Moreover, the prediction performance of ICC discrimination analysis using these three genes showed an efficient ROC curve (AUC = 0.987) and the equation allowed us to completely distinguish ICC from HCC, CLD and NL, although further evaluation is needed using, for example, metastatic liver cancers. The ROC curve was also generated for the test group and the AUC (0.963) was as high as that generated using the training set (data not shown). The discrimination of ICC using the known markers such as CEA and CA19-9 was less effective than this model. The AUC for CEA and CA19-9 was 0.676 and 0.722, respectively (data not shown).

However, the inclusion of other ICC markers will enhance the sensitivity and specificity; for example, both CEA and CA19-9 were enhanced in 50% and 62% of patients with ICC, respectively (Table 2). In addition, we found no correlation between the level of CEA or CA19-9 and the expression of the three genes, suggesting that they are independent markers of ICC. Thus, a combined evaluation of these markers could enhance sensitivity and specificity.

Despite the limited numbers of patients with ICC and SAGE libraries examined, this is the first report of ICC-specific, novel candidate tumour-marker genes. Further validation using a larger cohort is needed to confirm the expression of these genes in more patients with ICC and to examine the functional relevance of these genes in the development and prognosis of ICC.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhep.2008.03.025.

References

[1] Shaib Y, El-Serag HB. The epidemiology of cholangiocarcinoma. Semin Liver Dis 2004;24:115–125.
[2] Patel T. Increasing incidence and mortality of primary intrahepatic cholangiocarcinoma in the United States. Hepatology 2001;33:1353–1357.
[3] Taylor-Robinson SD, Toledano MB, Arora S, Keegan TJ, Hargreaves S, Beck A, et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968–1998. Gut 2001;48:816–820.
[4] Khan SA, Taylor-Robinson SD, Toledano MB, Beck A, Elliott P, Thomas HC. Changing international trends in mortality rates for liver, biliary and pancreatic tumours. J Hepatol 2002;37:806–813.
[5] Davila JA, El-Serag HB. Cholangiocarcinoma: the "other" liver cancer on the rise. Am J Gastroenterol 2002;97:3199–3200.
[6] Marrero JA, Lok AS. Newer markers of hepatocellular carcinoma. Gastroenterology 2004;127:S113–S119.
[7] Johnson DE, Herndier BG, Medeiros LJ, Warnke RA, Rouse RV. The diagnostic utility of the keratin profiles of hepatocellular carcinoma and cholangiocarcinoma. Am J Surg Pathol 1988;12:187–197.
[8] Lai YS, Thung SN, Gerber MA, Chen ML, Schaffner F. Expression of cytokeratins in normal and diseased livers and in primary liver carcinomas. Arch Pathol Lab Med 1989;113:134–138.
[9] Maeda T, Adachi E, Kajiyama K, Sugimachi K, Tsuneyoshi M. Combined hepatocellular and cholangiocarcinoma: proposed criteria according to cytokeratin expression and analyses of clinicopathologic features. Hum Pathol 1995;26:956–964.
[10] Altmannsberger M, Weber K, Holscher A, Schauer A, Osborn M. Antibodies to intermediate filaments as diagnostic tools: human

gastrointestinal carcinomas express prekeratin. Lab Invest 1982;46:520–526.
[11] Saintigny P, Coulon S, Kambouchner M, Ricci S, Martinot E, Danel C, et al. Real-time RT-PCR detection of CK19, CK7 and MUC1 mRNA for diagnosis of lymph node micrometastases in non small cell lung carcinoma. Int J Cancer 2005;115:777–782.
[12] Khan SA, Davidson BR, Goldin R, Pereira SP, Rosenberg WM, Taylor-Robinson SD, et al. Guidelines for the diagnosis and treatment of cholangiocarcinoma: consensus document. Gut 2002;51:VI1–VI9.
[13] Gores GJ. Cholangiocarcinoma: current concepts and insights. Hepatology 2003;37:961–969.
[14] Lódi C, Szabó E, Holczbauer A, Batmunkh E, Szijártó A, Kupcsulik P, et al. Claudin-4 differentiates biliary tract cancers from hepatocellular carcinomas. Modern Pathol 2006;19:460–469.
[15] Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:484–487.
[16] Yamashita T, Hashimoto S, Kaneko S, Nagai S, Toyoda N, Suzuki T, et al. Comprehensive gene expression profile of a normal human liver. Biochem Biophys Res Commun 2000;269:110–116.
[17] Yamashita T, Kaneko S, Hashimoto S, Sato T, Nagai S, Toyoda N, et al. Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma. Biochem Biophys Res Commun 2001;282:647–654.
[18] Kawai HF, Kaneko S, Honda M, Shirota Y, Kobayashi K. Alpha-fetoprotein-producing hepatoma cell lines share common expression profiles of genes in various categories demonstrated by cDNA microarray analysis. Hepatology 2001;33:676–691.
[19] Shirota Y, Kaneko S, Honda M, Kawai HF, Kobayashi K. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. Hepatology 2001;33:832–840.
[20] Makuuchi M, Belghiti J, Belli G, Fan ST, Lau JW, Ringe B, et al. IHPBA concordant classification of primary liver cancer: working group report. J Hepatobiliary Pancreat Surg 2003;10:26–30.
[21] Yano H, Maruiwa M, Iemura A, Mizogichi A, Kojiro M. Establishment and characterization of a new human extrahepatic bile duct carcinoma cell line (KMBC). Cancer 1992;69:1664–1673.
[22] Kim J, Kim J, Yoon S, Joo J, Lee Y, Lee K, et al. S100A6 protein as a marker for differential diagnosis of cholangiocarcinoma from hepatocellular carcinoma. Hepatol Res 2002;23:274–286.
[23] Lichtinghagen R, Helmbrecht T, Arndt B, Böker KH. Expression pattern of matrix metalloproteinases in human liver. Eur J Clin Chem Clin Biochem 1995;33:65–71.
[24] Miwa S, Miyagawa S, Soeda J, Kawasaki S. Matrix metalloproteinase-7 expression and biologic aggressiveness of cholangiocellular carcinoma. Cancer 2002;94:428–434.
[25] Kumble KD, Hirota M, Pour PM, Vishwanatha JK. Enhanced levels of annexins in pancreatic carcinoma cells of Syrian hamsters and their intrapancreatic allografts. Cancer Res 1992;52:163–167.
[26] Emoto K, Sawada H, Yamada Y, Fujimoto H, Takahama Y, Ueno M, et al. Annexin II overexpression is correlated with poor prognosis in human gastric carcinoma. Anticancer Res 2001;21:1339–1345.
[27] Emoto K, Yamada Y, Sawada H, Fujimoto H, Ueno M, Takayama T, et al. Annexin II overexpression correlates with stromal tenascin-C overexpression: a prognostic marker in colorectal carcinoma. Cancer 2001;92:1419–1426.
[28] Ohuchida K, Mizumoto K, Ohhashi S, Yamaguchi H, Konomi H, Nagai E, et al. S100A11, a putative tumour suppressor gene, is overexpressed in pancreatic carcinogenesis. Clin Cancer Res 2006;12:5417–5422.
[29] Cai L, Mu LN, Lu H, Lu QY, You NC, Yu SZ, et al. Dietary selenium intake and genetic polymorphisms of the GSTP1 and p53 genes on the risk of esophageal squamous cell carcinoma. Cancer Epidemiol Bio Prev 2006;15:294–300.
[30] Xu LL, Shi Y, Petrovics G, Sun C, Makarem M, Zhang W, et al. PMEPA1, an androgen-regulated NEDD4-binding protein, exhibits cell growth inhibitory function and decreased expression during prostate cancer progression. Cancer Res 2003;63:4299–4304.
[31]] Karahan N, Güney M, Oral B, Kapucuoglu N, Mungan T. CD24 expression is a poor prognostic marker in endometrial carcinoma. Eur J Gynaecol Oncol 2006;27:500–504.
[32] Wegrowski Y, Pillarisetti J, Danielson KG, Suzuki S, Iozzo RV. The murine biglycan: complete cDNA cloning, genomic organization, promoter function, and expression. Genomics 1995;30:8–17.
[33] Hogemann B, Edel G, Schwarz K, Krech R, Kresse H. Expression of biglycan, decorin and proteoglycan-100/CSF-1 in normal and fibrotic human liver. Pathol Res Pract 1997;193:747–751.
[34] Schneider MR, Wolf E, Hoeflich A, Lahm H. IGF-binding protein-5: flexible player in the IGF system and effector on its own. J Endocrinol 2002;172:423–440.
[35] Butt AJ, Dickson KA, McDougall F, Baxter RC. Insulin-like growth factor-binding protein-5 inhibits the growth of human breast cancer cells in vitro and in vivo. J Biol Chem 2003;278:29676–29685.
[36] Katahira J, Sugiyama H, Inoue N, Horiguchi Y, Matsuda M, Sugimoto N. *Clostridium perfringens* enterotoxin utilizes two structurally related membrane proteins as functional receptors in vivo. J Biol Chem 1997;272:26652–26658.

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 according 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

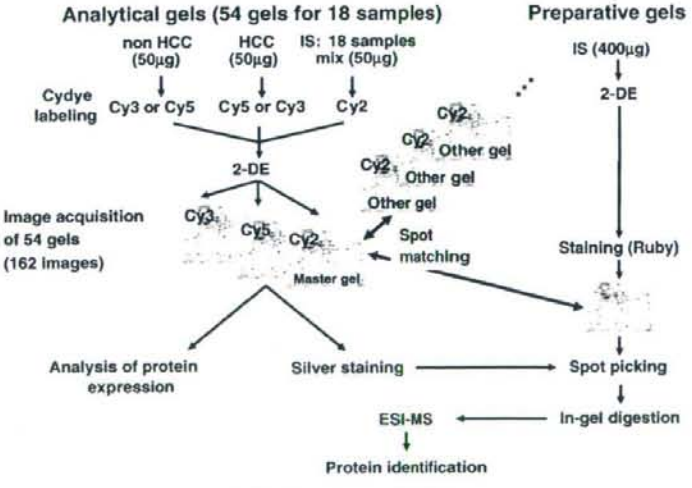

Fig. 1. The experimental strategy for 2D-DIGE.

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_{rl} 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

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

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

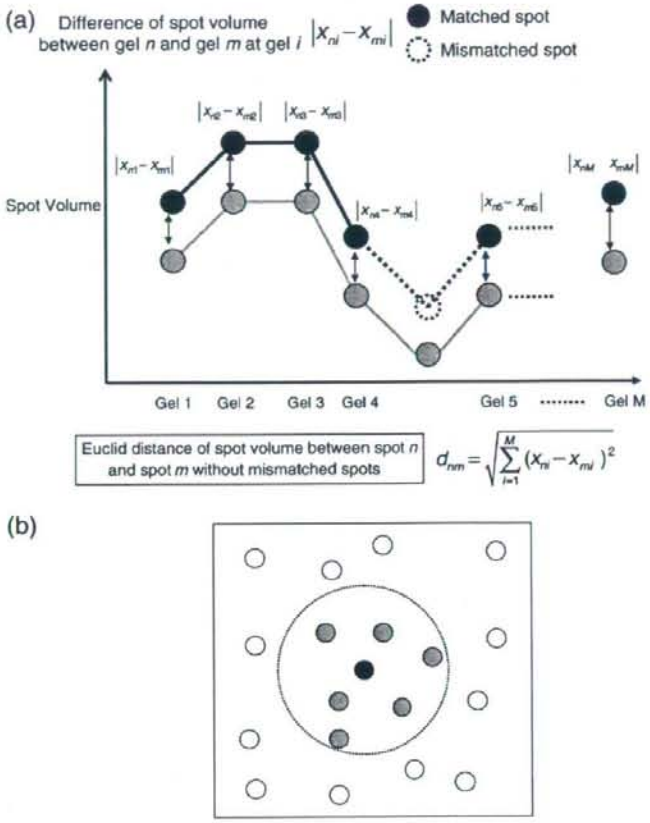

Fig. 2. Illustration of missing value estimation based on k-nearest neighbors. (a) Calculation of Euclid distance between spot n and m without mismatched spots. (b) Longitudinal axis and abscissa axis represent a vector space spanned by spot volumes. Black spot, gray spots and white spots represent missing 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 overfitting. The values r_{ti} in the fourth lines of the algorithms are the gradients of the loss function $L(Y,F(X))$:

$$r_{ti}=-\frac{\partial}{\partial F(X_i)}L\left(Y_i,F(X_i)|_{F=F_{m-1}(X_i)}\right)$$

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.

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

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 δ_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.