disappearance of hepatitis C virus. *Am J Gastroenterol* 2005; **100**: 1748–53.
17. Radkowski M, Gallegos-Orozco JF, Jablonska J, *et al.* Persistence of hepatitis C virus in patients successfully treated for chronic hepatitis C. *Hepatology* 2005; **41**: 106–14.
18. Tamori A, Yamanishi Y, Kawashima S, *et al.* Alteration of gene expression in human hepatocellular carcinoma with integrated hepatitis B virus DNA. *Clin Cancer Res* 2005; **11**: 5821–6.
19. Desmet VJ, Gerber M, Hoofnagle JH, *et al.* Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology* 1994; **19**: 1513–20.
20. King MP, Attardi G. Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* 1989; **246**: 500–3.
21. Hayashi H, Sugio K, Matsumata T, *et al.* The clinical significance of p53 gene mutation in hepatocellular carcinomas from Japan. *Hepatology* 1995; **22**: 1702–7.
22. Herman JG, Graff JR, Myohanen S, *et al.* Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* 1996; **93**: 9821–96.
23. Zysman MA, Chapman WB, Bapat B. Considerations when analyzing the methylation status of PTEN tumor suppressor gene. *Am J Pathol* 2002; **160**: 795–800.
24. To KF, Leung WK, Lee TL, *et al.* Promoter hypermethylation of tumor-related genes in gastric intestinal metaplasia of patients with and without gastric cancer. *Int J Cancer* 2002; **102**: 623–8.
25. Tamori A, Nishiguchi S, Nishikawa M, *et al.* Correlation between clinical characteristics and mitochondrial D-loop DNA mutations in hepatocellular carcinoma. *J Gastroenterol* 2004; **39**: 1063–8.
26. Nishikawa M, Nishiguchi S, Kioka K, *et al.* Interferon reduces somatic mutation of mitochondrial DNA in liver tissues from chronic viral hepatitis patients. *J Viral Hepat* 2005; **12**: 494–8.
27. Huang LR, Hsu HC. Cloning and expression of CD24 gene in human hepatocellular carcinoma: a potential early tumor marker gene correlates with p53 mutation and tumor differentiation. *Cancer Res* 1995; **55**: 4717–21.
28. Schagdarsurengin U, Wilkens L, Steinemann D, *et al.* Frequent epigenetic inactivation of the RASSF1A gene in hepatocellular carcinoma. *Oncogene* 2003; **22**: 1866–71.
29. Matsuda Y, Ichida T, Matsuzawa J, *et al.* p16(INK4) is inactivated by extensive CpG methylation in human hepatocellular carcinoma. *Gastroenterology* 1999; **116**: 394–400.
30. Liew CT, Li HM, Lo KW, *et al.* High frequency of p16INK4A gene alterations in hepatocellular carcinoma. *Oncogene* 1999; **18**: 789–95.
31. Li X, Hui AM, Sun L, *et al.* p16INK4A hypermethylation is associated with hepatitis virus infection, age, and gender in hepatocellular carcinoma. *Clin Cancer Res* 2004; **10**: 7484–9.
32. Fukai K, Yokosuka O, Imazeki F, *et al.* Methylation status of p14ARF, p15INK4b, and p16INK4a genes in human hepatocellular carcinoma. *Liver Int* 2005; **25**: 1209–16.
33. Jin M, Piao Z, Kim NG, *et al.* p16 is a major inactivation target in hepatocellular carcinoma. *Cancer* 2000; **89**: 60–8.
34. Wong IH, Lo YM, Yeo W, *et al.* Frequent p15 promoter methylation in tumor and peripheral blood from hepatocellular carcinoma patients. *Clin Cancer Res* 2000; **6**: 3516–21.
35. Forman D, Webb P, Parsonnet J. *H. pylori* and gastric cancer. *Lancet* 1994; **343**: 243–4.
36. Forman D, Newell DG, Fullerton F, *et al.* Association between infection with *Helicobacter pylori* and risk of gastric cancer: evidence from a prospective investigation. *BMJ* 1991; **302**: 1302–5.
37. Maekita T, Nakazawa K, Mihara M, *et al.* High levels of aberrant DNA methylation in *Helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. *Clin Cancer Res* 2006; **12**: 989–95.
38. Tatematsu M, Tsukamoto T, Mizoshita T. Role of *Helicobacter pylori* in gastric carcinogenesis: the origin of gastric cancers and heterotopic proliferative glands in Mongolian gerbils. *Helicobacter* 2005; **10**: 97–106.
39. Ushijima T, Nakajima T, Maekita T. DNA methylation as a marker for the past and future. *J Gastroenterol* 2006; **41**: 401–7.
40. Alter MJ. Epidemiology of hepatitis C in the West. *Semin Liver Dis* 1995; **15**: 5–14.

Liver International (2009)

Online Submissions: wjg.wjgnet.com
www.wjgnet.com
wjg@wjgnet.com

World J Gastroenterol 2008 April 7; 14(13): 2010-2022
World Journal of Gastroenterology ISSN 1007-9327

VIRAL HEPATITIS

Serial changes in expression of functionally clustered genes in progression of liver fibrosis in hepatitis C patients

Yoshiyuki Takahara, Mitsuo Takahashi, Qing-Wei Zhang, Hirotaka Wagatsuma, Maiko Mori, Akihiro Tamori, Susumu Shiomi, Shuhei Nishiguchi

Yoshiyuki Takahara, Mitsuo Takahashi, Qing-Wei Zhang, Hirotaka Wagatsuma, Maiko Mori, Exploratory & Applied Pharmaceutical Research Department, Pharmaceutical Company, Ajinomoto Co., Inc., 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan
Akihiro Tamori, Susumu Shiomi, Department of Nuclear Medicine, Graduate School of Medicine, Osaka City University, Osaka, Japan
Shuhei Nishiguchi, Division of Hepatobiliary and Pancreatic Diseases, Department of Internal Medicine, Hyogo College of Medicine, Nishinomiya, Japan
Correspondence to: Yoshiyuki Takahara, Exploratory & Applied Pharmaceutical Research Department, Pharmaceutical Company, Ajinomoto Co., Inc., 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan. yoshiyukitakahara@gmail.com
Telephone: +81-44-2105822
Received: April 23, 2007 Revised: July 18, 2007

Abstract

AIM: To investigate the relationship of changes in expression of marker genes in functional categories or molecular networks comprising one functional category or multiple categories in progression of hepatic fibrosis in hepatitis C (HCV) patients.

METHODS: Marker genes were initially identified using DNA microarray data from a rat liver fibrosis model. The expression level of each fibrosis associated marker gene was analyzed using reverse transcription-polymerase chain reaction (RT-PCR) in clinical biopsy specimens from HCV-positive patients (n = 61). Analysis of changes in expression patterns and interactions of marker genes in functional categories was used to assess the biological mechanism of fibrosis.

RESULTS: The profile data showed several biological changes associated with progression of hepatic fibrosis. Clustered genes in functional categories showed sequential changes in expression. Several sets of clustered genes, including those related to the extracellular matrix (ECM), inflammation, lipid metabolism, steroid metabolism, and some transcription factors important for hepatic biology showed expression changes in the immediate early phase (F1/F2) of fibrosis. Genes associated with aromatic amino acid (AA) metabolism, sulfur-containing AA metabolism and insulin/Wnt signaling showed expression changes in the middle phase (F2/F3), and some genes related to glucose metabolism showed altered expression in the late phase of fibrosis (F3/F4). Therefore, molecular networks showing serial changes in gene expression are present in liver fibrosis progression in hepatitis C patients.

CONCLUSION: Analysis of gene expression profiles from a perspective of functional categories or molecular networks provides an understanding of disease and suggests new diagnostic methods. Selected marker genes have potential utility for biological identification of advanced fibrosis.

Key words: Hepatitis C; Liver fibrosis; Marker gene; Gene expression; RT-PCR; Molecular network; Metabolism; Transcription factor; Diagnosis

Peer reviewer: Richard A Rippe, Dr, Department of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7038, United States

Takahara Y, Takahashi M, Zhang QW, Wagatsuma H, Mori M, Tamori A, Shiomi S, Nishiguchi S. Serial changes in expression of functionally clustered genes in progression of liver fibrosis in hepatitis C patients. *World J Gastroenterol* 2008; 14(13): 2010-2022 Available from: URL: http://www.wjgnet.com/1007-9327/14/2010.asp DOI: http://dx.doi.org/10.3748/wjg.14.2010

INTRODUCTION

Liver fibrosis is caused by liver disorders such as hepatitis C, hepatitis B, alcoholic hepatitis and non-alcoholic hepatitis. Fibrosis progresses gradually and finally disrupts liver structure and function over several decades, leading to fatal diseases such as cirrhosis and hepatocellular carcinoma (HCC). Classification of fibrosis progression is usually based on histological criteria using the METAVIR scoring system[1], which includes five stages: F0 (no fibrosis), F1, F2, F3, and F4 (cirrhosis). Such a classification is essential in decisions regarding treatment of liver fibrosis. Prominent subjective symptoms do not occur from F1 to F3, but patients begin to be aware of symptoms after F4. However, the F4 stage of cirrhosis is almost incurable and diagnosis of fibrosis at an earlier stage is desirable. The biology after F4 (cirrhosis) has been well studied, due to the interest in diagnosis and therapy for hepatocellular carcinoma (HCC),

but progression to HCC may begin in the early stage of fibrosis in hepatitis C[2].

Prevention of HCC and inhibition of fibrosis in the early phase is important, but the detailed hepatic biological changes corresponding to the F stage are unclear. To understand the background of the early stage of fibrosis, we previously identified genes that can be used as markers of biological changes in progression of hepatic fibrosis and diagnosis of fibrotic progression; these data were obtained from DNA microarray data from an experimental DMN (dimethylnitrosamine)-treated rat model of hepatic fibrosis[3]. This work led to marker genes that were arranged in functional categories related to fibrosis, based on genes associated with hepatic cell types such as Kupffer cells, hepatic stellate cells, and hepatocytes.

These marker genes give information on cell-specific and time-dependent behavior of each hepatic cell in fibrogenesis. In the current work, the behavior of marker genes associated with a particular F stage was analyzed using RT-PCR in clinical biopsy specimens from hepatitis C patients. This profile data revealed several biological changes in progression of hepatic fibrosis. Since many functionally clustered genes showed similar changes in expression, we propose serial expression changes in molecular networks associated with liver fibrosis progression in hepatitis C patients. Many functionally clustered genes showed large changes in expression in the early stage of fibrosis, suggesting the importance of therapy at this early stage. Alteration of gene expression also suggested qualitative changes in biological status in the transition from F3 to F4, which is a suspected risk factor for development of HCC. We conclude that analysis of gene expression profiles from a perspective of molecular networks provides improved understanding of disease and indicates potential methods of diagnosis.

MATERIALS AND METHODS

Patients in the clinical study

All patients were recruited from the Osaka City University Hospital (Osaka, Japan). Sixty-one patients with seropositive results in diagnosis using the third-generation hepatitis C virus enzyme-linked immunosorbent assay (Lumipulse II Ortho HCV, Ortho-Clinical Diagnostics, Tokyo, Japan) and positive serum HCV-RNA were included in the study. Informed consent was obtained from all patients. Liver biopsies were performed on all patients enrolled in the study, and the histological features of the liver specimens were analyzed and graded using the METAVIR scoring system[1]. The liver fibrotic stage (F stage) and inflammatory activity (A grade) were determined histologically: at least four subjects were found to be in each F stage classification. Determination for chymase 1 exceptionally has been done with three subjects in F4 stage due to the lack of appropriate samples.

Part of the biopsy sample from each patient was immediately immersed in RNAlater (QIAGEN, The Netherlands) to inhibit RNAase and then kept at 4℃ overnight before being transferred to another tube and frozen at -80℃.

RT-PCR analysis

Total RNA was extracted from liver biopsy samples using an ISOGEN kit (Nippongene) and reverse transcribed using a High Capacity cDNA Archive Kit (ABI, Foster City, CA), in each case according to the manufacturer's instructions. The total RNA in the final reaction mixture was 10 ng/μL. Real-time PCR was performed on an Applied Biosystems 7500 Real-Time PCR System (ABI) data collection system, and analyses were performed using the accompanying software. RT-PCR was performed using 0.8 μL cDNA in each well, with a final concentration of 1X the probes of the TaqMan® Gene Expression Assay and 1X the Taqman Universal PCR Master Mix (ABI). The final reaction volume was 20 μL. Each sample was analyzed in duplicate. The thermal cycler conditions were 2 min at 50℃ and 10 min at 95℃, followed by 40 cycles of 15 s at 95℃ and 1 min at 60℃.

Data were analyzed using the comparative CT method, in which the expression level of a target gene is normalized relative to an endogenous reference. GAPDH was used as the endogenous reference in all experiments. The target C_T and endogenous control C_T were calculated for each sample, and the target gene expression level was then calculated using the formula $2^{(34-CT)}$. The average of duplicate measurements was obtained, and the relative expression of each gene in a sample was calculated by setting the expression of GAPDH equal to 1000. PCR fluorogenic probes for all the target genes and the endogenous reference were purchased as TaqMan® Gene Expression Assays (ABI).

Statistical analysis

A Kruskal-Wallis test was applied to select marker genes with statistically significant changes ($P < 0.05$) in expression level during the fibrosis progression. This calculation was performed using SPSS (SPSS Inc., IL, USA). The gene expression data were subjected to hierarchical clustering analysis using Genowiz™ software (Ocimum Biosolutions).

Pathway analysis of PCR data

The behavior and relationships of marker genes in pathways associated with lipid metabolism were analyzed with bioSpace Explorer, a system for analysis of expression profile data. This system was produced collaboratively by Pharmafrontier Co. Ltd. and World Fusion Co. Ltd. to examine molecular interactions in expression profile data, using both manual and computational text mining.

RESULTS

Expression behavior of marker genes

All marker genes determined in this study are listed in Table 1. For each gene, since the quantitative limitation of biopsy specimens resulted in a difference in sample number for each probe, the type of samples, indicating the number of biopsy specimens, is listed in Table 2. Marker genes were selected as representative members of functional categories or molecular networks based on their expression changes in an experimental hepatic fibrosis model[3]. Marker genes with statistically significant changes in expression are listed in Table 1. Genes that showed statistically insignificant changes in expression are listed with the gene name only in Table 1. Marker genes that showed statistically significant changes by *t* test during a

Table 1 Expression profiles of marker genes with statistically significant changes during fibrosis progression

Functional category	Description	Gene name	Expression at F1	Group	Expression ratio			Serial	Type
					F2/F1	F3/F2	F4/F3		
ECM (or other HSC marker)	Decorin	DCN	217.6	2	1.2	1.1	1.1	1	1
	Matrix metalloproteinase 2	MMP2	11.7	2	1.5	1.2	0.9	2	1
	Hyaluronan-mediated motility receptor	HMMR	1.2	2	2.6	0.6	1.1	3	1
	Lysyl oxidase	LOX	1.0	2	1.4	1.2	0.9	4	1
	Lysyl oxidase-like 1	LOXL1	0.8	2	1.8	1.4	1.0	5	1
	Tropomyosin 1	TPM1	35.0	2	1.8	1.5	1.1	6	1
	Prion	PRNP	19.4	3	1.2	0.7	1.0	7	1
	Collagen, type I, alpha 1	COL1A1	16.2	2	1.3	2.2	0.6	8	1
	Collagen type Ⅲ alpha 1	COL3A1	143.9	2	1.3	1.5	0.8	9	1
	Collagen type alpha 1	COL4A1	16.8	2	1.3	1.9	0.6	10	1
	Humican	LUM	22.8	2	1.2	2.0	0.9	11	1
	Sialoprotein	SPP1	13.2	2	1.2	2.2	1.4	12	1
	Glypican 3	GPC3	23.5	2	1.8	2.5	1.0	13	1
	Proline 4-hydroxylase, alpha polypeptide I	P4HA1	34.0	1	1.1	0.4	1.2	14	1
	Insignificant change: MGP, BGN, TAGLN, LGALS1, EDG2, EDG5, TNNT2								
Inflammation (or apoptosis)	Lysozyme	LYZ	185.3	2	2.1	0.8	1.0	15	2
	TGF beta	TGFB1	51.9	3	1.5	0.8	0.9	16	1
	TGF beta 3	TGFB3	3.3	3	1.4	0.9	0.9	17	3
	TNF	TNF	2.6	3	1.7	0.6	1.2	18	3
	Natural killer cell proteinase 1	GZMB	1.7	3	1.6	0.5	1.1	19	1
	IL1 beta	IL1B	2.6	3	1.5	0.7	1.1	20	3
	Hemopoietic cell kinase	HCK	22.4	3	1.3	0.6	0.9	21	4
	Interleukin 6 receptor	IL6R	142.9	1	0.9	0.7	1.0	22	3
	BCL2-related ovarian killer	BOK	205.9	1	0.9	0.7	1.1	23	5
	Caspase 2	CASP2	1.3	1	1.2	0.7	0.8	24	5
	Chymase 1, mast cell	CMA1	0.2	2	0.7	1.8	2.2	25	6
	Insignificant change: LTBP1, LBP, TNFRSF1B, DEFB1, IL1RN, S100A8, BRIC3, CARD12, CASP1, CASP4, CASP8, PAWR, CD19, CD3Z, MS4A1, CD37, TRA[@]								
Growth factor	Growth hormone receptor	GHR	102.3	1	0.8	0.9	0.9	26	4
	IGF1	IGF1	56.5	1	0.9	0.7	1.1	27	1
	Insignificant change: PTN, FST, PRLR								
Insulin/ Wnt signal	Cyclin D1	CCND1	190.6	2	1.7	1.0	1.0	28	5
	Forkhead box M1	FOXM1	0.6	3	3.7	0.8	1.1	29	4
	Gap junction protein, alpha 1, 43 kDa (connexin 43)	GJA1	4.0	3	2.4	0.5	1.0	30	5
	V-akt murine thymoma viral oncogene homolog 1	AKT1	77.6	1	0.9	0.9	1.0	31	4
	V-akt murine thymoma viral oncogene homolog 2	AKT2	59.6	1	0.9	0.7	0.8	32	7
	Catenin (cadherin-associated protein), beta 1, 88 kDa	CTNNB1	159.0	1	1.1	0.7	1.0	33	5
	Catenin, beta interacting protein 1	CTNNBIP1	24.0	1	1.1	0.7	0.8	34	5
	Glycogen synthase kinase 3 beta	GSK3B	44.1	1	1.1	0.8	0.8	35	5
	Dishevelled, dsh homolog 1 (Drosophila)	DVL1	12.2	1	1.1	0.7	0.9	36	7
	Membrane-bound transcription factor peptidase, site 1	MBTPS1	72.4	1	0.9	0.7	1.0	37	7
	Membrane-bound transcription factor peptidase, site 2	MBTPS2	13.2	1	0.9	0.7	1.0	38	7
	Tribbles homolog 3 (Drosophila)	TRIB3	18.4	1	1.1	0.9	0.4	39	7
	Insignificant change: GSK3A, INSIG1, INSIG2, PRKCB1, PRKCD								
Others signal	Regucalcin (senescence marker protein-30)	RGN	697.8	1	0.9	0.8	0.8	56	4
	Insignificant change: DAB2, PMP22, S100A10, LCN2								
Transcription factors	CCAAT/enhancer binding protein (C/EBP), alpha	CEBPA	352.2	1	0.7	1.2	0.7	40	4
	Retinoid X receptor, alpha	RXRA	409.6	1	0.7	1.0	0.8	41	4
	Hepatocyte nuclear factor 4, alpha	HNF4A	708.7	1	0.7	1.1	0.9	42	4
	Transcription factor 1 (HNF1)	TCF1	26.7	1	0.8	0.9	0.8	43	4
	Nuclear receptor subfamily 0, group B, member 2	NR0B2	166.7	1	0.7	0.9	0.8	44	4
	Peroxisome proliferative activated receptor, alpha	PPARA	62.9	1	0.7	0.8	0.9	45	4
	Inhibitor of DNA binding 1 (splice variation)	ID1	463.4	3	1.6	0.5	1.4	46	1
	AE binding protein 1	AEBP1	24.6	2	1.4	1.2	1.0	47	4
	Nuclear receptor subfamily 1, group H, member 2	NR1H2 (LXRB)	9.0	1	1.2	0.7	1.0	48	7
	Nuclear receptor subfamily 1, group H, member 3	NR1H3 (LXRA)	27.8	1	1.0	0.7	0.8	49	7
	Nuclear receptor subfamily 1, group H, member 4	NR1H4 (FXR)	107.0	1	1.0	0.7	0.9	50	7
	c/EBPbeta	CEBPB	274.3	1	0.9	0.7	0.9	51	2
	Upstream transcription factor 2, c-fos interacting	USF2	323.7	1	0.9	0.7	0.9	52	4
	Estrogen-related receptor alpha	ESRRA	79.6	1	1.2	0.6	0.8	53	7
	C-met	MET	73.5	1	1.0	0.7	1.1	54	3
	Upstream transcription factor 1	USF1	30.8	1	1.0	0.9	0.8	55	4
	Insignificant change: ONECUT1, JUNB, NR3C1, PPARG, PPARGC1B, PPARGC1A, SREBF2, FHL2								
Transporter	Solute carrier family 6, member 6	SLC6A6	3.7	3	2.5	0.6	1.0	57	5
	Solute carrier family 7, member 1	SLC7A1	0.9	3	2.3	0.7	1.3	58	5
	Solute carrier family 38, member 2 Alanine-transporter)	SLC38A2	68.6	3	1.3	0.6	1.0	59	5
	Solute carrier family 25 member 15	SLC25A15	85.7	1	1.1	0.7	0.9	60	5
	Solute carrier family 7, member 7	SLC7A7	7.1	1	1.1	0.6	1.0	61	5
	Solute carrier family 17 (sodium phosphate), member 1	SLC17A1	23.6	1	0.9	0.7	1.0	62	4

	Insignificant change: SLC38A3, ABCB1, SLC15A4								
Redox	Catalase	CAT	1977.7	1	0.8	0.9	1.0	63	4
	Paraoxonase 1	PON1	191.1	1	0.8	0.7	0.9	64	4
Blood coagulation	Coagulation factor X	F10	187.5	1	0.8	1.0	0.9	65	4
	Angiotensinogen	AGT	1958.4	1	0.8	1.0	1.0	66	4
	Fibrinogen, A alpha polypeptide	FGA	8337.5	1	0.6	1.0	0.9	67	4
	Plasminogen	PLG	6156.1	1	0.8	0.8	1.0	68	4
	Pai I	SERPINE1	11.6	3	0.9	2.9	0.5	69	1
Lipid metabolism	Acyl-Coenzyme A oxidase 2, branched chain	ACOX2	122.2	1	0.8	0.9	0.9	70	4
	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	HADHSC	139.8	1	0.8	0.9	0.9	71	4
	Acyl-CoA synthetase long-chain family member 1	ACSL1	2800.8	1	0.8	0.9	0.9	72	4
	Acyl-Coenzyme A oxidase 1, palmitoyl	ACOX1	357.7	1	0.7	0.9	1.0	73	4
	Carnitine O-octanoyltransferase	CROT	10.3	1	0.7	0.8	1.1	74	4
	2,4-dienoyl CoA reductase 2, peroxisomal	DECR2	135.2	1	0.8	0.9	0.9	75	4
	Acetyl-Coenzyme A acyltransferase 2	ACAA2	1167.7	1	0.8	0.8	1.1	76	4
	Acetyl-Coenzyme A acetyltransferase 1	ACAT1	714.2	1	0.7	1.1	0.8	77	4
	Acyl-CoA synthetase long-chain family member 5	ACSL5	57.3	1	0.8	0.9	1.0	78	4
	Dodecenoyl-Coenzyme A delta isomerase	DCI	327.2	1	0.9	0.8	0.9	79	4
	Enoyl Coenzyme A hydratase, short chain, 1, mitochondrial	ECHS1	324.3	1	0.7	1.0	0.9	80	4
	Hydroxyacyl-Coenzyme A dehydrogenase, type II	HADH2	422.5	1	0.7	0.9	0.9	81	4
	Hydroxyacyl-Coenzyme A dehydrogenase, beta subunit	HADHB	398.8	1	0.8	0.9	1.1	82	4
	Lipase, hepatic	LIPC	561.5	1	0.5	1.5	0.9	83	4
	Hydroxyacyl-Coenzyme A dehydrogenase, alpha subunit	HADHA	103.2	1	0.9	0.9	1.0	84	4
	Palmitoyl-protein thioesterase 1	PPT1	82.1	1	1.0	0.8	0.8	85	4
	Fatty acid synthase	FASN	105.3	3	0.8	1.8	0.6	86	4
	Peroxisomal D3, D2-enoyl-CoA isomerase	PECI	292.8	1	0.8	0.8	0.8	87	4
	Acyl-CoA synthetase long-chain family member 4	ACSL4	5.8	2	1.1	1.4	2.3	88	4
	Insignificant change: BHHADH, ACAA1, CPT1A, ACADM, ACACA, CPT2								
Steroid (or drug) metabolism	Aldo-keto reductase family 1, member D1	AKR1D1	248.3	1	0.6	0.8	0.6	89	4
	HMT1 hnRNP methyltransferase-like 2	HRMT1L2	6.6	1	0.8	1.0	0.8	90	4
	Hydroxysteroid (11-beta) dehydrogenase 1	HSD11B1	1064.6	1	0.7	0.8	0.8	91	4
	Hydroxysteroid (17-beta) dehydrogenase 4	HSD17B4	66.7	1	0.8	0.9	0.9	92	4
	Steroid-5-alpha-reductase, alpha polypeptide 1	SRD5A1	74.3	1	0.7	0.8	0.9	93	4
	UDP glycosyltransferase 2 family, polypeptide B7	UGT2B7	530.3	1	0.8	0.8	0.9	94	4
	Sulfotransferase family 1E, estrogen-preferring, member 1	SULT1E1	52.8	1	0.5	0.5	1.4	95	7
	Aldo-keto reductase family 1, member C4	AKR1C4	137.1	1	0.9	0.6	1.2	96	4
	Hydroxysteroid (17-beta) dehydrogenase 2	HSD17B2	496.9	1	1.0	0.8	1.1	97	4
	Sulfotransferase family, cytosolic, 2A, member 1	SULT2A1	689.0	1	0.9	0.8	0.9	98	4
	Hydroxysteroid (17-beta) dehydrogenase 8	HSD17B8	72.2	1	0.7	0.7	1.0	99	7
	Steroid sulfatase (microsomal), arylsulfatase C, isozyme S	STS	10.9	1	1.0	1.0	0.6	100	4
	Emopamil binding protein (sterol isomerase)	EBP	197.0	1	1.0	0.7	0.7	101	5
	Farnesyl-diphosphate farnesyltransferase 1	FDFT1	252.6	1	0.7	0.9	0.7	102	5
	Insignificant change: HSD17B2, HSD3B1, LCMT1, SULT2A1, HMGCR, DHCR7, CES2								
Bile acid metabolism	Sterol O-acyltransferase 1	SOAT1	9.1	1	1.0	0.8	0.9	103	5
	Alcohol dehydrogenase 1C (class I), gamma polypeptide	ADH1C	312.7	1	0.7	0.6	1.1	104	5
	Alcohol dehydrogenase, iron containing, 1	ADHFE1	112.7	1	0.7	0.7	0.8	105	5
	Cytochrome P450, family 7, subfamily A, polypeptide 1	CYP7A1	104.1	1	0.7	0.6	2.9	106	5
Prostanoid	Arachidonate 5-lipoxygenase-activating protein	ALOX5AP	7.9	2	1.7	0.7	1.0	107	3
	Leukotriene B4 receptor 2	LTB4R2	2.5	1	1.1	0.7	1.0	108	3
	Insignificant change: LTA4H, CYSLTR1, CYSLTR2, LTC4S, PPT1								
Aromatic amino acid metabolism	Dopa decarboxylase	DDC	51.4	1	1.0	0.8	0.8	109	4
	Monoamine oxidase B	MAOB	426.6	1	0.8	0.8	1.0	110	4
	Kynurenine 3-monooxygenase	KMO	51.7	1	1.0	0.7	1.0	111	4
	Kynureninase	KYNU	63.3	1	0.9	0.7	0.8	112	4
	Tyrosine aminotransferase	TAT	658.2	1	0.9	0.5	1.3	113	4
	GTP cyclohydrolase 1	GCH1	49.6	1	0.9	0.7	0.9	114	4
	Insignificant change: HPD								
Sulfur-containing amino acid metabolism	MAT2	MAT2B	111.7	1	1.1	0.7	1.1	115	3
	Cystathionase (cystathionine gamma-lyase)	CTH	112.0	1	1.0	0.7	1.1	116	3
	Cystathionine-beta-synthase	CBS	350.4	1	1.1	0.7	0.9	117	3
	Betaine-homocysteine methyltransferase	BHMT	1032.1	1	1.0	0.7	1.0	118	3
	Methionine adenosyltransferase I, alpha	MAT1A	547.6	1	1.0	0.5	1.1	119	2
	Cysteine dioxygenase, type I	CDO1	84.4	1	0.9	0.7	1.0	120	1
	Glutamate-cysteine ligase, catalytic subunit	GCLC	175.9	1	1.1	0.6	1.1	121	1
	Glutathione S-transferase A1	GSTA1	2812.3	1	1.0	0.7	1.0	122	1
	Alanyl (membrane) aminopeptidase	ANPEP	501.9	1	0.9	0.8	0.9	124	5
	Bile acid Coenzyme A: amino acid N-acyltransferase	BAAT	284.4	1	1.0	0.6	1.0	125	5
	Glutathione synthetase	GSS	70.1	1	1.1	0.8	0.8	126	5
	Lactate dehydrogenase A	LDHA	716.3	1	0.8	0.7	1.0	127	5
	Mercaptopyruvate sulfurtransferase	MPST	684.3	1	0.9	0.7	0.9	128	5
	Serine dehydratase	SDS	440.4	1	0.9	0.4	1.0	129	5
	Methionine adenosyltransferase II, alpha	MAT2A	122.9	1	1.0	0.9	0.8	130	5
	Insignificant change: GGT1, GSR, MTR, DNMMT1, CSAD, GCLM, LDHB								

Energy source amino acid metabolism	Phosphoenolpyruvate carboxykinase 2 (mitochondrial)	PCK2	791.8	1	0.8	0.9	1.0	131	4
	Alanine-glyoxylate aminotransferase	AGXT	3352.3	1	0.7	0.9	1.0	132	4
	Alanine-glyoxylate aminotransferase 2	AGXT2	122.8	1	0.6	0.8	1.0	133	4
	Aldehyde dehydrogenase 2 family (mitochondrial)	ALDH2	2027.7	1	0.7	0.9	0.9	134	4
	Aldehyde dehydrogenase 9 family, member A1	ALDH9A1	160.9	1	0.8	0.8	0.9	135	4
	Pyruvate kinase, liver and RBC	PKLR	211.3	1	0.6	1.1	0.7	136	4
	Aldehyde dehydrogenase 3 family, member A2	ALDH3A2	237.4	1	0.9	0.8	1.1	137	4
	Phosphoenolpyruvate carboxykinase 1 (soluble)	PCK1	3166.8	1	1.0	0.5	1.2	138	4
	Dihydrolipoamide dehydrogenase	DLD	72.3	1	0.9	0.8	0.9	139	4
	Glutaminase 2 (liver, mitochondrial)	GLS2	133.8	1	0.8	0.7	1.0	140	4
	Glutamate-ammonia ligase	GLUL	6.2	1	1.0	0.6	1.2	141	4
	Glutamic-oxaloacetic transaminase 1, soluble	GOT1	388.2	1	0.9	0.7	0.8	142	4
	Glutamic-pyruvate transaminase	GPT	298.8	1	0.8	0.6	1.1	143	4
	Pyruvate carboxylase	PC	27.8	1	0.7	0.7	0.9	144	4
	Phosphoglucomutase 1	PGM1	218.8	1	0.8	0.8	1.0	145	4
	Pyruvate dehydrogenase kinase, isoenzyme 2	PDK2	24.0	1	0.9	0.8	0.9	146	7
	Pyruvate dehydrogenase kinase, isoenzyme 4	PDK4	85.6	1	0.7	0.5	1.4	147	7
	Fumarylacetoacetate hydrolase	FAH	85.8	1	0.9	0.9	0.8	148	4
	Malic enzyme 1, NADP (+) -dependent, cytosolic	ME1	4.9	2	1.0	1.0	1.3	149	4
	Insignificant change: ALDOA, ASNS, GOT2, MGC33309, PDHB, PDK1								
Glucose metabolism	Phosphorylase, glycogen; liver	PYGL	37.2	1	0.8	1.0	1.0	150	4
	Aldolase B, fructose-bisphosphate	ALDOB	13141.8	1	0.7	0.8	0.9	151	4
	Hexokinase 3 (white cell)	HK3	3.3	1	1.2	0.6	0.9	152	4
	Glycogen synthase 2 (liver)	GYS2	276.6	1	0.9	0.6	0.9	153	7
	Sterol regulatory element binding transcription factor 1	SREBF1	10.6	2	1.1	1.2	0.9	154	4
	Glycerol-3-phosphate dehydrogenase 1 (soluble)	GPD1	130.8	1	0.8	1.0	0.8	155	4
	Ketohexokinase (fructokinase)	KHK	411.8	1	0.7	1.0	0.8	156	4
	Glucokinase (hexokinase 4) regulator	GCKR	204.8	1	0.7	1.0	0.8	157	4
	Aldolase C, fructose-bisphosphate	ALDOC	36.8	1	0.7	0.9	0.6	158	4
	Insignificant change: GCK, PFKFB1, G6PC, HK2								
Urea cycle	Carbamoyl-phosphate synthetase 1, mitochondrial	CPS1	1230.8	1	0.7	0.9	0.9	159	7
	Ornithine aminotransferase (gyrate atrophy)	OAT	94.8	1	0.5	0.6	1.6	160	7
	Insignificant change: OTC								

Liver biopsy samples were analyzed by RT-PCR using a TaqMan® Gene Expression Assay probe, as described in the Materials and Methods. The genes were selected from DNA microarray data from a rat fibrosis model presented in a previous paper 3. Genes in the ECM category (or other HSC markers) were selected as marker genes for HSC. Genes in the inflammation category were selected as markers of inflammatory cells in liver. Other genes were selected as markers of hepatocytes. Expression profiles of marker genes that showed statistically significant changes during fibrosis by a Kruskal-Wallis test are listed. Genes that showed statistically insignificant changes in expression are listed with the gene name only in Table 1. Columns from left to right indicate functional category, gene description, gene name, expression intensity at F1 stage, group classification (expression of marker genes in group 1 decreased almost linearly along with the F stages, expression of marker genes in group 2 increased almost linearly along with the F stages, and expression of marker genes in group 3 had a peak in the middle of the F stages), expression ratio between neighbouring F stages (a decreased ratio is shown in light gray and an increased ratio in deep gray), serial gene number and types of samples, which indicates the number of biopsy specimens, as shown in the Table 2. Actual alteration behavior of some characteristic genes are shown in Figure 1.

Table 2 Types of samples indicating the number of biopsy specimens

Type of sample	Number of samples			
	F1	F2	F3	F4
1	25	13	14	9
2	23	13	14	8
3	22	12	14	8
4	12	6	9	4
5	7	4	5	4
6	6	4	4	3
7	5	4	5	5

Quantitative limitation of biopsy specimens resulted in a difference in sample number for each probe, according to the interest in the particular gene.

transition to an adjacent F stage are similar to the selected genes in Table 1. An additional four marker genes selected in the *t* test analysis were added to the genes in Table 1. Over half of the selected marker genes from the DNA microarray data from the animal model showed statistically significant changes in the human samples, showing the effectiveness of the animal model for selection of genes of significance in humans.

Genes showing changes in expression were roughly divided into three groups: genes in group 1 showed an almost linear decrease in expression along with an increase in F stage; genes in group 2 showed an almost linear increase with an increase in F stage; and genes in group 3 showed a peak in the middle of the F stage scale. Almost all genes in one category belonged to one or two groups, suggesting that genes in one category showed similar changes in expression during progression of fibrosis. The expression ratio between F stages is also shown in Table 1, with the peak ratio shown in bold. The peak for almost all genes in a given category occurred at the transition to the same F stage, again suggesting that genes in one category underwent changes in expression under similar mechanistic control.

Expression changes in gene clusters in the early phase (F1 to F2) of fibrosis

The functional categories showing a peak change in expression ratio in the early phase of fibrosis were inflammation, ECM, blood coagulation, lipid metabolism, half of the

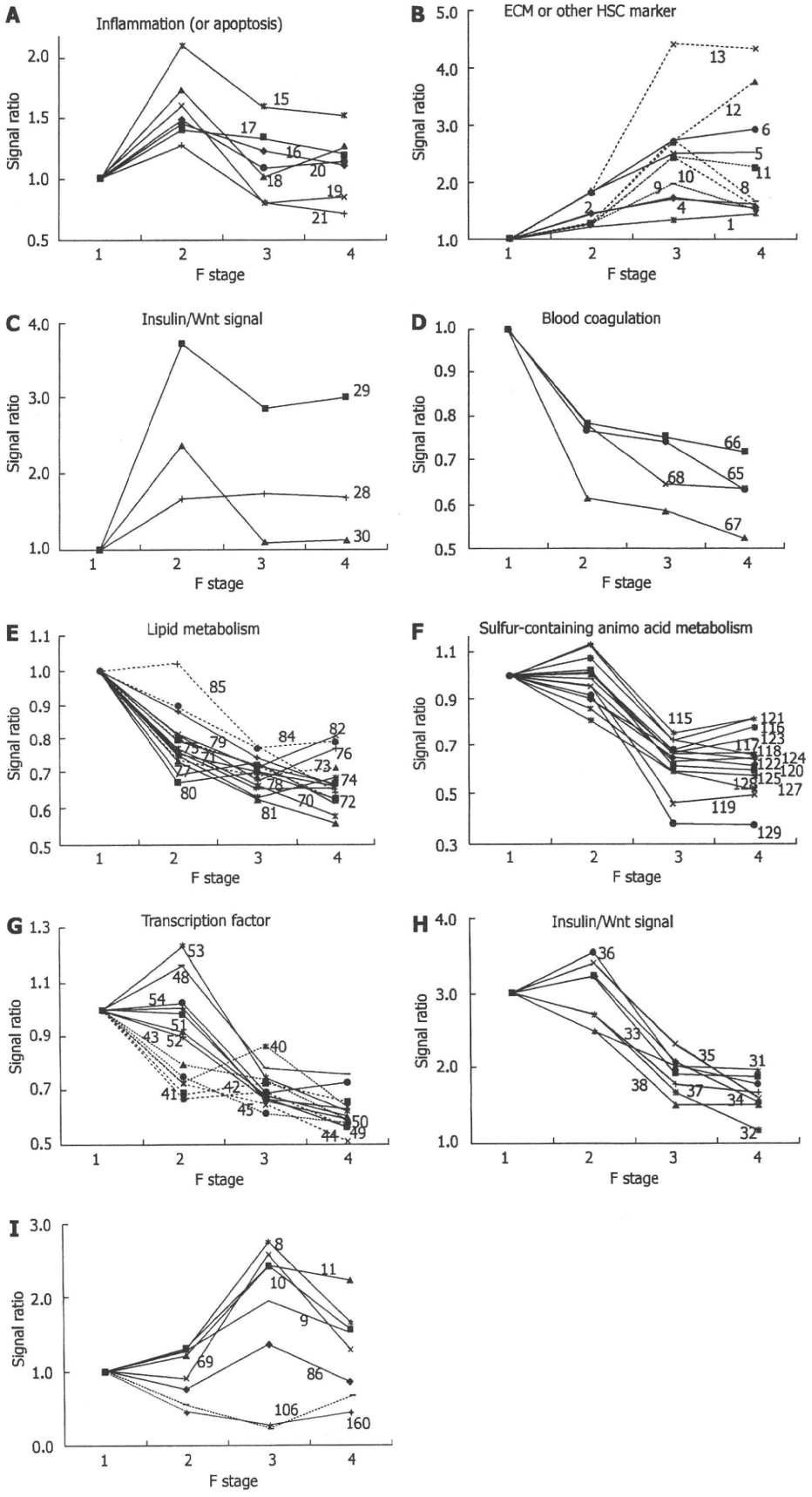

Figure 1 Clusters of genes in functional categories. The characteristic behavior of gene clusters for each functional category is shown separately in Figure 1A-I. The numbers in each figure refer to the serial number of genes in Table 1. The expression ratio relative to the F1 stage is plotted in each graph. **A**: Inflammation gene cluster with a peak at F2; **B**: ECM gene cluster showing increased expression along with fibrosis progression; **C**: Insulin/Wnt signaling gene cluster with increased expression in the early phase of fibrosis; **D**: Blood coagulation factor gene cluster showing decreased expression along with fibrosis progression; **E**: Lipid metabolism gene cluster showing decreased expression along with fibrosis progression; **F**: Sulfur-containing amino acid metabolism gene cluster showing a synchronous decrease in expression from F2 to F3; **G**: Two transcription factor gene clusters showing sequential decreases in expression; **H**: Insulin/Wnt signaling gene cluster showing decreased expression from F2 to F3; **I**: A gene cluster showing a peak or bottom at F3.

genes in steroid metabolism, half of those in energy source amino AA metabolism, and half of transcription factors. Expression of marker genes in the inflammation and ECM categories started to increase in the early phase. Expression

Table 3 Classification of gene clusters based on serial alteration of gene expression along with the fibrosis progression

F1-F2	F2-F3	F3-F4
Inflammation		
Wound-healing (ECM)		
Blood coagulation		
Transcription factors (cluster 1)	Transcription factors (cluster 2)	
	Insulin/Wnt signal	
Lipid metabolism		
Steroid metabolism (cluster 1)	Steroid metabolism (cluster 1)	
	Bile acid metabolism	
Energy source AA metabolism (cluster 1)	Energy source AA metabolism (cluster 1)	
	Aromatic AA metabolism	
	Sulfur-containing AA metabolism	
	Glucose metabolism (cluster 1)	Glucose metabolism (cluster 2)

Gene clusters are indicated under the name of gene category. The pattern of serial alteration for gene cluster was determined based on the maximum changing point of the expression ratio shown in Table 1.

of marker genes related to inflammation, such as LYZ, GZMB, IL1B, TNF and TGFB1, occurred in clusters and reached a peak at fibrotic stage F2, as shown in Figure 1A, suggesting that inflammatory events are particularly active at the F2 stage. However, histological classification of inflammatory activity shows a tendency for an increase in inflammation that is correlated with an increase in F stage; therefore, the conclusion regarding inflammatory events based on expression of marker genes appears to differ from that based on histological classification. Gene expression in the ECM category also increased until F3 or F4, as shown in Figure 1B; expression of such genes might indicate an inflammatory response for wound healing. Increased expression of some clustered genes related to the cell cycle, CCND1, FOXM1 and GJA1 (Connexin 43), was also found at an early stage, as shown in Figure 1C, and might reflect a response to hepatic cell injury.

Almost all other genes were linearly down-regulated. Several genes related to blood coagulation, i.e. F10, AGT, FGA and PLG, were down-regulated as a cluster in the early phase, as shown in Figure 1D; this down-regulation may be linked to prolongation of the blood coagulation time in cirrhosis. An early response of many clustered genes associated with lipid metabolism was also found, as shown in Figure 1E. Expression of these genes decreased consecutively during fibrosis and the early response of genes affecting lipid metabolism is of interest.

Expression changes in gene clusters in the middle phase (F2 to F3) of fibrosis

The peak change in the expression ratio of marker genes in metabolism of sulfur-containing AA and aromatic AA was delayed, compared to genes associated with other kinds of metabolism. Marker genes for sulfur-containing AA metabolism decreased remarkably as a cluster in the phase from F2 to F3, as shown in Figure 1F. Decreased expression of marker genes for steroid metabolism, energy source AA, and transcription factors were separable into two groups: early-response and middle-response genes.

All the down-regulated transcription factors, including TCF1 (HNF-1), HNF4A (HNF-4), CEBPA (C/EBP alpha), CEBPB (C/EBP beta), PPARA (PPAR alpha), RXRA (RXR alpha), NR1H3 (LXRA), NR1H2 (LXRB), NR1H4 (FXR), USF-1, USF-2, and NR0B2 are important in hepatic metabolism and other regulatory mechanisms. Alteration of expression of these genes might be related to abnormal expression of metabolic enzymes. Two clusters of transcription factors were clearly distinguishable based on their expression pattern, as shown in Figure 1G: the first cluster, including HNF-4, C/EBPA, RXR, TCF1 (HNF1), PPARA, and NR0B2, which showed altered expression in the early phase, might influence expression of the second cluster, including CEBPB, NR1H2(LXRB), NR1H3(LXRA), NR1H4(FXR), ESRRA and USF2, which showed altered expression in the middle phase of fibrosis.

A cluster of genes associated with insulin/Wnt signaling were down-regulated, as shown in Figure 1H, suggesting a common regulatory mechanism. These expression changes are likely to be related to changes in expression of transcription factors and genes in metabolic networks. The down-regulated genes included GSK3B and CTNNB1 (catenin beta 1), which participates in Wg/Wnt signaling for regulation of cell proliferation and differentiation[4]; GJA1 (connexin 43), which forms gap junctions and is regulated by Wg/Wnt signaling[5]; and FOXM1, which is associated with cell proliferation[6] and liver regeneration[7]. All of these genes had peak expression at F2 in a clustered manner, as shown in Figure 1C and G. Enhancement of cell proliferation for wound healing might be linked with a peak of inflammation at F2, and expression of genes such as CCND1, GJA1 and FOXM1 in the downstream part of the insulin/Wnt pathway were altered ahead of genes involved in insulin/Wnt signaling. The relationship between these genes requires further study.

Expression changes in gene clusters in the late phase (F3 to F4) of fibrosis

Few genes showed altered expression in the late phase of fibrosis, but a cluster of genes in the glucose metabolism category showed decreased expression. It was also of interest that expression of several genes reversed direction or abruptly altered in the late phase, as shown in Figure 1I. These results suggest different biological changes from the start of the late stage in the transition from F3 to F4.

Serial expression changes for the functional categories are summarized in Table 3 and these data indicate associations between clustered genes in one category and inter-category relationships.

Down-regulated individual marker genes (group 1)

Regeneration of hepatic cells is suppressed during fibrosis and such suppression is thought to then cause further fibrosis. IGF1, GHR and IL6R (inflammation) support the regeneration of hepatocytes and down-regulation of the expression of these genes may be linked directly to formation of fibrosis. Down-regulation of PON1, which associates with HDL (high-density lipoprotein)

and regulates the cellular redox state, and PPT, which is known as a lysosomal hydrolase of long chain fatty acyl CoA and has a role in maintenance of synaptic function, may be related to mitochondrial damage, as we and others have suggested[3,8,9]. RGN, a calcium-binding protein that plays a pivotal role in maintaining cell homeostasis and function, was also down-regulated. Down-regulation of these genes may impair liver function. TRIB3 inhibits Akt/PKB activation by insulin[10], and this gene was suddenly down-regulated from the F3 to F4 phase, suggesting new conditions in the insulin signaling pathway in the transition from F3 to F4.

Up-regulated individual marker genes (group 2)

Most inflammatory marker genes showed peak expression in the middle phase of fibrosis, as shown in Figure 1A, but CMA1 (chymase 1), which is produced by mast cells, underwent a linear increase in expression with progression of fibrosis. This is of note, since chymase has been reported to be involved in chronic hepatic fibrosis in autoimmune hepatitis and primary biliary cirrhosis[11], and mast cells may have a special role in fibrogenesis. The role of mast cells in chronic inflammation, however, deserves further study, because of the shortage of determined sample and low expression of CMA1.

Individual marker genes with peak expression in the middle phase of fibrosis (group 3)

The only hepatic stellate cell (HSC)-specific marker gene to show peak expression at the F2 stage was PRNP, which is reported to be a marker for the early phase of HSC activation. The amino acid transporters SLC38A2 (ATA2), SLC6A6 (TAUT) and SLC7A1 (CAT-1) showed peak expression at F2, which may also suggest enhancement of cell proliferation at this stage of fibrosis. Increased expression of SLC38A2, which preferentially transports alanine, has been reported in regeneration of hepatocytes[12], since hepatocytes require alanine as an energy source[13]. Based on our results, down-regulation of SLC38A2 in the late phase of fibrosis suggests that utilization of alanine as an energy source decreases at this stage of fibrosis. Up-regulation of SLC6A6, a taurine transporter, in the early phase of fibrosis can be understood as protective behavior against injury of hepatocytes[14]. FASN, a marker gene for fatty acid metabolism, showed peak expression at the F3 stage of fibrosis. We have observed suppression of biosynthesis and degradation of fatty acids in the liver in a CCl4-induced cirrhotic rat model, and clinical results show temporal enhancement of fatty acid biosynthesis before subsequent suppression in an advanced phase of fibrosis; this may be a compensative action related to suppression of other forms of energy metabolism in fibrosis, as we have previously suggested[3].

Candidate marker genes for diagnosis of fibrosis and discrimination of fibrotic stages

Diagnosis in the early stage of fibrosis may be important for monitoring progression of fibrosis and hepatic biological changes. Candidate marker genes at each step of fibrosis were selected based on an expression change ratio > 1.5 and an intensity > 10. Up-regulation of LYZ and down-regulation of FGA, OAT and AGXT2 were noteworthy in the transition from F1 to F2, and up-regulation of genes in the ECM category and down-regulation of genes in metabolism of energy source AA, aromatic AA, steroids and sulfur-containing AA occurred in the transition from F2 to F3. Diagnosis of the late stage of fibrosis (the transition from F3 to F4) is important because of the risk of tumorigenesis. Some genes showed a reversal of expression in the F3 to F4 transition, suggesting that biological changes in the stage from F3 to F4 are qualitatively different from those at earlier stages. Remarkable down-regulation of TRIB3, which inhibits insulin signaling and NFκB signaling, was noteworthy in the F3 to F4 transition. The reversal in expression of FASN in this phase may indicate changes in lipid metabolism and that of OAT indicates changes in ornithine metabolism or the urea cycle. The early increase in collagen expression began to decline in the F3 to F4 stage and the similar decline of SERPINE1 (Pai1) expression may reflect increased fibrolysis. Consecutive analysis of marker molecules in plasma will be important for monitoring progression of fibrosis at each step, and the data in this paper provides useful information for the selection of serum markers and interpretation of changes in the levels of these markers.

Coordinated regulation of functional categories

All genes in Figure 1 were subjected to hierarchical clustering analysis. Statistical clustering of the expression ratio between neighboring F stages for the genes in Table 1 was combined with functional categories, as shown in Figure 2. Functional categories clearly corresponded to the statistical clusters, suggesting coordinated regulation of genes in one functional category. Overlap of functional categories in statistical clusters suggested that these categories might be regulated by correlated mechanisms. In contrast, separation of members of a category into multiple positions of a statistical cluster suggests that the functional category may be divided into subgroups with respect to regulation.

DISCUSSION

Biological interpretation of changes in gene expression

Changes in expression of hepatic cell-specific marker genes reflects biological changes in the progression of hepatic fibrosis. Shimizu *et al*[15] reported co-localization of chymase with fibrotic tissue, and we have reported increased expression of marker genes for mast cells, including chymase, in progression of fibrosis in a DNM-induced rat fibrosis model. The results reported here also show correlation of the expression of chymase with the stage of fibrosis. Therefore, these results suggest that a certain cell population producing chymase has an important role in the pathogenesis of fibrosis.

Many marker genes related to inflammation showed peak expression at F2. Inflammation has been reported to induce activation of HSCs[16], and we have shown a peak in activated inflammatory cells in a DMN-induced fibrosis model. These results suggest that a temporal

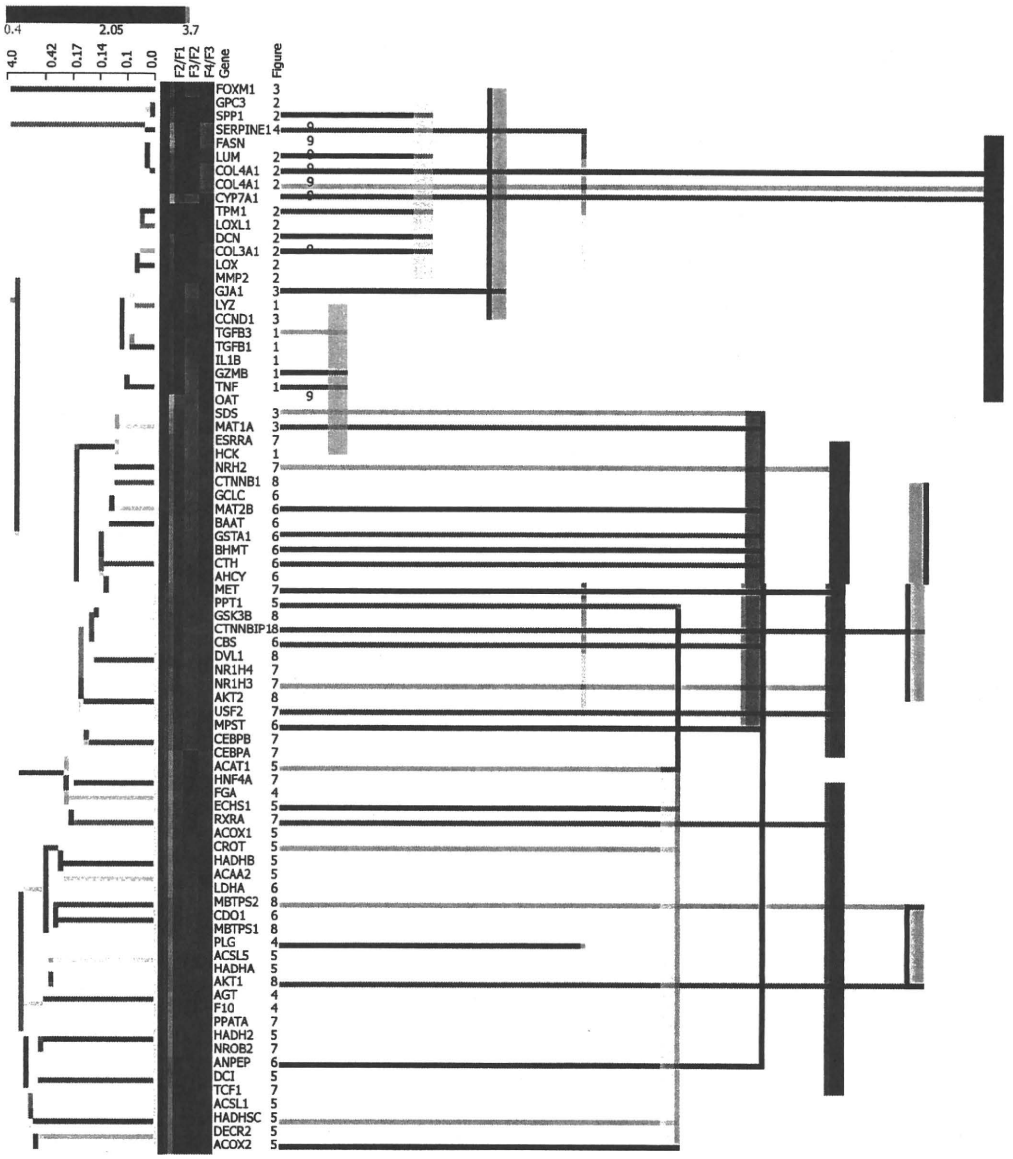

Figure 2 Comparison between statistical clustering analysis and functional cluster. All genes in Figure 1 were subjected to hierarchical clustering analysis using Genowiz™. Statistical clustering of the expression ratio between neighboring F stages for the genes in Table 1 was combined with functional categories. The color coding in the center column shows the expression ratio between neighboring F stages for the genes in Table 1: Ratio increases from green to red. The left dendrogram shows the result of hierarchical clustering by Genowiz™ software. The numbers on the right of gene names indicate the subgraph numer in Figure 1, which represents functional categories, and the same numbers are tied with color bars.

strong inflammatory reaction is required for initiation of auto-stimulatory activation of HSCs. In the DMN-induced fibrosis model, inflammation was suppressed after it reached a peak, but weaker inflammation might still be sufficient to maintain activation of HSCs.

Expression of HSC-specific marker genes such as COL1A1, LUM, TAGLN, MGP, SERPINE1 (PAI-1) and LOX did not increase or decrease continuously from F3 to F4, as shown in Figure 1B and 1I, whereas expression of DCN, BGN and GPC3 increased comparatively from F3 to F4, as shown in Figure 1B. This suggests some changes in the behavior of HSCs between F3 and F4, and it has been hypothesized that the perpetuation phase[17] may change to a new phase in fibrosis of stage F4. Therefore, profiles of HSC-specific marker genes may be useful for indication of the phase of fibrosis.

A cluster of down-regulated genes associated with sulfur-containing AA metabolism is shown in Figure 1F. Changes in expression of the clustered genes occurred simultaneously between the F2 and F3 stages; this transition may reflect a biological change causing fibrotic progression, and the change in expression of the clustered genes seems to suggest an increased demand for glutathione[18]. Abnormalities in metabolism of sulfur-containing AA in thioacetamide-induced cirrhosis in rat liver have been found in a proteomics analysis[19], and similar abnormalities were also reported based on hepatic gene expression changes in patients with alcoholic hepatitis[20].

There is only limited information on altered expression of transcription factors in fibrotic liver, but down-regulation

of HNF-4 in human cirrhosis[21] and of PPARs in hepatitis C virus genotype 3[22] have been reported. C/EBP alpha and C/EBP beta regulate proliferation of hepatocytes[23] and glucose and lipid homeostasis[24-27], and HNF-1 and HNF-4 broadly regulate hepatic functions such as carbohydrate metabolism[28], lipid metabolism[29], bile acid metabolism and HDL-cholesterol metabolism[30]; expression of HNF-1 is also regulated by HNF-4[31]. USF1 and USF2 have been reported as glucose signals[32,33], and the USF1 and USF2 homodimers and the USF1-USF2 heterodimer regulate expression of liver-specific genes such as apolipoprotein A2 and pyruvate kinase. HNF-4 and USF2a bind to the enhancer sequence cooperatively[34]. HNF-4 also regulates PPAR alpha[35], which in turn regulates glucose[36], lipid[37] and cholesterol metabolism[38]. RXR alpha regulates lipid, bile acid and cholesterol homeostasis[39], and LXR alpha, LXR beta and FXR are associated with lipid[40,41], bile acid and cholesterol homeostasis[40,42]. RXR and FXR form heterodimers with other transcription factors, including other members of the same family or with PPAR alpha[40,43,44]. NR0B2 (SHP) regulates cholesterol metabolism[45], glucose metabolism[45] and bile acid synthesis[46], and interacts with LXR[47]. AKT1 and AKT2 are important kinases in the pathway of insulin regulation of glucose homeostasis[48] and in fatty acid synthesis[49]. Hence, the altered expression of these transcription factors may relate to altered expression of metabolic enzymes in glucose and lipid metabolism in the fibrotic liver of hepatitis C patients.

The continuous increase in expression of cyclin D1 (CCND1) correlated with F stage, as shown in Figure 1C, and appears to be important for hepatic tumorigenesis. The association of Wnt signaling[50,51] and cyclin D[52] with tumorigenesis is well known. Catenin beta 1 (CTNNB1) regulates cyclin D expression[53] and is itself regulated through phosphorylation by GSK3B[54] or other kinases[55]. Down-regulation of expression of CTNNB1 occurred in advanced fibrotic stages, as shown in Figure 1H, and beta interacting protein 1 (CTNNBIP1), which is a negative regulator of CTNNB1, was also down-regulated, as also shown in Figure 1H. From this perspective, it was interesting that alterations in expression of genes associated with insulin signaling, including GSK3B, CTNNB1, CTNNBIP1 and downstream genes such as CCND1 and GJA1 (connexin 43), were clustered, as shown in Figures 1C and H. Suppression of insulin signaling has been reported in cirrhosis[56]; however, the response of downstream genes in this pathway was inconsistent with suppression of the insulin signal. Therefore, further studies will be necessary to clarify whether this inconsistency arises from differences between expression levels and the activity of the protein products of these genes, or if another signal[57] is involved in the Wnt and insulin signaling pathways.

The bioSpace Explorer is a system for analysis of DNA microarray data that may allow an understanding of the molecular relationships underlying the above results. The network of genes from previous reports revealed integral relationships among insulin/Wnt signal, CEBPA, PPARA, RXRA, glucose metabolism, lipid synthesis and lipid metabolism. An initial version of bioSpace Explorer was constructed based on molecular networks related to lipid metabolism, with entries showing relationships between molecules *via* a line between the molecules. This bird's eye view of the pathway including lipid metabolism with the input PCR data is illustrated in Figure 3; molecules related to inflammation were up-regulated and many genes related to lipid metabolism were down-regulated.

Hypothetical causes of biological changes in progression of fibrosis

Biological changes in fibrosis can be summarized as follows. Initially, Kupffer cells or other inflammatory cells are activated in the transition from F1 to F2. This event immediately influences production of ECM and cell cycle genes for wound-healing[58]. Blood coagulation is quickly suppressed in moving from F1 to F2, as shown by down-regulation of coagulation factor genes and up-regulation of the inhibitor, PAI-1. Several transcription factor genes are also immediately influenced, probably due to inflammation, as shown by the down-regulation of CEBPA, HNF4A, TCF1 and NR0B2. Expression of many genes associated with lipid metabolism also changed quickly in the transition from F1 to F2. Down-regulation of these genes may be controlled by down-regulation of the transcription factors, especially RXRA, PPARA, LXRs and FXR. Some genes related to steroid metabolism also responded quickly for control of inflammation.

Expression of genes associated with sulfur-containing AA metabolism and aromatic AA metabolism changed simultaneously in the transition from F2 to F3. The first type of metabolism relates to the redox state[59] and the second is associated with production of active metabolites such as catecholamines and serotonin. Such important biological states are controlled to maintain homeostasis through several mechanisms[60] and this may explain the delayed change in expression of these genes. The expression of many genes related to intracellular signaling, including insulin/Wnt signaling, also changed simultaneously in the transition from F2 to F3. This delayed change may also reflect compensative action for hepatic cellular defects on metabolism for energy supply and/or hepatic cellular proliferation. The main molecules in fibrosis, such as collagens, increase in expression from F2 to F3 and cause development of fibrosis through activation of HSCs through a stimulatory cycle involving inflammatory cells, HSCs and hepatocytes, as described previously[3].

Some quantitative biological changes started in the transition from F3 to F4. Down-regulation of genes associated with sugar metabolism and fatty acid synthesis at this stage might induce persistent defects in energy storage and supply to the liver. The liver transits into an inescapable negative cycle between this defect in the hepatic energy state and mitochondrial damage in cirrhosis. This negative cycle will be discussed in a future paper describing DNA microarray analysis of an animal model of cirrhosis.

Coordinated regulation of functional categories

Statistical cluster analysis showed coordinated regulation of functional categories in liver fibrosis. These regulatory mechanisms can be examined prospectively using bioSpace Explorer, and these results will be discussed in a future paper. In the current work, the clinical gene expression profiles assembled using RT-PCR, using genes

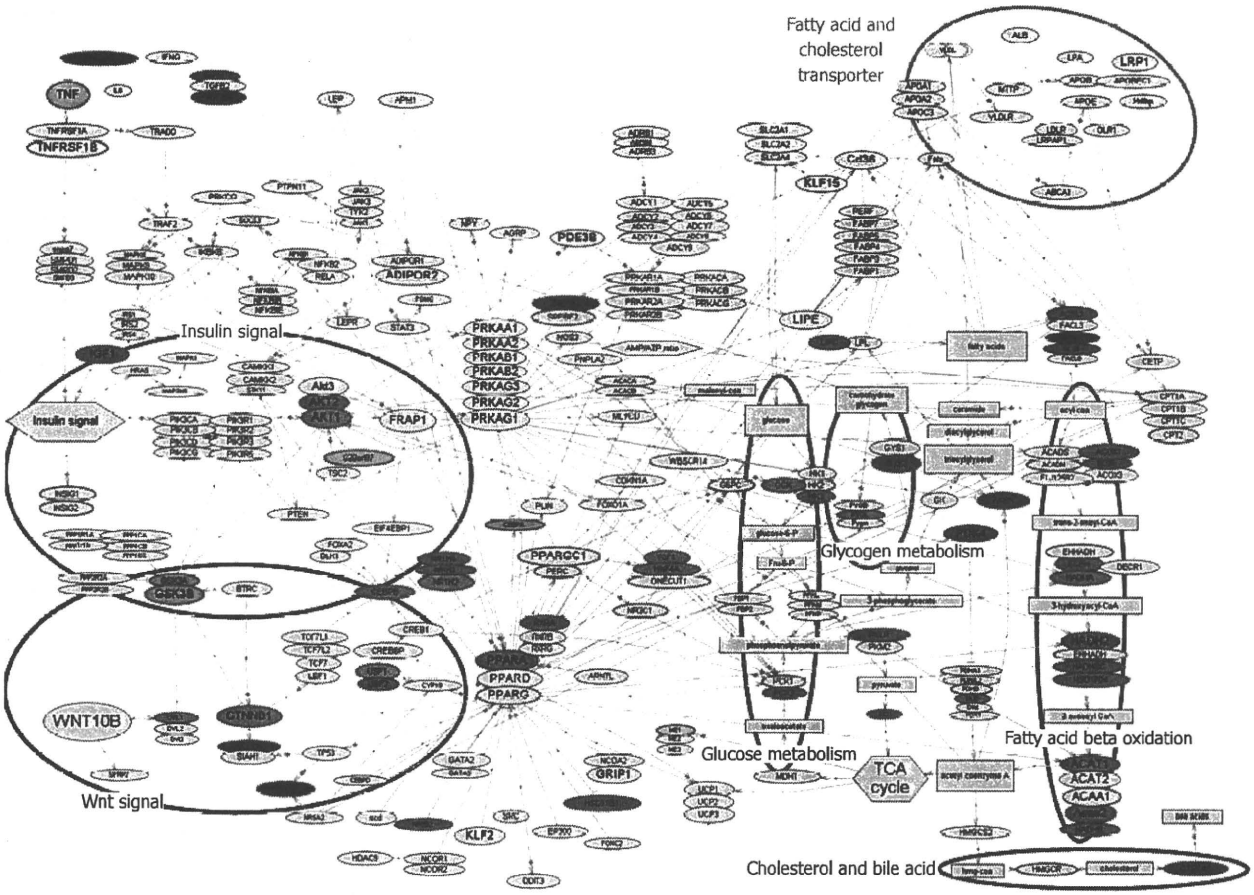

Figure 3 Molecular network associated with lipid metabolism. Gene expression changes in pathways related to lipid metabolism, illustrated with bioSpace Explorer (a system for analysis of DNA microarray data for lipid metabolism; Pharma Frontier Co. Ltd / World Fusion Co. Ltd.; see texts for details). The bird's-eye view of the lipid metabolism is displayed. The up-regulated and down-regulated gene expression ratios at F3 *vs* F1 in Table 1 are displayed in red and green, respectively, with the color gradation proportional to the ratio. Genes in Table 1 that did not show a statistically significant change in expression are indicated with a blue circle with gray background. An entry with a gray background only indicates no input data. Most of the entity names in Figure 2 are the same as the gene name in Table 1, but the names "C20orf97", "SCEH", "Acaa2", and "ACS5" in Figure 2 refer to "TRIB3", "ECHS1", "ACAA2", and "ACSL5", respectively, in Table 1. These differences are due to the software used in bioSpace Explorer.

originally selected based on DNA microarray data from an experimental animal model, provided an improved understanding of disease and suggested new methods of diagnosis. Therefore, statistical analysis and functional clustering or network analysis of the transcriptome, alone or in combination, can provide an overview of a dynamic biological system.

ACKNOWLEDGMENTS

We thank Ms. Tomoko Kobayashi of Pharmafrontier Co. Ltd. and Mr. Ken Osaki of World Fusion Co. Ltd for statistical analysis and operation of the bioSpace Explorer for analysis of DNA microarray data for lipid metabolism.

COMMENTS

Background

Information from clinical specimens is very important, but there is a limitation in sample preparation. With DNA microarray, it is difficult to determine gene expression profile from a small amount of sample such as a clinical needle biopsy sample. RT-PCR with TaqMan probe can make it possible with high quality. We have to get maximum information with a minimum number of TaqMan probes because the amount of samples is limited and probes are expensive. Selection of probes is a key factor. We extensively determined gene expression profiles from animal models of liver fibrosis with DNA microarray (*WJG* 2006; 12: 6473). The gene marker sets were arranged to show the change of gene expression of a molecular network or functionally clustered gene sets. In this paper, the selected gene marker sets effectively showed the dynamic behavior of global gene network change during liver fibrosis.

Research frontiers

Dynamic behavior of genome-wide genes expression is now measured with DNA microarray. This technology must be applied to clinical samples. Such information can greatly advance the study of disease pathogenesis, diagnosis and therapy. One problem is the interpretation of huge expression profile data from DNA microarray. Advanced technology of computational text-mining has recently shown the genome-wide molecular networks or functional molecular clusters. This genome-wide network is going to be applied to the analysis of DNA microarray data. When expression profile data are arranged as a change in their networks of functional clusters, these huge data are expected to suggest effectively the biological meaning in terms of broad aspects of research interest. Therefore, we are developing an analysis system genome-wide molecular network which was made possible by a combination system based on both computational and manual text-mining. This system was partially applied on the interpretation of gene expression profile in this paper. Prevention therapy for individual patients at an early stage is required because genomic polymorphism is going to reveal the personal risk of diseases. The biological background of progression to disease onset must be understood for development of prevention therapy. Prediction of liver cancer risk is going to be tried in the early

stage of liver fibrosis before cirrhosis. The accumulation process of hepatic stress which leads to onset of liver cancer has to be elucidated. For example, there is a question why BCAA (branched chain amino acids), which improves hepatic metabolism, reduces onset of liver cancer.

Innovations and breakthroughs

We have already prepared the gene marker sets which can show the change of molecular networks or functionally clustered genes by DNA microarray experiment on liver fibrosis of animal models. The gene marker sets were linked to biological events in each hepatic cell such as hepatocytes, immune cells and stellate cells. Application of gene marker sets and RT-PCR with TaqMan probe technology on clinical specimens successfully showed the serial change of gene expression in molecular network or functionally clustered gene sets in the progression of liver fibrosis. It was proved that network analysis is a powerful tool for biological research. Our sequential approach (animal model/DNA microarray → selection of appropriate gene marker sets in molecular networks → clinical samples/RT-PCR → analysis functionally clustered gene markers in molecular networks → analysis of the relation between molecular networks) can effectively advance clinical research. Serial change of clustered gene expression during liver fibrosis progression, which was made clear in this paper, will reveal the molecular mechanism of many symptoms before and after the onset of cirrhosis and liver cancer.

Application

Gene marker sets and RT-PCR on clinical specimens as well as analytical methods with genome-wide gene networks can be applied to get the information of dynamic biological progression on various diseases. Serial change of clustered gene expression during liver fibrosis progression can provide a method of diagnosis and therapy in liver fibrosis. For example, the question why BCAA, which improves hepatic metabolism, reduces onset of liver cancer, will be solved.

Terminology

Text-mining: all published information about molecular interaction and their function are collected and arranged to make new valuable information such as gene-wide molecular networks; TaqMan-PCR probe: Applied Biosystems offers more than 700 000 TaqMan® Gene Expression Assays, the most comprehensive set of pre-designed Real-Time PCR assays available. All TaqMan® Gene Expression Assays have been designed using validated bioinformatics pipeline of Applied Biosystems, and run with the same PCR protocol, eliminating the need for primer design or PCR optimization.

Peer review

This paper revealed that metabolic deficiency occurs before the onset of cirrhosis. It had already been found in animal models with hepatic toxic substances in the preceding paper. Metabolic deficiency in hepatitis which was caused by a virus was found to be the same as animal models in this paper. Gene marker sets, selected from analysis of animal models, and analysis methods using molecular networks can lead to success in finding the serial change of functionally clustered genes expression during liver fibrosis progression. A set of appropriate gene markers in each network was a key to analysis. The sequential approach (animal model/DNA microarray → appropriate gene marker sets in molecular networks → clinical samples/RT-PCR → analysis functionally clustered gene markers in molecular network → analysis of the relation between molecular networks) is useful to elucidate the molecular mechanism of disease.

REFERENCES

1 **Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C**. The French METAVIR Cooperative Study Group. *Hepatology* 1994; **20**: 15-20

2 **Gobel T**, Vorderwulbecke S, Hauck K, Fey H, Haussinger D, Erhardt A. New multi protein patterns differentiate liver fibrosis stages and hepatocellular carcinoma in chronic hepatitis C serum samples. *World J Gastroenterol* 2006; **12**: 7604-7612

3 **Takahara Y**, Takahashi M, Wagatsuma H, Yokoya F, Zhang QW, Yamaguchi M, Aburatani H, Kawada N. Gene expression profiles of hepatic cell-type specific marker genes in progression of liver fibrosis. *World J Gastroenterol* 2006; **12**: 6473-6499

4 **Dierick H**, Bejsovec A. Cellular mechanisms of wingless/Wnt signal transduction. *Curr Top Dev Biol* 1999; **43**: 153-190

5 **Ai Z**, Fischer A, Spray DC, Brown AM, Fishman GI. Wnt-1 regulation of connexin43 in cardiac myocytes. *J Clin Invest* 2000; **105**: 161-171

6 **Wang IC**, Chen YJ, Hughes D, Petrovic V, Major ML, Park HJ, Tan Y, Ackerson T, Costa RH. Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. *Mol Cell Biol* 2005; **25**: 10875-10894

7 **Wang X**, Bhattacharyya D, Dennewitz MB, Kalinichenko VV, Zhou Y, Lepe R, Costa RH. Rapid hepatocyte nuclear translocation of the Forkhead Box M1B (FoxM1B) transcription factor caused a transient increase in size of regenerating transgenic hepatocytes. *Gene Expr* 2003; **11**: 149-162

8 **Ferre N**, Marsillach J, Camps J, Mackness B, Mackness M, Riu F, Coll B, Tous M, Joven J. Paraoxonase-1 is associated with oxidative stress, fibrosis and FAS expression in chronic liver diseases. *J Hepatol* 2006; **45**: 51-59

9 **Kim SJ**, Zhang Z, Lee YC, Mukherjee AB. Palmitoyl-protein thioesterase-1 deficiency leads to the activation of caspase-9 and contributes to rapid neurodegeneration in INCL. *Hum Mol Genet* 2006; **15**: 1580-1586

10 **Du K**, Herzig S, Kulkarni RN, Montminy M. TRB3: a tribbles homolog that inhibits Akt/PKB activation by insulin in liver. *Science* 2003; **300**: 1574-1577

11 **Satomura K**, Yin M, Shimizu S, Kato Y, Nagano T, Komeichi H, Ohsuga M, Katsuta Y, Aramaki T, Omoto Y. Increased chymase in livers with autoimmune disease: colocalization with fibrosis. *J Nippon Med Sch* 2003; **70**: 490-495

12 **Fowler FC**, Banks RK, Mailliard ME. Characterization of sodium-dependent amino acid transport activity during liver regeneration. *Hepatology* 1992; **16**: 1187-1194

13 **Freeman TL**, Ngo HQ, Mailliard ME. Inhibition of system A amino acid transport and hepatocyte proliferation following partial hepatectomy in the rat. *Hepatology* 1999; **30**: 437-444

14 **Warskulat U**, Borsch E, Reinehr R, Heller-Stilb B, Monnighoff I, Buchczyk D, Donner M, Flogel U, Kappert G, Soboll S, Beer S, Pfeffer K, Marschall HU, Gabrielsen M, Amiry-Moghaddam M, Ottersen OP, Dienes HP, Haussinger D. Chronic liver disease is triggered by taurine transporter knockout in the mouse. *FASEB J* 2006; **20**: 574-576

15 **Shimizu S**, Satomura K, Aramaki T, Katsuta Y, Takano T, Omoto Y. Hepatic chymase level in chronic hepatitis: co-localization of chymase with fibrosis. *Hepatol Res* 2003; **27**: 62-66

16 **Baroni GS**, Pastorelli A, Manzin A, Benedetti A, Marucci L, Solforosi L, Di Sario A, Brunelli E, Orlandi F, Clementi M, Macarri G. Hepatic stellate cell activation and liver fibrosis are associated with necroinflammatory injury and Th1-like response in chronic hepatitis C. *Liver* 1999; **19**: 212-219

17 **Gressner AM**. Transdifferentiation of hepatic stellate cells (Ito cells) to myofibroblasts: a key event in hepatic fibrogenesis. *Kidney Int Suppl* 1996; **54**: S39-S45

18 **Wu G**, Fang YZ, Yang S, Lupton JR, Turner ND. Glutathione metabolism and its implications for health. *J Nutr* 2004; **134**: 489-492

19 **Low TY**, Leow CK, Salto-Tellez M, Chung MC. A proteomic analysis of thioacetamide-induced hepatotoxicity and cirrhosis in rat livers. *Proteomics* 2004; **4**: 3960-3974

20 **Lee TD**, Sadda MR, Mendler MH, Bottiglieri T, Kanel G, Mato JM, Lu SC. Abnormal hepatic methionine and glutathione metabolism in patients with alcoholic hepatitis. *Alcohol Clin Exp Res* 2004; **28**: 173-181

21 **Berasain C**, Herrero JI, Garcia-Trevijano ER, Avila MA, Esteban JI, Mato JM, Prieto J. Expression of Wilms' tumor suppressor in the liver with cirrhosis: relation to hepatocyte nuclear factor 4 and hepatocellular function. *Hepatology* 2003; **38**: 148-157

22 **de Gottardi A**, Pazienza V, Pugnale P, Bruttin F, Rubbia-Brandt L, Juge-Aubry CE, Meier CA, Hadengue A, Negro F. Peroxisome proliferator-activated receptor-alpha and -gamma mRNA levels are reduced in chronic hepatitis C with steatosis and genotype 3 infection. *Aliment Pharmacol Ther* 2006; **23**: 107-114

23 **Greenbaum LE**, Cressman DE, Haber BA, Taub R. Coexistence of C/EBP alpha, beta, growth-induced proteins and DNA synthesis in hepatocytes during liver regeneration. Implications

for maintenance of the differentiated state during liver growth. *J Clin Invest* 1995; **96**: 1351-1365

24 **Arizmendi C**, Liu S, Croniger C, Poli V, Friedman JE. The transcription factor CCAAT/enhancer-binding protein beta regulates gluconeogenesis and phosphoenolpyruvate carboxykinase (GTP) gene transcription during diabetes. *J Biol Chem* 1999; **274**: 13033-13040

25 **Gautier-Stein A**, Mithieux G, Rajas F. A distal region involving hepatocyte nuclear factor 4alpha and CAAT/enhancer binding protein markedly potentiates the protein kinase A stimulation of the glucose-6-phosphatase promoter. *Mol Endocrinol* 2005; **19**: 163-174

26 **Qiao L**, MacLean PS, You H, Schaack J, Shao J. knocking down liver ccaat/enhancer-binding protein alpha by adenovirus-transduced silent interfering ribonucleic acid improves hepatic gluconeogenesis and lipid homeostasis in db/db mice. *Endocrinology* 2006; **147**: 3060-3069

27 **Wang ND**, Finegold MJ, Bradley A, Ou CN, Abdelsayed SV, Wilde MD, Taylor LR, Wilson DR, Darlington GJ. Impaired energy homeostasis in C/EBP alpha knockout mice. *Science* 1995; **269**: 1108-1112

28 **Pontoglio M**. Hepatocyte nuclear factor 1, a transcription factor at the crossroads of glucose homeostasis. *J Am Soc Nephrol* 2000; **11** Suppl 16: S140-S143

29 **Louet JF**, Hayhurst G, Gonzalez FJ, Girard J, Decaux JF. The coactivator PGC-1 is involved in the regulation of the liver carnitine palmitoyltransferase I gene expression by cAMP in combination with HNF4 alpha and cAMP-response element-binding protein (CREB). *J Biol Chem* 2002; **277**: 37991-38000

30 **Shih DQ**, Bussen M, Sehayek E, Ananthanarayanan M, Shneider BL, Suchy FJ, Shefer S, Bollileni JS, Gonzalez FJ, Breslow JL, Stoffel M. Hepatocyte nuclear factor-1alpha is an essential regulator of bile acid and plasma cholesterol metabolism. *Nat Genet* 2001; **27**: 375-382

31 **Miura N**, Tanaka K. Analysis of the rat hepatocyte nuclear factor (HNF) 1 gene promoter: synergistic activation by HNF4 and HNF1 proteins. *Nucleic Acids Res* 1993; **21**: 3731-3736

32 **Corre S**, Galibert MD. Upstream stimulating factors: highly versatile stress-responsive transcription factors. *Pigment Cell Res* 2005; **18**: 337-348

33 **Kahn A**. From the glycogenic function of the liver to gene regulation by glucose. *C R Seances Soc Biol Fil* 1998; **192**: 813-827

34 **Moriizumi S**, Gourdon L, Lefrancois-Martinez AM, Kahn A, Raymondjean M. Effect of different basic helix-loop-helix leucine zipper factors on the glucose response unit of the L-type pyruvate kinase gene. *Gene Expr* 1998; **7**: 103-113

35 **Pineda Torra I**, Jamshidi Y, Flavell DM, Fruchart JC, Staels B. Characterization of the human PPARalpha promoter: identification of a functional nuclear receptor response element. *Mol Endocrinol* 2002; **16**: 1013-1028

36 **Xu J**, Chang V, Joseph SB, Trujillo C, Bassilian S, Saad MF, Lee WN, Kurland IJ. Peroxisomal proliferator-activated receptor alpha deficiency diminishes insulin-responsiveness of gluconeogenic/glycolytic/pentose gene expression and substrate cycle flux. *Endocrinology* 2004; **145**: 1087-1095

37 **Lee SS**, Chan WY, Lo CK, Wan DC, Tsang DS, Cheung WT. Requirement of PPARalpha in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *J Lipid Res* 2004; **45**: 2025-2037

38 **Chakravarthy MV**, Pan Z, Zhu Y, Tordjman K, Schneider JG, Coleman T, Turk J, Semenkovich CF. "New" hepatic fat activates PPARalpha to maintain glucose, lipid, and cholesterol homeostasis. *Cell Metab* 2005; 1: 309-322

39 **Wan YJ**, Cai Y, Lungo W, Fu P, Locker J, French S, Sucov HM. Peroxisome proliferator-activated receptor alpha-mediated pathways are altered in hepatocyte-specific retinoid X receptor alpha-deficient mice. *J Biol Chem* 2000; **275**: 28285-28290

40 **Edwards PA**, Kennedy MA, Mak PA. LXRs; oxysterol-activated nuclear receptors that regulate genes controlling lipid homeostasis. *Vascul Pharmacol* 2002; **38**: 249-256

41 **Sinal CJ**, Tohkin M, Miyata M, Ward JM, Lambert G, Gonzalez FJ. Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* 2000; **102**: 731-744

42 **Lambert G**, Amar MJ, Guo G, Brewer HB Jr, Gonzalez FJ, Sinal CJ. The farnesoid X-receptor is an essential regulator of cholesterol homeostasis. *J Biol Chem* 2003; **278**: 2563-2570

43 **Vidal H**. PPAR receptors: recent data. *Ann Endocrinol* (Paris) 2005; **66**: 1S5-1S9

44 **Yoshikawa T**, Shimano H, Amemiya-Kudo M, Yahagi N, Hasty AH, Matsuzaka T, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Osuga J, Harada K, Gotoda T, Kimura S, Ishibashi S, Yamada N. Identification of liver X receptor-retinoid X receptor as an activator of the sterol regulatory element-binding protein 1c gene promoter. *Mol Cell Biol* 2001; **21**: 2991-3000

45 **Kim HJ**, Kim JY, Kim JY, Park SK, Seo JH, Kim JB, Lee IK, Kim KS, Choi HS. Differential regulation of human and mouse orphan nuclear receptor small heterodimer partner promoter by sterol regulatory element binding protein-1. *J Biol Chem* 2004; **279**: 28122-28131

46 **Ito S**, Fujimori T, Furuya A, Satoh J, Nabeshima Y, Nabeshima Y. Impaired negative feedback suppression of bile acid synthesis in mice lacking betaKlotho. *J Clin Invest* 2005; **115**: 2202-2208

47 **Brendel C**, Schoonjans K, Botrugno OA, Treuter E, Auwerx J. The small heterodimer partner interacts with the liver X receptor alpha and represses its transcriptional activity. *Mol Endocrinol* 2002; **16**: 2065-2076

48 **Cho H**, Thorvaldsen JL, Chu Q, Feng F, Birnbaum MJ. Akt1/PKBalpha is required for normal growth but dispensable for maintenance of glucose homeostasis in mice. *J Biol Chem* 2001; **276**: 38349-38352

49 **Ono H**, Shimano H, Katagiri H, Yahagi N, Sakoda H, Onishi Y, Anai M, Ogihara T, Fujishiro M, Viana AY, Fukushima Y, Abe M, Shojima N, Kikuchi M, Yamada N, Oka Y, Asano T. Hepatic Akt activation induces marked hypoglycemia, hepatomegaly, and hypertriglyceridemia with sterol regulatory element binding protein involvement. *Diabetes* 2003; **52**: 2905-2913

50 **Colnot S**, Decaens T, Niwa-Kawakita M, Godard C, Hamard G, Kahn A, Giovannini M, Perret C. Liver-targeted disruption of Apc in mice activates beta-catenin signaling and leads to hepatocellular carcinomas. *Proc Natl Acad Sci USA* 2004; **101**: 17216-17221

51 **Murata M**, Miyoshi Y, Ohsawa M, Shibata K, Ohta T, Imai Y, Nishikawa M, Iwao K, Tateishi H, Shimano T, Kobayashi T, Nakamura Y. Accumulation of beta-catenin in the cytoplasm and the nuclei during the early hepatic tumorigenesis. *Hepatol Res* 2001; **21**: 126-135

52 **Yamaoka H**, Ohtsu K, Sueda T, Yokoyama T, Hiyama E. Diagnostic and prognostic impact of beta-catenin alterations in pediatric liver tumors. *Oncol Rep* 2006; **15**: 551-556

53 **Kolligs FT**, Bommer G, Goke B. Wnt/beta-catenin/tcf signaling: a critical pathway in gastrointestinal tumorigenesis. *Digestion* 2002; **66**: 131-144

54 **Zeng X**, Tamai K, Doble B, Li S, Huang H, Habas R, Okamura H, Woodgett J, He X. A dual-kinase mechanism for Wnt co-receptor phosphorylation and activation. *Nature* 2005; **438**: 873-877

55 **Taurin S**, Sandbo N, Qin Y, Browning D, Dulin NO. Phosphorylation of beta-catenin by cyclic AMP-dependent protein kinase. *J Biol Chem* 2006; **281**: 9971-9976

56 **Picardi A**, D'Avola D, Gentilucci UV, Galati G, Fiori E, Spataro S, Afeltra A. Diabetes in chronic liver disease: from old concepts to new evidence. *Diabetes Metab Res Rev* 2006; **22**: 274-283

57 **Gotoh J**, Obata M, Yoshie M, Kasai S, Ogawa K. Cyclin D1 over-expression correlates with beta-catenin activation, but not with H-ras mutations, and phosphorylation of Akt, GSK3 beta and ERK1/2 in mouse hepatic carcinogenesis. *Carcinogenesis* 2003; **24**: 435-442

58 **Kershenobich Stalnikowitz D**, Weissbrod AB. Liver fibrosis and inflammation. A review. *Ann Hepatol* 2003; **2**: 159-163

59 **Nagahara N**, Katayama A. Post-translational regulation of mercaptopyruvate sulfurtransferase via a low redox potential cysteine-sulfenate in the maintenance of redox homeostasis. *J Biol Chem* 2005; **280**: 34569-34576

60 **Fitzpatrick PF**. The aromatic amino acid hydroxylases. *Adv Enzymol Relat Areas Mol Biol* 2000; **74**: 235-294

S- Editor Liu Y **L- Editor** Roberts SE **E- Editor** Liu Y

Tissue Antigens ISSN 0001-2815

BRIEF COMMUNICATION

PTPRC (CD45) variation and disease association studied using single nucleotide polymorphism tagging

B. J. Hennig[1], A. E. Fry[1], K. Hirai[2], H. Tahara[2], A. Tamori[2], M. Moller[3], J. Hopkin[3], A. V. Hill[1], W. Bodmer[4], P. Beverley[5] & E. Tchilian[6]

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
2 Department of Gynaecology, Osaka City University Graduate School of Medicine, Osaka, Japan
3 Experimental Medicine, School of Medicine, Swansea University, Swansea, UK
4 Cancer and Immunogenetics Laboratory, Weatherall Institute of Molecular Medicine, Cancer Research UK, John Radcliffe Hospital, Oxford, UK
5 CD45 Group, The Jenner Institute for Vaccine Research, Compton, UK
6 Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Oxford, UK

Key words
A138G; C77G; case control; disease association; Graves' disease; PTPRC (CD45); tuberculin response; type 1 diabetes

Correspondence
Elma Tchilian
Centre for Clinical Vaccinology and Tropical Medicine
University of Oxford
Old Road, Churchill Hospital
Oxford OX3 7LJ
UK
Tel: +44 1865 857429
Fax: +44 1865 857471
e-mail: elma.tchilian@ndm.ox.ac.uk

Received 2 October 2007; revised 4 January 2008; accepted 22 January 2008

doi: 10.1111/j.1399-0039.2008.01014.x

Abstract

CD45 is a haemopoietic tyrosine phosphatase, crucial for lymphocyte signalling. Two polymorphisms (C77G and A138G), which alter CD45 isoform expression, are associated with autoimmune and infectious diseases. Using HapMap data, we show that there is substantial linkage disequilibrium across the CD45 gene (PTPRC), with similar patterns in different populations. Employing a set of single nucleotide polymorphisms, correlated with a substantial proportion of variation across this gene, we tested for association with type 1 diabetes, Graves' disease in a Japanese population, hepatitis C in UK population and tuberculin response in a Chinese population. A limited number of common haplotypes was found. Most 138G alleles are present on only one haplotype, which is associated with Graves' disease, supporting previous data that A138G is a functionally important CD45 polymorphism.

The CD45 leukocyte common antigen, encoded by the PTPRC gene, is essential for normal lymphocyte function, and its absence is a cause of severe combined immunodeficiency (1). CD45 is subjected to complex alternative splicing of exons 4, 5 and 6. Polymorphisms in PTPRC that affect CD45 alternative splicing are associated with infectious and autoimmune diseases (2). The two best described human genetic variants, in exons 4 and 6, affect splicing and have contrasting geographical distribution, phenotypes and disease associations. The exon 6 A138G polymorphism (rs4915154 A191T) is present at around 20% in the East Asian populations, promotes splicing towards low-molecular-weight isoforms (CD45R0) and is associated with protection against Graves' disease, type 1 diabetes and hepatitis B and C virus infections (3, 4). In contrast, the exon 4 C77G variant (rs17612648 P57P) is found at around 2% in populations of European origin and prevents splicing to CD45R0 isoform. An excess of 77G individuals has been reported in HIV (5), multiple sclerosis (6–8), autoimmune hepatitis (9) and systemic sclerosis (10). PTPRC is a relatively large gene (118.1 kb), it has recently been surveyed as part of the International Haplotype Map (HapMap) Project (http://www.hapmap.org/) and dozens of single nucleotide polymorphisms (SNPs) have been validated at this locus. Most research to date has focused on the two SNPs A138G and C77G, but it is also possible that they are markers correlated with nearby, and as yet ungenotyped, functional polymorphisms with stronger signals of disease association. Therefore, our objectives were: (i) to use HapMap data as a resource to examine the

linkage disequilibrium (LD) structure around the A138G and C77G variants in relevant populations, (ii) to choose a subset of tagSNPs across the whole length of PTPRC, which could be used to test disease associations and (iii) to assess multiple-marker haplotypes across the PTPRC gene, which may be correlated with untyped variants that are associated with disease and immune response.

We began by examining the LD structure across the PTPRC locus in three HapMap populations (Release 21 July 2006), namely Japanese (JPT), Han Chinese (CHB) and US Caucasians (CEU) (11). HapMap data were visualised using HAPLOVIEW (version 3.32) (12), and tagging was carried out using the 'Tagger' algorithm (13). We appreciate that caution is required when making inferences from the HapMap datasets because of, for example, limited sample size. Furthermore, different LD metrics can affect the interpretation of LD structure. Using a common block definition (13), we found substantial LD across the whole length of the PTPRC gene (Figure 1), with similar patterns in the East Asian samples and to some extent in the CEU samples. In these populations, there appear to be three major LD blocks, although the LD break points beyond block 1 differ between the Asian and CEU populations. We also noted an area of somewhat higher LD at the 5′end of the gene in CEU and possibly lower LD in the central CHB block. The first block spans the 5′ untranslated region, exons 1 and 2 with the intervening first intron, which includes elements with promoter activity (14), as well as most of the 50 kb second intron. Exons 3–15, coding for the extracellular domain, are covered by the second block, which extends further in JPT/CHB populations than in the CEU population. Most of the known coding variations occur in this region, including C77G and A138G. The extracellular domain is the platform for complex polysaccharides and is known to bind several lectins (15). There is evidence suggesting that this region is under selection by pathogens in a number of vertebrate species including primates (16, 17). The 3′ end of the gene (exons 17–33), covered by the third LD block and which is larger in the CEU population than in the JPT/CHB populations, codes for two phosphatase domains. The first membrane proximal domain has phosphatase activity, while the distal domain is involved in molecular interactions (18). We also assessed whether the PTPRC locus may have experienced selection pressure but found no significant signal in Tajima's D, Fay and Wu's H, F_{ST} or extended haplotype homozygosity, i.e. no departure from neutrality was observed.

Using the HapMap as a resource, we selected several SNPs for further study in sample sets that had previously only been genotyped for the key functional SNPs C77G or A138G. We aimed to choose SNPs that were well distributed across the gene, taking advantage of the pronounced LD to tag dozens of SNPs across the PTPRC locus and to include variants with a potential functional role

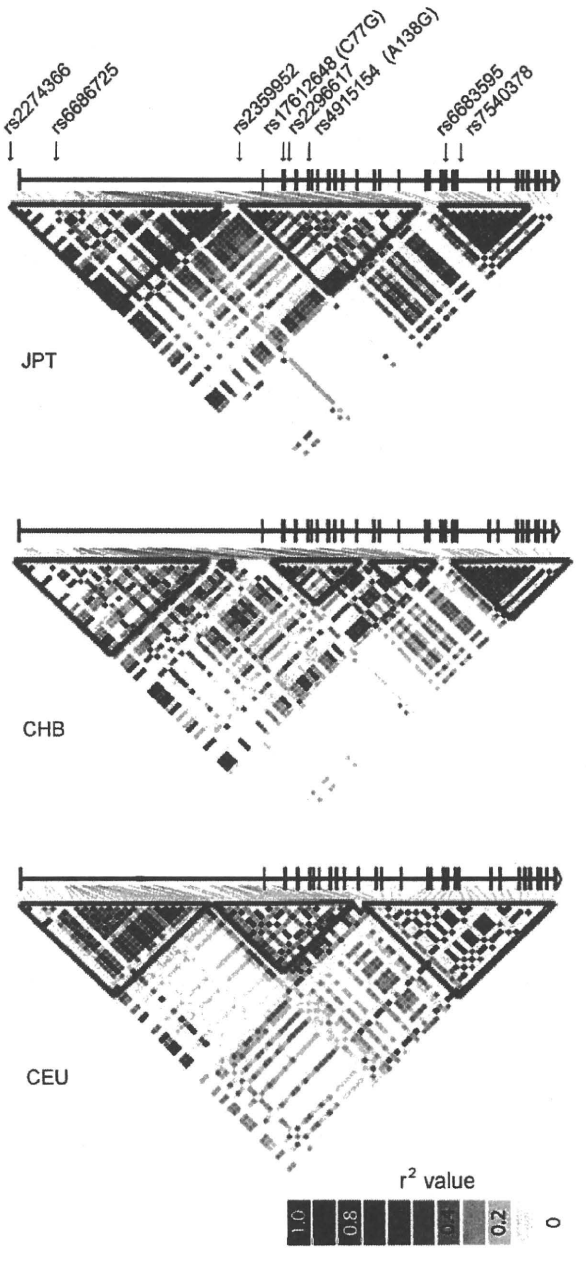

Figure 1 Linkage disequilibrium (LD) across the PTPRC gene. A model of the PTPRC gene with the location of our genotyped markers and LD (as r^2 ranging from 0 to 1 by greyscale generated using the HAPLOVIEW programme) derived from HapMap genotyping data (release 21 July 2006) from 45 unrelated Han Chinese from Beijing (CHB), 45 unrelated Japanese from Tokyo (JPT) and 30 US family trios of Northern and Western European origin collected by the Centre d'Etude du Polymorphisme Humain (CEU) (11). The PTPRC gene is located on chromosome 1 from coordinates 196874848 to 196992953 (NCBI build 36). Haplotype blocks are bound by red lines. The original genotyping data are available from the HapMap website (www.hapmap.org).

(e.g. promoter region, exonic SNPs or SNPs near splicing boundaries). SNP tagging is particularly effective in regions of strong LD, where statistical correlations between markers mean a single marker can be genotyped as a surrogate for several nearby markers, although the small sample size for the HapMap populations does pose limitations to this approach (13, 19). A total of eight markers was chosen, all had a minor allele frequency (MAF) greater than 5% in the JPT, CHB and CEU HapMap populations (apart for the C77G SNP at 1.7% in the CEU population). Across the 118.1 kb PTPRC locus, our genotyped markers tagged 36 common SNPs in the JPT population with an r^2 of greater than 0.8 (average r^2 of 0.96) and 25 SNPs in the CHB population (average r^2 of 0.99). Using the same criteria, we tagged 28 common SNPs in the CEU population with an r^2 of greater than 0.8 (average r^2 of 0.99). Details of the chosen SNPs with allele frequencies in our study populations are shown in Table 1.

We genotyped 355 JPT population controls, 204 individuals with type 1 diabetes and 174 patients with Graves' disease for our PTPRC SNP set (Table 2). The majority of these samples have previously been genotyped for A138G but not for other SNPs in PTPRC (3, 4). We performed single SNP and multiple-marker analysis. A total of 59 haplotypes was present, including 6 common haplotypes (>5% frequency) that had a combined frequency of 88.5% in our JPT controls. The 138G allele is found predominantly on the second most common haplotype (1121222), present at 19.7% in controls. The A138G SNP was associated with protection against Graves' disease (dominant model GG + GA *vs* AA: $P = 0.007$, $\chi^2 = 7.23$, OR = 0.57, 95% CI = 0.37–0.88). Comparing diabetes patients with controls showed a trend towards protection with 138G (dominant model: $P = 0.029$, $\chi^2 = 4.74$, OR = 0.67, 95% CI = 0.48–0.98). In contrast, the intron 20 (rs6683595) SNP was shown to be associated with disease in this patient group (CC *vs* CT + TT: $P = 0.01$, $\chi^2 = 6.64$, OR = 1.59, 95% CI = 1.10–2.29). Because this intron 20 polymorphism is close to the intron 20/exon 21 boundary, we considered whether this marker is correlated with altered splicing events at the 3′ end of the gene. We sequenced exons 19–25 of complementary DNA from peripheral blood mononuclear cells of nine JPT individuals with different A138G genotypes. No change in splicing was observed. The intron 20 rs6683595 SNP is located ~2.5 kb downstream from a homologous position in humans, where an engineered mutation in mice has been shown to have functional consequences on CD45 dimerisation and signalling (20). However, this mutation has never been found naturally. The intronic SNP rs6683595 shows some degree of LD with A138G (r^2 ~0.55, corresponding to an r of 0.74), reflected by the fact that the majority of both of their minor alleles are present in the 1121222 haplotype. The fact that some recombination occurs between these two polymorphisms could explain the seeming paradox of opposite single SNP associations with 138G being protective for diabetes, while the rs6683595 minor allele increases risk. This results from a reduced frequency of their shared haplotype in diabetes (from 19.7% in controls to 15% in cases) with enrichment in the normally rare species of haplotypes carrying the rs6683595 minor allele but not 138G (such that rs6683595 reaches 28.5% in cases). Thus, there is the possibility that this is a spurious finding (because of genotyping error, population stratification or chance) or that this is because of recombination occurring between the two SNPs. Either way, we consider this result an issue for further investigation. The common 1121222 haplotype bearing 138G was associated with protection from diabetes ($P = 0.046$, $\chi^2 = 3.99$, OR = 0.72,

Table 1 Details of genotyped single nucleotide polymorphisms (SNPs)[a]

RefSNP ID	Location	Residue	JPT (HapMap)	CHB (HapMap)	CEU (HapMap)	Japanese (controls)	Chinese (all)	UK (controls)
rs2274366	5UTR	–	10.2	15.6	0.0	15.1	12.0	0.0
rs6686725	int-2	–	29.1	19.3	40.0	24.5	24.8	34.3
rs2359952	int-2	–	37.5	37.8	22.0	38.6	27.7	31.0
rs17612648	ex-4 (C77G)	P57P	0.0	0.0	1.7	0.0	0.0	2.2
rs2296617	int-4	–	14.8	5.6	0.0	8.0	11.5	0.0
rs4915154	ex-6 (A138G)	T191A	15.6	21.1	0.0	21.7	20.6	0.0
rs6683595	int-20	–	17.8	28.9	10.8	22.7	25.2	12.6
rs7540378	ex-23	G770G	17.8	27.8	10.0	26.9	29.1	10.4

[a] The table shows the SNP identification number, location in the PTPRC gene, amino acid residue affected and minor allele frequencies (MAF as percentage) in the HapMap Japanese (JPT), Han Chinese (CHB) and US Caucasian (CEU) samples (11) and our own Japanese controls (4), whole Chinese sample cohort (21) and UK blood donors (22). All polymorphisms were genotyped by Sequenom hME Mass-Array primer extension assay (http://www.sequenom.de/) (24) under standard conditions. Primer sequences for genotyping and sequencing are available upon request. The average genotype success rate was 88.9% (ranging from 69.9 to 98.9%) across all SNPs and all populations. The distribution of genotypes was tested in controls and did not deviate from Hardy–Weinberg equilibrium ($P > 0.05$), except for rs6686725 in our Chinese population ($P = 0.018$). Differences between HapMap and our own results most probably reflect the smaller sample size of the HapMap datasets causing statistical fluctuation in sample allele frequency estimates and/or a degree of genetic heterogeneity between our populations and the relevant HapMap reference populations.

Table 2 Genetic association data for single nucleotide polymorphisms (SNPs) and haplotypes across PTPRC. A total of 355 Japanese population controls, 204 individuals with type 1 diabetes (DM) and 174 patients with Graves' disease[a]

Japanese population									
				Dominant model				Dominant model	
RefSNP ID	Control (n = 355)	DM (n = 204)	3 × 2 genotype model Pvalues	OR (95% CI)	Pvalues	Graves (n = 174)	3 × 2 genotype model Pvalues	OR (95% CI)	Pvalues
rs2274366	15.1	12.6	0.484			14.9	0.979		
rs6686725	24.5	23.0	0.781			21.9	0.099		
rs2359952	38.6	37.7	0.679			31.4	0.078		
rs2296617	8.0	9.0	0.574			9.3	0.449		
rs4915154 (A138G)	21.7	17.0	0.090	**0.67** (0.48–0.98)[b]	**0.029**	14.6	**0.025**	**0.57** (0.37–0.88)[c]	**0.007**
rs6683595	22.6	28.5	**0.029**	**1.59** (1.10–2.29)[d]	**0.01**	21.0	0.789		
rs7540378	26.9	24.6	0.776			22.6	0.057		

Haplotypes[e]	Control	DM	OR (95% CI)	Pvalues	Graves	OR (95% CI)	Pvalues
1111111	31.7	30.9			41.7		
1121222	19.7	15.0	**0.72** (0.51–1.01)[f]	**0.046**	14.1	**0.67** (0.46–0.96)[g]	**0.024**
1221111	13.1	13.5			11.2		
2111111	10.0	9.6			10.6		
1112111	7.2	7.8			5.7		
1211111	6.8	9.3			5.7		
All others	11.5	14.0			10.9		

[a] Genotype and allele frequencies in different patient groups were analysed using standard 3 × 2 or 2 × 2 χ^2 tests or Fisher exact tests (F-exact) applying SPSS 12.0 for Windows or the Statcalc programme (EPI2000; http://www.cdc.gov/epiinfo/index.htm). Odds ratios (OR) were used to describe the size of effect by allele/haplotype. No correction for multiple testing is made in the reported Pvalues. Haplotypes were estimated using the SNPHAP software programme (version 1.3; http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt) and inferred individual haplotypes imported into SPSS for statistical analysis (with 1 representing the major and 2 the minor allele for each SNP). Common haplotypes (>5%) are shown individually, the 'all others' category contains all haplotypes present at lower frequencies combined. Significant associations with SNPs and haplotypes are shown in bold.

[b] AA *vs* AG + GG: G carriers controls = 140 (39.8%) *vs* DM = 62 (30.5%).

[c] AA *vs* AG + GG: G carriers controls = 140 (39.8%) *vs* Graves' disease = 43 (27.4%).

[d] CC *vs* CT + TT: T carriers controls = 133 (39.1%) *vs* DM = 101 (50.1%).

[e] Order of SNPs in haplotypes: rs2274366, rs6686725, rs2359952, rs2296617, rs4915154, rs6683595, rs7540378 (with 1111111 = ctaaacc and 2222222 = tggggtt).

[f] 1121222 carriers: controls = 140 (19.7%) *vs* DM = 49 (14.1%).

[g] 1121222 carriers: controls = 140 (19.7%) *vs* Graves' disease = 61 (15.0%).

95% CI = 0.51–1.01) and Graves' disease (P = 0.024, χ^2 = 5.06, OR = 0.67, 95% CI = 0.46–0.96), but this does not add to the association with 138G on its own.

We genotyped a Chinese population of 604 children, aged 11–15 years, from the Xing-Chang Province for whom tuberculin response and *Ascaris lumbricoides* stool egg count had been measured (Table 1) (21). This population had not previously been screened for A138G. As in the JPT population, we found six common haplotypes (out of a total of 81) with the 138G containing haplotype being the second most common and present at 19.2% (Table 3). The distribution of common haplotypes in the JPT and Chinese populations was very similar, indicating that the A138G SNP is a good ancestry informative marker for the East Asian *vs* European populations but not for the Chinese *vs* JPT populations, i.e. that A138G has arisen prior to the separation of these two latter groups. No evidence of association was found between any individual SNP or haplotype with tuberculin response or egg count.

We also genotyped a group of 379 Caucasoid samples from the UK, comprising 192 hepatitis C virus (HCV) patients from Oxfordshire and 187 blood donor controls (Table 1) (22). HCV patients (including resolvers and persistently infected individuals) were a subset of the samples previously screened for C77G only, where we had seen an indication of association with susceptibility to infection and disease progression, i.e. severity of fibrosis (23). Four common haplotypes (frequency >5.0%, out of a total of 21 haplotypes present) were found across the gene with a combined frequency of 90.4% (Table 4). The 77G allele is present on a rare haplotype (21211) at a frequency of 1.7% (data not shown). No single SNP was correlated with susceptibility to HCV, and because of the low frequency of the C77G SNP, we were unable to analyse

Table 3 Genetic association data for haplotypes across PTPRC in a total of 604 Chinese individuals, 306 presented with a tuberculin response of <5 mm and 298 with >5 mm response; 492 individuals had worm egg counts [eggs per gram (epg) of stool] of <96 and 112 worm egg counts of >96[a]

Chinese population				
Haplotypes[b]	Tuberculin response <5mm (n = 306)	Tuberculin response >5mm (n = 298)	*Ascaris* epg <96 (n = 112)	*Ascaris* epg >96 (n = 492)
1111111	31.0	26.5	28.8	29.0
1121222	18.1	20.3	19.0	20.1
1211111	9.6	10.2	9.6	11.6
1221111	9.8	9.9	9.8	10.3
1112111	7.5	8.6	7.8	8.9
2111111	7.5	8.1	8.2	6.7
All others	16.3	16.3	17.0	13.4

[a] Analysis as detailed for Table 2. There was no association of any haplotype with either of these outcome measures.
[b] Order of SNPs in haplotypes: as given in Table 2.

the association of the 77G containing haplotypes with disease outcomes.

In summary, previous disease association studies of PTPRC have focused on only two variants, A138G and C77G. Employing a set of tagSNP markers covering the whole length of the gene, we have investigated whether other variations in the gene were associated with diseases such as type 1 diabetes, Graves' disease in the JPT population, tuberculin response and worm burden in the Chinese population and HCV in CEU population. No stronger or novel signals of association than previously reported for Graves' disease and diabetes (3, 4) were discovered, apart from one intronic SNP (rs6683595) that showed some LD with the exonic SNP A138G. We also showed that most of the 138G allele is present in one common haplotype. Given the existing evidence of altered splicing and disease associations with A138G, which is observed at frequency of around 20% in East Asian populations, this polymorphism continues to be the leading candidate as a functional variant in the PTPRC gene.

Table 4 Genetic association data for haplotypes across PTPRC. A total of 379 samples from the UK, comprising 192 HCV patients and 187 blood donor controls[a]

European population		
Haplotypes[b]	Control (n = 192)	HCV (n = 187)
11111	28.8	34.0
21111	37.0	33.3
22111	17.2	15.6
22122	7.4	8.7
All others	9.6	8.4

[a] Analysis as detailed for Table 2. No association with susceptibility to HCV was observed.
[b] Order of SNPs in haplotypes: rs6686725, rs2359952, rs17612648, rs6683595, rs7540378 (with 11111 = taccc and 22222 = gggtt).

References

1. Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol* 2003: **21**: 107–37.
2. Tchilian EZ, Beverley PC. Altered CD45 expression and disease. *Trends Immunol* 2006: **27**: 146–53.
3. Boxall S, Stanton T, Hirai K et al. Disease associations and altered immune function in CD45 138G variant carriers. *Hum Mol Genet* 2004: **13**: 2377–84.
4. Ward V, Hennig BJ, Hirai K et al. Geographical distribution and disease associations of the CD45 exon 6 138G variant. *Immunogenetics* 2006: **58**: 235–9.
5. Tchilian EZ, Wallace DL, Dawes R et al. A point mutation in CD45 may be associated with HIV-1 infection. *AIDS* 2001: **15**: 1892–4.
6. Ballerini C, Rosati E, Salvetti M et al. Protein tyrosine phosphatase receptor-type C exon 4 gene mutation distribution in an Italian multiple sclerosis population. *Neurosci Lett* 2002: **328**: 325–7.
7. Barcellos LF, Caillier S, Dragone L et al. PTPRC (CD45) is not associated with the development of multiple sclerosis in U.S. patients. *Nat Genet* 2001: **29**: 23–4.
8. Cocco E, Murru MR, Melis C et al. PTPRC (CD45) C77G mutation does not contribute to multiple sclerosis susceptibility in Sardinian patients. *J Neurol* 2004: **251**: 1085–8.
9. Vogel A, Strassburg CP, Manns MP. 77 C/G mutation in the tyrosine phosphatase CD45 gene and autoimmune hepatitis: evidence for a genetic link. *Genes Immun* 2003: **4**: 79–81.
10. Schwinzer R, Witte T, Hundrieser J et al. Enhanced frequency of a PTPRC (CD45) exon A mutation (77C–>G) in systemic sclerosis. *Genes Immun* 2003: **4**: 168–9.
11. Consortium IH. A haplotype map of the human genome. *Nature* 2005: **437**: 1299–320.
12. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005: **21**: 263–5.
13. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005: **37**: 1217–23 [Epub 23 Oct 2005].
14. Timon M, Beverley PC. Structural and functional analysis of the human CD45 gene (PTPRC) upstream region: evidence for a functional promoter within the first intron of the gene. *Immunology* 2001: **102**: 180–9.
15. van Vliet SJ, Gringhuis SI, Geijtenbeek TB, van Kooyk Y. Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. *Nat Immunol* 2006: **7**: 1200–8.
16. Filip LC, Mundy NI. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. *Mol Biol Evol* 2004: **21**: 1504–11.

17. Ballingall KT, Waibochi L, Holmes EC et al. The CD45 locus in cattle: allelic polymorphism and evidence for exceptional positive natural selection. *Immunogenetics* 2001: **52**: 276–83.
18. Mustelin T, Vang T, Bottini N. Protein tyrosine phosphatases and the immune response. *Nat Rev Immunol* 2005: **5**: 43–57.
19. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004: **74**: 106–20.
20. Majeti R, Xu Z, Parslow TG et al. An inactivating point mutation in the inhibitory wedge of CD45 causes lymphoproliferation and autoimmunity. *Cell* 2000: **103**: 1059–70.
21. Peisong G, Mao XQ, Enomoto T et al. An asthma-associated genetic variant of STAT6 predicts low burden of ascaris worm infestation. *Genes Immun* 2004: **5**: 58–62.
22. Hellier S, Frodsham AJ, Hennig BJ et al. Association of genetic variants of the chemokine receptor CCR5 and its ligands, RANTES and MCP-2, with outcome of HCV infection. *Hepatology* 2003: **38**: 1468–76.
23. Dawes R, Hennig B, Irving W et al. Altered CD45 expression in C77G carriers influences immune function and outcome of hepatitis C infection. *J Med Genet* 2006: **43**: 678–84.
24. Jurinke C, van den Boom D, Cantor CR, Koster H. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* 2002: **187**: 179–92.

