

Genetic structural differences between responders and non-responders to interferon therapy for chronic hepatitis-B patients

Pei-Jer Chen · Cherry Guan-Ju Lin ·
Felicia Yi-Fang Lin · Ellson Chen ·
Lawrence Shih-Hsin Wu

Received: 27 June 2006 / Accepted: 5 August 2006 / Published online: 22 September 2006
© The Japan Society of Human Genetics and Springer 2006

Abstract Interferon- α therapy has become a main stay of treatment for hepatitis-B patients. The sustained remission rates are around 30%, and the factors determining response are poorly defined. Our study aimed to search for the genetic differences between responder and non-responder patients. We have found 13 short tandem repeat markers (STR) that display different allele and/or genotype frequency between the two patient groups. Eleven out of 13 STR markers were selected to perform principal component analysis and hierarchical clustering. The study subjects could be further divided into six groups based on their genetic similarity, which correlated with the drug response rate. In conclusion, this pilot study has developed a new approach to identify genetic markers that allows us to predict the drug response in hepatitis B patients. Our study utilizing STR markers may provide an alternative approach to the utilized SNP markers in pharmacogenetic study.

Keywords Sustained response of interferon therapy · Short tandem repeat markers · Principal component analysis · Hierarchical clustering

Introduction

Although vaccines have been available for almost 2 decades, chronic hepatitis-B virus (HBV) infection remains a major health problem worldwide. Current chronic hepatitis B is treated with either nucleotide analog or interferon. Interferon therapy successfully controls infection in about one-third of the chronically infected individuals (Delaney et al. 2001). Treatment with interferon-alpha often leads to the cessation of viral replication in 30–40% of the patients with chronic hepatitis B (Heintges et al. 2001). In addition, the 6–12 months of therapy is not only expensive, but also includes many side effects that can be debilitating in some patients (Liaw 2002; Wai and Lok 2002; Feld and Locarnini 2002). These tailorings of the therapies to individual patient problems have led to the search for predictors of the drug response rate to treatment. Current candidate predicting factors include viral genotypes, the ALT level, serum HBV DNA, female gender, fibrosis on liver biopsy and the serum fibronectin level (Kao 2002; Sakai et al. 2002; Kao et al. 2000a, b; Neudorf-Grauss et al. 2000; Helvacı et al. 1999).

With the advent of pharmacogenetics, and considering the host's unique genetic background, the role of DNA polymorphisms (including STRP and SNP) in association with treatment response has become increasingly appreciated and supported in a variety of illnesses. Hence, looking into such a topic may lead to important predictions of treatment response for interferon therapy, with its many displeasing side effects, in HBV patients. In particular in hepatitis-B disease, MHC I and MHC II class polymorphisms, TNF- α , mannose-binding protein, as well as eIF-2 α and MxA in

P.-J. Chen
Hepatitis Research Center,
National Taiwan University Hospital,
Taipei, Taiwan

C. G.-J. Lin · F. Y.-F. Lin · E. Chen · L. S.-H. Wu (✉)
Research Development Division,
Vita Genomics Inc., 7Fl., No.6, Sec.1,
Jungshing Rd., Wugu Shiang,
Taipei County 248, Taiwan
e-mail: lawrence.wu@vitagenomics.com

the interferon pathway, the SNPs have been suggested to affect the host immune and antiviral response and thus are associated with variable disease progression and treatment responses for previous studies (Hohler et al. 1997; Thursz et al. 1995; McNicholl et al. 2000; Scully et al. 1990; King et al. 2002).

Genetically structured populations may be composed of two or more subpopulations with distinct drug-reaction profiles and thus may be better considered separately in some contexts. Sometimes the population substructure is not obvious, and as a result, a sample may consist of heterogeneous subsamples from the population. This raises the question of the appropriate way to find the genetic background difference between distinct drug-reaction groups and of how to relate the inferred genetic structure to drug response.

Our study focuses on researching the differences of host genetic backgrounds between interferon responders and non-responders by highly polymorphic STR markers. We found 13 STR markers with different allele frequencies between the responder and non-responder groups. Principal component (PC) analysis and hierarchical clustering analysis were performed to find the genetically different patient subgroups with differing responsiveness by STR markers. It could provide a method for analyzing host genetic factors of distinct drug-reaction groups and a different view of the candidate gene SNP approach.

Materials and methods

Study subjects

We retrospectively enrolled 104 Chinese Han patients with chronic hepatitis B from our outpatient clinics at the National Taiwan University Hospital and Taipei Municipal Jen-Ai Hospital. Informed consent was signed and collected from each patient. All patients' blood samples were HbsAg (+) and HBeAg (+) with an elevated ALT that was at least twofold higher than the upper limits of normal for 6 months. HBV genotypes were determined using PCR-restriction fragment-length polymorphisms of the surface gene of HBV as in previous reports (Kao et al. 2000a, b; Lindh et al. 1999). Patients were excluded from receiving interferon therapy if they had any of the following criteria: neutrophil count $<1,500$ cells/mm³, Hgb <12 g/dl in women or 13 g/dl in men, or a platelet count of $<90,000$ cells/mm³, a history of poorly controlled thyroid disease, and a serum creatinine level >1.5 times the upper limit of normal at the initial screening. Live biopsy was performed in some patients

to exclude cases with severe cirrhosis. Eligible patients received interferon-alpha (2a or 2b) at a dosage of 5–10 MU three times per week for 4–6 months, and were subsequently followed by a series of standard tests to assess the treatment response for more than 1 year. The definition of sustained responders to IFN treatment for chronic hepatitis B disease included the loss of HBeAg at 1 year after completing the treatment. Patients with concurrent hepatitis C or D infection were also excluded from the study. Our study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected by the approval of our (National Taiwan University Hospital) institutional review committee.

DNA preparation

Genomic DNA was extracted from 99 unrelated Taiwanese chronic hepatitis patients. DNA was isolated from blood samples using QIAamp DNA Blood kit (QIAGEN, Valencia, CA) according to the manufacturer's instructions. The quality of the isolated genomic DNA was checked using agarose gel electrophoresis analysis, and the quantity was determined by spectrophotometer and stored at -80°C until use.

Microsatellite genotyping

Genotyping was performed using the ABI PRISM Linkage Mapping Sets MD-10 (400 markers; Applied Biosystems: ABI) and provided coverage of the human genome at 10-cM average resolution. Each marker set included a fluorescence-labeled forward primer and a tailing reverse primer. The PCR amplifications were carried out according to the manufacturer's instructions. PCR products were separated on ABI 3700 DNA analyzers. The use of the GeneScan 500 LIZ as the internal size standard assists in polymorphic fragment length calling and allows more accurate allele calling and unambiguous comparison of data across experimental conditions. Genotypes were initially scored using Genescan and Genotyper (ABI) software and were then verified independently by three individuals without prior knowledge of the phenotype.

Statistical analysis

The associations of STR markers with the interferon treatment response were analyzed by Monte Carlo simulation (Sham and Curtis 1995). Various alleles of the significant markers were tested for linkage disequilibrium by the analysis of the contingency table. The significant alleles were tested for risk factors by

odds ratio. The genotype contingency tables are constructed according to the specific allele with the significant P value and odds ratio. The meaningful genotypes are generated by testing with χ^2 and significant odds ratio to the genotype contingency table. All significant marker genotype information was collected and the dataset transformed into a binary category; the meaningful genotype and the other genotype were represented with “1” and “0,” respectively. PC analysis and hierarchical clustering were performed on the transformed binary dataset. Major parts of the statistical analyses were performed by SAS 8.0.

Results

Patient characteristics

Sera were collected from 99 chronic hepatitis patients. Forty-six patients were responsive and 53 patients non-responsive to IFN α treatment. Demographic and clinical characteristics of IFN α -treated chronic hepatitis B patients are shown in Table 1. The sums of patients were not consistent in some comparing of the clinical outcome due to several patients with unavailable characteristics.

Markers correlated with the efficacy of interferon therapy

Two hundred twenty-one STR markers in chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 17, 21 and 22 were genotyped. The other chromosomes were not genotyped due to the subjects' DNA shortages. After the data were analyzed, 13 STR markers were sought by allele and/or genotype frequency with a significant difference between the responder and non-responder groups (Table 2). For example, the D1S2890 226bp allele, defined as the E allele, was more frequent in the non-responder group and showed a significant odds ratio. Homozygosity of the D1S2890 E allele and

heterozygosity of the D1S413 A allele showed significantly different frequencies between the responder and non-responder groups. The alleles of each locus were tested with the odds ratios and $2 \times 2 \chi^2$ test (data not shown), otherwise with alleles of the same locus. In Table 2, the column “Associated allele test” lists the alleles that were statistically significant in the test. Some loci, for example, D1S2890, have more than one associated allele. The genotype for each associated allele was classified as homozygous and heterozygous then tested with the odds ratio and $2 \times 2 \chi^2$ test (data not shown), otherwise with genotypes of the same locus. In Table 2, the column “Meaningful genotype test” lists the results of the test. The genotype of locus D1S2890 was shown to be a recessive model. Other loci, except D6S292, were shown to be dominant models. The marker D6S292 could not define a specific allele or genotype with different frequencies between the two patient groups and therefore was discarded from further analysis. Along with D6S292, D2S367 was also discarded because the Monte Carlo estimation did not reach a significant level. D17S785 was another marker not reaching a significant level with the Monte Carlo estimation, but its C-allele and genotype contained with the C-allele showed good differentiation between the two patient groups. We selected 11 of 13 STR markers, discarding D2S367 and D6S292, for further analysis.

Principal component analysis

The significant genotype datasets were transformed into the binary category; the meaningful genotype and the others genotype were represented with “1” and “0,” respectively. In loci with more than one meaningful genotype, a higher frequency than one was selected.

The principal component analysis is a technique that analyzes the information of uncorrelated composite variables without any significant loss from the original data sets. The scores resulting from the PC can also be used as input variables for further analysis of the data using multivariate techniques such as cluster analysis, regression analysis and discriminate analysis. The advantages of using PC scores is that the new variables are not correlated, and the problem of multicollinearity is avoided (Subhash 1996). With these characteristics, we use this method to analyze STR data that may exist multicollinearly.

Some common criteria were used to define the number of PCs. First, the eigenvalue of the components should be larger than one. Second, fewer components and larger culminate variance are better. SAS

Table 1 Demographic and clinical characteristics of patients with response or non-response to IFN treatment

	Responder	Non-responder	P value
Age, years	30.5±8.39	33.9±8.08	0.158
Gender, male/female	34/10	43/5	0.094
HBV genotype B/C	25/6	18/22	0.003
Pretreatment ALT level, U/l	214.39±156.45	196.71±169.107	0.632
Pretreatment HBV DNA, log _e copies/ml	18.69±2.17	18.98±1.90	0.581

Table 2 Testing results of the allele and/or genotype frequency of 13 STR markers that were significantly different between the responder and non-responder groups

Locus	R/NR ^a	Allele no.	P value ^b	Associated allele test		Meaningful genotype test	
				Allele-bp (R/NR)	OR (95% CI)	Genotype (R/NR)	OR (95% CI)
D1S2890	46/53	8	0.0279	E-226 (28/56)	0.39 (0.22–0.70)	EE (3/16)	0.39 (0.16–0.91)
				G-230 (25/15)	2.26 (1.1–4.62)	GG + Gx (22/14)	2.55 (1.10–5.92)
D1S413	46/53	10	0.003	A-252 (23/10)	3.2 (1.43–7.15)	AA + Ax (20/10)	3.31 (1.34–8.15)
				C-256 (6/18)	0.34 (0.13–0.90)	CC + Cx (6/16)	0.35 (0.12–0.98)
D2S367	43/52	12	0.147	E-321 (9/22)	0.44 (0.19–1.01)	EE + Ex (9/21)	0.39 (0.16–0.98)
				H-327 (18/7)	3.67 (1.45–9.26)	HH + Hx (16/7)	3.33 (1.20–9.29)
D2S319	43/52	8	0.0471	F-136 (21/12)	2.48 (1.14–5.39)	FF + Fx (20/12)	2.90 (1.20–6.99)
D3S1289	43/51	8	0.0116	F-213 (12/3)	5.35 (1.46–19.65)	FF + Fx (12/3)	6.19 (1.62–23.73)
D4S391	46/53	10	0.0424	D-162 (26/16)	2.22 (1.10–4.46)	DD + Dx (22/14)	2.55 (1.10–5.92)
D5S406	46/53	11	0.0043	H-182 (24/45)	0.49 (0.17–0.90)	HH + Hx (19/34)	0.41 (0.18–0.93)
D6S1581	46/53	8	0.0241	B-261 (31/19)	2.33 (1.20–4.49)	BB + Bx (27/17)	3.00 (1.32–6.85)
D6S292	46/52	11	0.0067	Not found	Not found	Not found	Not found
D7S515	43/51	14	0.0035	H-173 (29/16)	2.73 (1.36–5.49)	HH + Hx (26/15)	3.67 (1.56–8.66)
D8S260	43/51	10	0.043	F-208 (16/34)	0.46 (0.23–0.90)	FF + Fx (13/29)	0.33 (0.14–0.77)
D9S288	46/53	13	0.00019	D-137 (19/6)	4.34 (1.65–11.40)	DD + Dx (17/6)	4.59 (1.62–12.98)
				K-151 (1/15)	0.07 (0.01–0.52)	KK + Kx (1/14)	0.06 (0.01–0.49)
D17S785	46/54	13	0.1322	C-173 (12/3)	5.25 (1.43–19.23)	CC + Cx (12/3)	6.00 (1.57–22.86)

OR odds ratio; CI confidence interval

^a The numbers in the R/NR column are represented responder/non-responder, which were genotyped by each locus

^b Monte Carlo estimates of exact P values

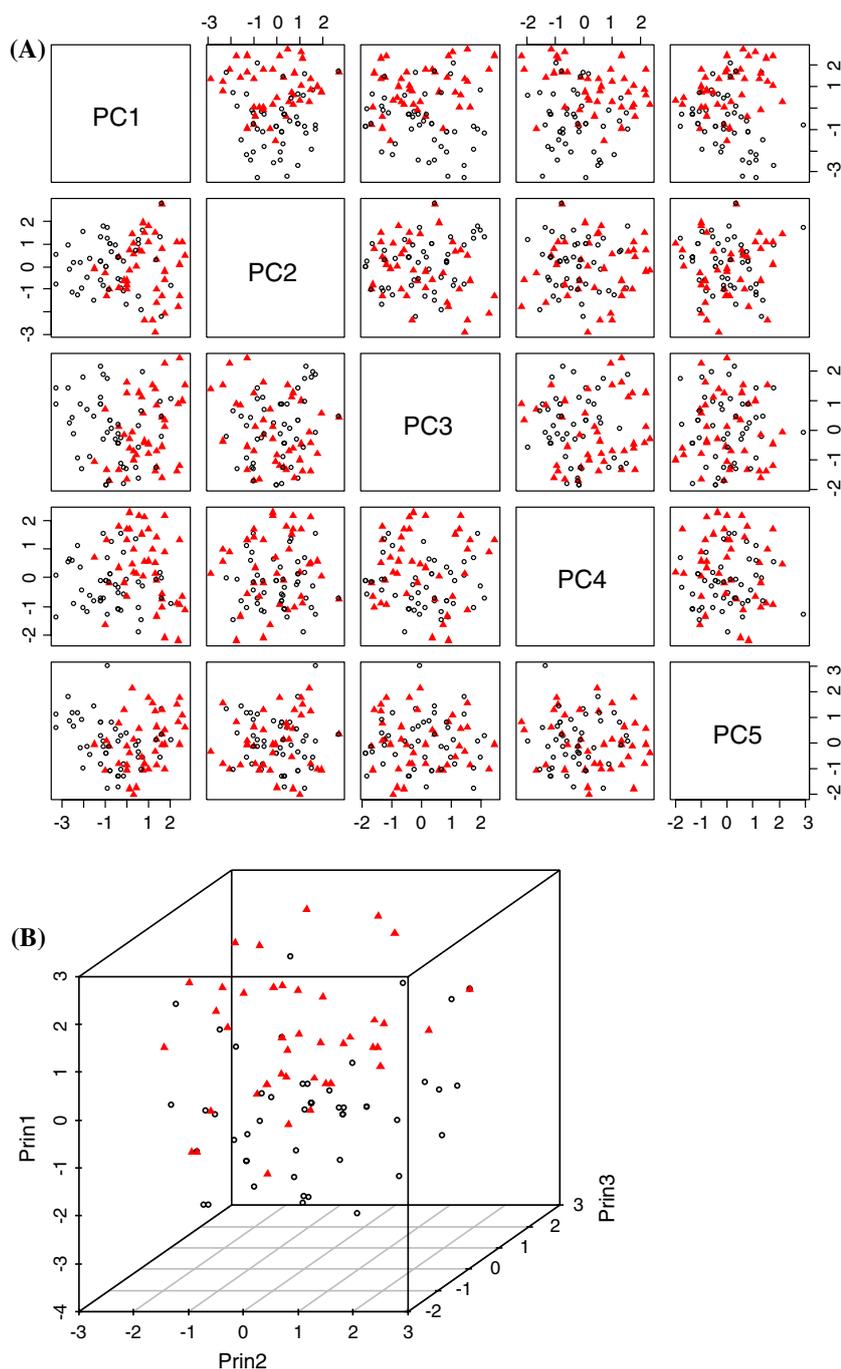
and R (the R Project for Statistical Computing, <http://www.r-project.org>) were used to perform PC analysis of significant STR markers. In Table 3, we chose five components that explain 62% of the total variability amount of these markers. Table 3 shows the weights of all markers for the first five components. The first component always has the largest proportion. In our case, this was 18%. D8S260 and D9S288 provide more information than the others in component 1, and so does D5S406 in component 2, D7S515 in component 3 and D4S391 in component 4. The positive and negative weights indicate the different directions of their information.

Figure 1a shows the pair-wise scatter plot of the five components. The non-responders are shown in the red triangle, and the black circle represents the responders. There are some patterns between component 1 and the others. These scatter plots correspond with the two groups in drug response, although several subjects are misgrouped. Other components do not have the same pattern as component 1, but they still provide some information. Figure 1b shows the scatter plot of the first three components. The two groups represented by the red triangle and black circle can be divided into obvious different parts. The ambiguous region is at the cross part between the two distinguishing parts in this

Table 3 Principal component score of the first five components

	Prin1	Prin2	Prin3	Prin4	Prin5
D1S2890	-0.3123	0.1325	0.3977	-0.1962	-0.0105
D1S413	0.3714	-0.3129	0.1776	0.0541	0.2190
D2S319	0.1901	0.3833	0.3909	0.3669	-0.3169
D3S1289	0.3355	-0.4646	0.3346	-0.1218	-0.0006
D4S391	-0.0787	-0.1453	0.1820	0.7447	0.1115
D5S406	-0.0605	0.5021	-0.0881	0.0256	-0.1421
D6S1581	0.3366	0.3753	0.0948	0.1992	0.4824
D7S515	0.1271	-0.0790	-0.6891	0.2651	0.0023
D8S260	-0.4096	-0.1131	-0.0284	0.2273	0.4509
D9S288	-0.4474	0.0028	0.1327	-0.0856	0.3697
D17S785	0.3296	0.2978	-0.0482	-0.2915	0.4961
Eigenvalue	1.9845	1.4066	1.2468	1.15459	1.0213
Cumulative proportion	0.1804	0.3083	0.4216	0.5266	0.6194

Fig. 1 Principal component scores. **a** Two-dimensional scatter plot of each two of the first five component scores. **b** Three-dimensional scatter plot of the first three component scores. *Open triangle*: non-responder; *open circle*: responder



plot. The PC scores were used to perform hierarchical clustering.

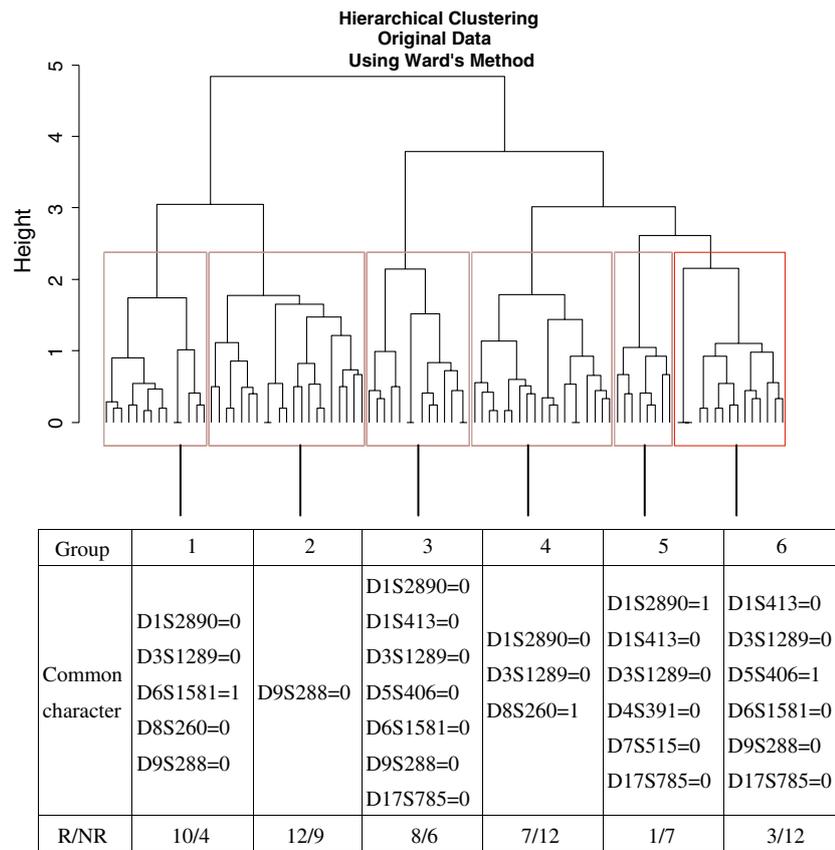
Hierarchical clustering analysis

Cluster analysis is a set of methods that can divide samples into several subgroups. These methods were often used to observe the nature of the features between/within the divided subgroups. If samplings are from several different groups, the members of the same

group usually share some common attributes (or character). The attributes of subjects are similar in the same group and different between groups. Most of the cluster analysis methods divide subjects by their character similarities. It often considers the distances between two subjects as a measurement of similarity.

Obviously, some subdivisions were found in our data, which were performed by several distance matrixes and Ward's cluster methods using SAS software. In Figs. 2, 3 these subjects could be separated into six

Fig. 2 Hierarchical clustering using original binary dataset. Every group has their common character. R/NR is the count of responders and non-responders belonging to the same group



groups unless the data were transformed by the PC score. Although the number of groups is the same, the subjects in each group are not exactly the same. The PC score is uncorrelated. It may imply some potential relations existed among markers when clustering subjects by these markers.

For the purpose of finding homogeneous responder groups, clustering using the PC score should be better at grouping than using the original binary data. The R/NR ratio of Fig. 2 shows that group 1 has a higher R/NR ratio (10/4). In Fig. 3, groups 3, 4 and 5 show higher R/NR ratios: 11/4, 7/2 and 11/2, respectively.

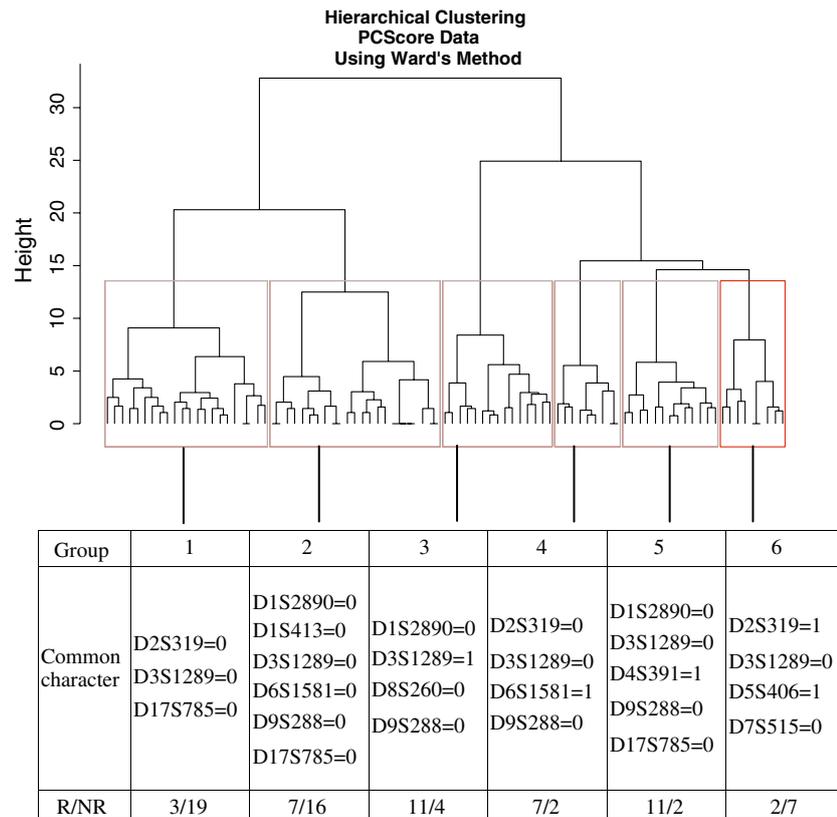
Discussion

Variation in allelic frequency among subpopulations, or clinal variation, is often thought to result from differential selection in various parts of the population in combination with population substructuring. Such spatial variation in allelic frequency may be temporally stable, but changing with time. Limited gene flow between the subpopulation gives a stable polymorphism that gradually disappears as the amount of gene flow increases (Hedrick 2000).

The patients’ inclusion criteria may act as a selection power for picking a specific genotype pool. All patients have the same phenotype (HBV carrier) except for the efficacy of interferon treatment. Variation in allelic frequency between responders and non-responders may present clinal variation and/or linkage disequilibrium with polymorphism of the causative genes. Theoretically, STR markers have more power to detect LD than diallelic markers such as SNP (Ott and Rabinowitz 1997; Chapman and Wijsman 1998). The sufficiently dense sets of microsatellite data can be used to make initial predictions about the level of short-range LD present in susceptibility regions identified by linkage study (Schulze et al. 2002).

Because of their underlying complexity, constructing predictive models of interferon-treated sustained response requires (1) more samples to perform guaranteed robust solutions for hierarchical clustering, (2) more genotyping to find more suitable markers, and (3) including environmental deviation or other non-genetic affection factors (i.e., gender, age, viral type, viral titer, etc.) for the analysis. Our pilot study suggests HBV infection patients could be divided into genetic subgroups with variable sustained responses. The discoveries yield a powerful knowledge base to develop tools

Fig. 3 Hierarchical clustering using PC score transformed dataset. Every group has its common character. R/NR is the count of the responders and non-responders belonging to the same group



for predicting and changing the course of HB patient care and treatment.

Acknowledgements We thank Dr. Ying-Jye Wu and Dr. Ling-Mei Wang for comments on the manuscript, Ms. Julia Yu for data management, and Mr. Lichih Huang for technical support.

References

- Chapman NH, Wijsman EM (1998) Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *Am J Hum Genet* 63:1872–1885
- Delaney WE IV, Locarnini S, Shaw T (2001) Resistance of hepatitis B virus to antiviral drugs: current aspects and directions for future investigation. *Antivir Chem Chemother* 12:1–35
- Feld J, Locarnini S (2002) Antiviral therapy for hepatitis B virus infections: new targets and technical challenges. *J Clin Virol* 25:267–283
- Hedrick PW (2000) Genetics of populations. Jones and Bartlett, MA
- Heintges T, Petry W, Kaldewey M, Erhardt A, Wend UC, Gerlich WH, Niederau C, Haussinger D (2001) Combination therapy of active HBsAg vaccination and interferon-alpha in interferon-alpha nonresponders with chronic hepatitis B. *Dig Dis Sci* 46:901–906
- Helvacı M, Ozkaya B, Ozbal E, Ozinel S, Yaprak I (1999) Efficacy of interferon therapy on serum fibronectin levels in children with chronic hepatitis B infection. *Pediatr Int* 41:270–273
- Hohler T, Gerken G, Notghi A, Lubjuhn R, Taheri H, Protzer U, Lohr HF et al (1997) HLA-DRB1*1301 and *1302 protect against chronic hepatitis B. *J Hepatol* 26:503–507
- Kao JH (2002) Hepatitis B viral genotypes: clinical relevance and molecular characteristics. *J Gastroenterol Hepatol* 17:643–650
- Kao JH, Chen PJ, Lai MY, Chen DS (2000a) Hepatitis B genotypes correlate with clinical outcomes in patients with chronic hepatitis B. *Gastroenterology* 118:554–559
- Kao JH, Wu NH, Chen PJ, Lai MY, Chen DS (2000b) Hepatitis B genotypes and the response to interferon therapy. *J Hepatol* 339:998–1002
- King JK, Yeh SH, Lin MW, Liu CJ, Lai MY, Kao JH, Chen DS, Chen PJ (2002) Genetic polymorphisms in interferon pathway and response to interferon treatment in hepatitis B patients: a pilot study. *Hepatology* 36:1416–1424
- Liaw YF (2002) Therapy of chronic hepatitis B: current challenges and opportunities. *J Viral Hepatol* 9:393–399
- Lindh M, Hannoun C, Dhillon AP, Norkrans G, Horal P (1999) Core promoter mutations and genotypes in relation to viral replication and liver damage in East Asian hepatitis B virus carriers. *J Infect Dis* 179:775–782
- McNicholl JM, Downer MV, Udhayakumar V, Alper CA, Swerdlow DL (2000) Host–pathogen interactions in emerging and re-emerging infectious diseases: a genomic perspective of tuberculosis, malaria, human immunodeficiency virus infection, hepatitis B, and cholera. *Annu Rev Public Health* 21:15–46
- Neudorf-Grauss R, Bujanover Y, Dinari G, Broide E, Neveh Y, Zahavi I, Reif S (2000) Chronic hepatitis B virus in children in Israel: clinical and epidemiological characteristics and response to interferon therapy. *Isr Med Assoc J* 2:164–168

- Ott J, Rabinowitz D (1997) The effect of marker heterozygosity on the power to detect linkage disequilibrium. *Genetics* 147:927–930
- Sakai T, Shiraki K, Inoue H, Okano H, Deguchi M, Sugimoto K, Ohmori S, Murata K, Nakano T (2002) Efficacy of long-term interferon therapy in chronic hepatitis B patients with HBV genotype C. *Int J Mol Med* 10:201–204
- Schulze TG, Chen YS, Akula N, Hennessy K, Badner JA, McInnis MG, DePaulo JR, Schumacher J, Cichon S, Propping P, Maier W, Rietschel M, Nothen MM, McMahon FJ (2002) Can long-range microsatellite data be used to predict short-range linkage disequilibrium? *Hum Mol Genet* 11:1363–1372
- Scully LJ, Brown D, Lloyd C, Shein R, Thomas HC (1990) Immunological studies before and during interferon therapy in chronic HBV infection: identification of factors predicting response. *Hepatology* 12:1111–1117
- Sham PC, Curtis D (1995) Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 59:97–105
- Subhash S (1996) *Applied multivariate techniques*. Wiley, New York
- Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, Hill AV (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *N Engl J Med* 332:1065–1069
- Wai CT, Lok AS (2002) Treatment of hepatitis B. *J Gastroenterol* 37:771–778