to the dramatic increase in HCV RNA replication efficiency obtained by introducing the adaptive mutation alone. A similar but more moderate phenotype was observed with cm 252 when the adaptive mutation was present. Our result is different from the reported enhancement of G418-resistant colony formation caused by a large 47-aa deletion in HCV-Con1 (4) and by a natural 4-aa insertion in HCV-N (29). Region 3 contains the IFN sensitivity-determining region, within which mutations have been reported to be associated with sensitivity to IFN therapy in patients with chronic HCV infection in Japan (14, 15). Region 3 is almost the same as a protein kinase R (PKR)-binding domain (18). It has been reported that NS5A can disrupt the dimerization of PKR through binding, resulting in the repression of PKR function, and efficient HCV RNA replication may involve a block in PKR-dependent signaling (18, 46). In this context, the adaptive mutations can greatly augment HCV RNA replication and thus may induce PKR, which could be inactivated via interaction with the PKR-binding region of NS5A, region 3, but not with those mutants defective in PKR-binding, such as del-3 and cm 252.

Second, cm 290 was less replication competent than the other cm mutants in region 2 when the adaptive mutation, S232I or S225P, was present. The highly charged sequence mutated in cm 290 may be critical for HCV RNA replication in the presence of the adaptive mutation. To address this point, four different combinations of three to four alanine substitutions in 7 aa residues were introduced into the M1LE/S232I construct. All of these mutants were more efficient in HCV RNA replication than M1LE/S232I plus cm 290 but still less so than the other replication-competent mutants in region 2 (data not shown), suggesting that all or most amino acids in the sequence contribute to its critical role in HCV RNA replication in the presence of the adaptive mutations. This result may suggest a functional linkage of the sequence mutated in cm 290 to the adaptive mutations.

In HCV subgenomic replicons, some groups have found that cured cell clones showed a high permissiveness for HCV RNA replication (5, 45) while another has not (40). Between the recipient sublines we used, the cured cells (Huh-7-KV-C) shared the nonpermissive property for wild-type M1LE; however, the DMB subline was permissive for wild-type M1LE and/or adaptive mutations with lower efficiency than the KV-C subline. These results suggest that several different genetic or phenotypic alterations in recipient cells emerge under IFN treatment or multiple ways for cells to be permissive to HCV RNA replication.

In summary, we established a highly efficient HCV replicon system derived from the isolate M1LE and demonstrated that the two regions critical for the interaction between NS5A and NS5B are also indispensable for HCV RNA replication in an HCV replicon system. Our results strongly suggest that NS5A is involved in the HCV replication complex and acts as a positive modulator of HCV RNA replication through its interaction with NS5B. The molecular mechanism of this positive effect by NS5A remains to be elucidated and may lead to the design of new drugs that inhibit HCV RNA replication.

ACKNOWLEDGMENTS

We are grateful to N. Hayashi and T. Nomura for encouraging discussion. We also thank M. Yasukawa and K. Kuwabara for technical assistance.

This work was supported in part by the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research, Grants-in-aid for scientific research (B) and development, and Grants-in-aid for scientific research on priority areas (C) in oncogenesis from the Ministry of Education, Sports, Culture, and Technology.

REFERENCES

1. **Alter, H. J., R. H. Purcell, J. W. Shih, J. C. Melpolder, M. Houghton, Q. L. Choo, and G. Kuo.** 1989. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. N. Engl. J. Med. **321:**1494–1500.
2. **Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen.** 1993. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J. Virol. **67:**3835–3844.
3. **Behrens, S. E., L. Tomei, and R. De Francesco.** 1996. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. EMBO J. **15:**12–22.
4. **Blight, K. J., A. A. Kolykhalov, and C. M. Rice.** 2000. Efficient initiation of HCV RNA replication in cell culture. Science **290:**1972–1974.
5. **Blight, K. J., J. A. McKeating, and C. M. Rice.** 2002. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. J. Virol. **76:**13001–13014.
6. **Blight, K. J., J. A. McKeating, J. Marcotrigiano, and C. M. Rice.** 2003. Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. J. Virol. **77:**3181–3190.
7. **Bordo, D., and P. Argos.** 1991. Suggestion for safe residue substitutions in site-directed mutagenesis. J. Mol. Biol. **217:**721–729.
8. **Brass, V., E. Bieck, R. Montserret, B. Wölk, J. A. Hellings, H. E. Blum, F. Penin, and D. Moradpour.** 2002. An amino-terminal amphipathic α-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. J. Biol. Chem. **277:**8130–8139.
9. **Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244:**359–362.
10. **Chung, K. M., J. Lee, J. E. Kim, O. K. Song, S. Cho, J. Lim, M. Seedorf, B. Hahm, and S. K. Jang.** 2000. Nonstructural protein 5A of hepatitis C virus inhibits the function of karyopherin β3. J. Virol. **74:**5233–5241.
11. **Dimitrova, M., I. Imbert, M. P. Kieney, and C. Schuster.** 2003. Protein-protein interactions between hepatitis C virus nonstructural proteins. J. Virol. **77:**5401–5414.
12. **Egger, D., B. Wölk, R. Gosert, L. Bianchi, H. E. Blum, D. Moradpour, and K. Bienz.** 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. J. Virol. **76:**5974–5984.
13. **Elazar, M., K. H. Cheong, P. Liu, H. B. Greenberg, C. M. Rice, and J. S. Glenn.** 2003. Amphipathic helix-dependent localization of NS5A mediates hepatitis C virus RNA replication. J. Virol. **77:**6055–6061.
14. **Enomoto, N., I. Sakuma, Y. Asahina, M. Kurosaki, T. Murakami, C. Yamamoto, N. Izumi, F. Marumo, and C. Sato.** 1995. Comparison of full-length sequence of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J. Clin. Investig. **96:**224–230.
15. **Enomoto, N., I. Sakuma, Y. Asahina, M. Kurosaki, T. Murakami, C. Yamamoto, Y. Ogura, N. Izumi, F. Marumo, and C. Sato.** 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N. Engl. J. Med. **334:**77–82.
16. **Ferrari, E., J. Wright-Minogue, J. W. S. Fang, B. M. Baroudy, J. Y. N. Lau, and Z. Hong.** 1999. Characterization of soluble hepatitis C virus RNA-dependent RNA polymerase expressed in *Escherichia coli.* J. Virol. **73:**1649–1654.
17. **Friebe, P., and R. Bartenschlager.** 2002. Genetic analysis of sequence in the 3′ nontranslated region of hepatitis C virus that are important for RNA replication. J. Virol. **76:**5326–5338.
18. **Gale, M., Jr., C. M. Blakely, B. Kwieciszewski, S. L. Tan, M. Dossett, N. M. Tang, M. J. Korth, S. J. Polyak, D. R. Gretch, and M. G. Katze.** 1998. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. Mol. Cell. Biol. **18:**5208–5218.
19. **Ghosh, A. K., M. Majumder, R. Steele, P. Yaciuk, J. Chrivia, R. Ray, and R. B. Ray.** 2000. Hepatitis C virus NS5A protein modulates transcription through a novel cellular transcription factor SRCAP. J. Biol. Chem. **275:** 7184–7188.
20. **Gosert, R., D. Egger, V. Lohmann, R. Bartenschlager, H. E. Blum, K. Bienz, and D. Moradpour.** 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. J. Virol. **77:**5487–5492.
21. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** 1993. Characterization of the hepatitis C virus-encoded serine protease: determination of proteinase-dependent polyprotein cleavage sites. J. Virol. **67:**2832–2843.
22. **Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone, and C. M. Rice.** 1993.

Expression and identification of hepatitis C virus polyprotein cleavage products. J. Virol. 67:1385–1395.

23. **Gu, B., A. T. Gates, O. Isken, S. E. Behrens, and R. T. Sarisky.** 2003. Replication studies using genotype 1a subgenomic hepatitis C virus replicons. J. Virol. 77:5352–5359.
24. **Guo, J. T., V. V. Bichko, and C. Seeger.** 2001. Effect of alpha interferon on the hepatitis C virus replicon. J. Virol. 75:8516–8523.
25. **Hijikata, M., H. Mizushima, Y. Tanji, Y. Komoda, Y. Hirowatari, T. Akagi, N.Kato, K. Kimura, and K. Shimotohno.** 1993. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proc. Natl. Acad. Sci. USA 90:10773–10777.
26. **Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
27. **Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, and K. Shimotohno.** 1991. Hypervariable regions in the putative glycoprotein of hepatitis C virus. Biochem. Biophys. Res. Commun. 175:220–228.
28. **Hügle, T., F. Fehrmann, E. Bieck, M. Kohara, H. G. Kräusslich, C. M. Rice, H. E. Blum, and D. Moradpour.** 2001. The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. Virology 284:70–81.
29. **Ikeda, M., M. Yi, K. Li, and S. M. Lemon.** 2002. Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. J. Virol. 76:2997–3006.
30. **Ivashkina, N., B. Wölk, V. Lohmann, R. Bartenschlager, H. E. Blum, F. Penin, and D. Moradpour.** 2002. The hepatitis C virus RNA-dependent RNA polymerase membrane insertion sequence is a transmembrane segment. J. Virol. 76:13088–13093.
31. **Kato, N., T. Nakazawa, T. Mizutani, and K. Shimotohno.** 1995. Susceptibility of human T-lymphotropic virus type I-infected cell line MT-2 to hepatitis C virus infection. Biochem. Biophys. Res. Commun. 206:863–869.
32. **Kato, N., Y. Ootsuyama, H. Sekiya, S. Ohkoshi, T. Nakazawa, M. Hijikata, and K. Shimotohno.** 1994. Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. J. Virol. 68:4776–4784.
33. **Kishine, H., K. Sugiyama, M. Hijikata, N. Kato, H. Takahashi, T. Noshi, Y. Nio, M. Hosaka, Y. Miyanari, and K. Shimotohno.** 2002. Subgenomic replicon derived from a cell line infected with the hepatitis C virus. Biochem. Biophys. Res. Commun. 293:993–999.
34. **Kolykhalov, A. A., K. Mihalik, S. M. Feinstone, and C. M. Rice.** 2000. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3′ nontranslated region are essential for virus replication in vivo. J. Virol. 74:2046–2051.
35. **Krieger, N., V. Lohmann., and R. Bartenschlager.** 2001. Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. J. Virol. 75:4614–4624.
36. **Lin, Y., H. Tang, T. Nomura, D. Dorjsuren, N. Hayashi, W. Wei, T. Ohta, R. Roeder, and S. Murakami.** 1998. The hepatitis B virus X protein is a coactivator of activated transcription that modulates the transcription machinery and distal binding activators. J. Biol. Chem. 273:27097–27103.
37. **Lohmann, V., F. Körner, A. Dobierzewska, and R. Bartenschlager.** 2001. Mutations in hepatitis C virus RNAs conferring cell culture adaptation. J. Virol. 75:1437–1449.
38. **Lohmann, V., F. Körner, J. O. Koch, U. Herian, L. Theilmann, and R. Bartenschlager.** 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. Science 285:110–113.
39. **Lohmann, V., F. Körner, U. Herian, and R. Bartenschlager.** 1997. Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. J. Virol. 71:8416–8428.
40. **Lohmann, V., S. Hoffmann, U. Herian, F. Penin, and R. Bartenschlager.** 2003. Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. J. Virol. 77:3007–3019.
41. **Major, M. E., and S. M. Feinstone.** 1997. The molecular virology of hepatitis C. Hepatology 25:1527–1538.
42. **Majumder, M., A. K. Ghosh, R. Steele, R. Ray, and R. B. Ray.** 2001. Hepatitis C virus NS5A physically associates with p53 and regulates p21/waf1 gene expression in a p53-dependent manner. J. Virol. 75:1401–1407.
43. **Mizutani, T., N. Kato, S. Saito, M. Ikeda, K. Sugiyama, and K. Shimotohno.** 1996. Characterization of hepatitis C virus replication in cloned cells obtained from a human T-cell leukemia virus type 1-infected cell line, MT-2. J. Virol. 70:7219–7223.
44. **Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers.** 1995. Classification and nomenclature of viruses: sixth report of the internal committee on taxonomy of viruses, p. 424–426. Springer-Verlag, Vienna, Austria.
45. **Murray, E. M., J. A. Grobler, E. J. Markel, M. F. Pagnoni, G. Paonessa, A. J. Simon, and O. A. Flores.** 2003. Persistent replication of hepatitis C virus replicons expressing the β-lactamase reporter in subpopulations of highly permissive Huh7 cells. J. Virol. 77:2928–2935.
46. **Pflugheber, J., B. Fredericksen, R. Sumpter, Jr., C. Wang, F. Ware, D. L. Sodora, and M. Gale, Jr.** 2002. Regulation of PKR and IRF-1 during hepatitis C virus RNA replication. Proc. Natl. Acad. Sci. USA 99:4650–4655.
47. **Piccininni, S., A. Varaklioti, M. Nardelli, B. Dave, K. D. Raney, and J. E. G. McCarthy.** 2002. Modulation of the hepatitis C virus RNA-dependent RNA polymerase activity by the non-structural (NS) 3 helicase and the NS4B membrane protein. J. Biol. Chem. 277:45670–45679.
48. **Qin, W., H. Luo, T. Nomura, N. Hayashi, T. Yamashita, and S. Murakami.** 2002. Oligomeric interaction of hepatitis C virus NS5B is critical for catalytic activity of RNA-dependent RNA polymerase. J. Biol. Chem. 277:2132–2137.
49. **Qin, W., T. Yamashita, Y. Shirota, Y. Lin, W. Wei, and S. Murakami.** 2001. Mutational analysis of the structure and functions of hepatitis C virus RNA-dependent RNA polymerase. Hepatology 33:728–737.
50. **Saito, I., T. Miyamura, A. Ohbayashi, H. Harada, T. Katayama, S. Kikuchi, Y. Watanabe, S. Koi, M. Onji, Y. Ohta, Q. Choo, M. Houghton, and G. Kuo.** 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA 87:6547–6549.
51. **Santolini, E., L. Pacini, C. Fipaldini, G. Migliaccio, and N. Monica.** 1995. The NS2 protein of hepatitis C virus is a transmembrane polypeptide. J. Virol. 69:7461–7471.
52. **Schmidt-Mende, J., E. Bieck, T. Hügle, F. Penin, C. M. Rice, H. E. Blum, and D. Moradpour.** 2001. Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. J. Biol. Chem. 276:44052–44063.
53. **Seeff, L. B.** 1997. Natural history of hepatitis C. Hepatology 26:21S–28S.
54. **Shirota, Y., H. Luo, W. Qin, S. Kaneko, T. Yamashita, K. Kobayashi, and S. Murakami.** 2002. Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRp) NS5B and modulates RNA-dependent RNA polymerase activity. J. Biol. Chem. 277:11149–11155.
55. **Sugiyama, K., N. Kato, T. Mizutani, M. Ikeda, T. Tanaka, and K. Shimotohno.** 1997. Genetic analysis of the hepatitis C virus (HCV) genome from HCV-infected human T cells. J. Gen. Virol. 78:329–336.
56. **Tan, S. L., H. Nakao, Y. He, S. Vijaysri, P. Neddermann, B. L. Jacobs, B. J. Mayer, and M. G. Katze.** 1999. NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. Proc. Natl. Acad. Sci. USA 96:5533–5538.
57. **Tsukiyama-Kohara, K., N. Iizuka, M. Kohara, and A. Nomoto.** 1992. Internal ribosome entry site within hepatitis C virus RNA. J. Virol. 66:1476–1483.
58. **Tu, H., L. Gao, S. T. Shi, D. R. Taylor, T. Yang, A. K. Mircheff, Y. Wen, A. E. Gorbalenya, S. B. Hwang, and M. M. C. Lai.** 1999. Hepatitis C virus RNA polymerase and NS5A Complex with a SNARE-like protein. Virology 263:30–41.
59. **Wang, C., P. Sarnow, and A. Siddiqui.** 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by internal ribosome-binding mechanism. J. Virol. 67:3338–3344.
60. **Wang, Q. M., M. A. Hockman, K. Staschke, R. B. Johnson, K. A. Case, J. Lu, S. Parsons, F. Zhang, R. Rathnachalam, K. Kirkegaard, and J. M. Colacino.** 2002. Oligomerization and cooperative RNA synthesis activity of hepatitis C virus RNA-dependent RNA polymerase. J. Virol. 76:3865–3872.
61. **Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford, M. Brunetto, P. J. Barr, T. Miyamura, J. McHutchinson, and M. Houghton.** 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. Proc. Natl. Acad. Sci. USA 89:3468–3472.
62. **Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q. L. Choo, M. Houghton, and J. H. Han.** 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180:842–848.
63. **Wölk, B., D. Sansonno, H. G. Kräusslich, F. Dammacco, C. M. Rice, H. E. Blum, and D. Moradpour.** 2000. Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines. J. Virol. 74:2293–2304.
64. **Yamashita, T., S. Kaneko, Y. Shirota, W. Qin, T. Nomura, K. Kobayashi, and S. Murakami.** 1998. RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus NS5B protein truncated at the C-terminal region. J. Biol. Chem. 273:15479–15486.
65. **Yanagi, M., M. St. Claire, S. U. Emerson, R. H. Purcell, and J. Bukh.** 1999. *In vivo* analysis of the 3′ untranslated region of the hepatitis C virus after *in vitro* mutagenesis of an infectious cDNA clone. Proc. Natl. Acad. Sci. USA 96:2291–2295.
66. **Yi, M., and S. M. Lemon.** 2003. 3′ Nontranslated RNA signals required for replication of hepatitis C virus RNA. J. Virol. 77:3557–3568.
67. **Zech, B., A. Kurtenbach, N. Krieger, D. Strand, S. Blencke, M. Morbitzer, K. Salassidis, M. Cotten, J. Wissing, S. Obert, R. Bartenschlager, T. Herget, and H. Daub.** 2003. Identification and characterization of amphiphysin II as a novel cellular interaction partner of the hepatitis C virus NS5A protein. J. Gen. Virol. 84:555–560.

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Biochemical and Biophysical Research Communications 315 (2004) 1088–1096

BBRC

www.elsevier.com/locate/ybbrc

A low-density cDNA microarray with a unique reference RNA: pattern recognition analysis for IFN efficacy prediction to HCV as a model☆

Akito Daiba,[a] Niro Inaba,[a] Satoshi Ando,[a] Naoki Kajiyama,[a] Hiroshi Yatsuhashi,[b] Hiroshi Terasaki,[a] Atsushi Ito,[a] Masanori Ogasawara,[a] Aki Abe,[a] Junichi Yoshioka,[a] Kazuhiro Hayashida,[c,1] Shuichi Kaneko,[d] Michinori Kohara,[e] and Satoru Ito[a,*,2]

[a] *JGS Japan Genome Solutions, Inc. 51 Komiya-cho, Hachioji, Tokyo 192-0031, Japan*
[b] *National Nagasaki Medical Center, Kubara 2-1001-1 Omura City, Nagasaki 856-8562, Japan*
[c] *Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan*
[d] *Kanazawa University Graduate School of Medical Sciences, 13-1 Muro-machi, Kanazawa 920-8641, Japan*
[e] *Tokyo Metropolitan Institute of Medical Science, 3-18-22 Hon-komagome, Bunkyo-ku, Tokyo 113-8613, Japan*

Received 9 January 2004

Abstract

We have designed and established a low-density (295 genes) cDNA microarray for the prediction of IFN efficacy in hepatitis C patients. To obtain a precise and consistent microarray data, we collected a data set from three spots for each gene (mRNA) and using three different scanning conditions. We also established an artificial reference RNA representing pseudo-inflammatory conditions from established hepatocyte cell lines supplemented with synthetic RNAs to 48 inflammatory genes. We also developed a novel algorithm that replaces the standard hierarchical-clustering method and allows handling of the large data set with ease. This algorithm utilizes a standard space database (SSDB) as a key scale to calculate the Mahalanobis distance (MD) from the center of gravity in the SSDB. We further utilized sMD (divided by parameter k: MD/k) to reduce MD number as a predictive value. The efficacy prediction of conventional IFN mono-therapy was 100% for non-responder (NR) vs. transient responder (TR)/sustained responder (SR) ($P < 0.0005$). Finally, we show that this method is acceptable for clinical application.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Low-density microarray; Artificial reference RNA; Efficacy prediction; Mahalanobis distance

Large data sets can be collected from cDNA microarrays for the study of expression profiles in biological systems. The large amount of data generated can be especially useful to help cluster genes into interest groups. Such genome-wide information can be used for clinical applications (see reviews, [1–5]), for example, for the identification of the cDNA expression patterns associated with different stages of tumor development. However, translating complex microarray data into practical clinical applications has been difficult. New algorithms are being developed to solve this problem, for example for the prognosis of cancer treatment [6,7]. Also, low-density microarrays with selected genes of interest can simplify the analysis of microarray data.

Another critical issue in understanding microarray data is the level of precision in the data set. Solid phase DNA hybridization is not as quantitative as hybridization in solution, and scanners have limited dynamic

☆ *Abbreviations:* SAGE, serial analysis of gene expression; SSDB, standard space database; MD, Mahalanobis distance; sMD, scaled MD; FL, firefly luciferase gene; RL, *Renilla* luciferase gene; GP, baculovirus glycoprotein gene; LD, lambda DNA; MEP, microcapillary electrophoretic; aRNA, amplified RNA; NR, non-responder; SR, sustained responder; TR, transient responder.

* Corresponding author. Fax: +81-426-45-0461.
E-mail address: sr-itou@jgs-inc.co.jp (S. Ito).

[1] Present address: Sasebo Kyosai Hospital, 10-17 Shimaji, Sasebo, Nagasaki 857-8575, Japan.

[2] Visiting position: Medical Research Institute, Tokyo Medical and Dental University.

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.01.160

ranges. In addition, sample variability can result from variations in sampling conditions, RNA amplification, RNA degradation, and cDNA labeling conditions. These factors are not well understood, and, currently, precision is primarily controlled at the level of data collection [8–10].

One means of enhancing the precision of data collection is to use reference controls for each individual. This can be accomplished by laser-captured microdissection of tissues into diseased and counter part areas [11,12]. In case of hepatitis C, this is difficult because inflammatory damage occurs throughout the liver. Another possible approach is to utilize artificial reference RNAs as a reference [13–15] in conjunction with RNA from established cell lines, such as hepatocellular carcinoma cell lines. However, the stages and characteristics of the disease in vivo and in vitro can differ, and expression of some RNAs of interest, especially some inflammatory genes, may be too low in the cultured cells to produce a satisfactory signal-to-noise ratio.

In the current studies, we identified inflammatory genes that can be used for a low-density microarray to predict the efficacy of INF treatment in hepatitis C patients. We found sufficient levels of expression for these genes in patient samples, but very low levels of expression in established cell lines. We overcame this problem by using a low-density cDNA microarray system in conjunction with a unique artificial reference RNA. In addition, we describe an algorithm for analysis of the microarray data.

Methods

cDNA Microarray. We selected 295 genes for the cDNA microarray based on publicly available data, including SAGE analysis and other DNA microarray data from HCV patients and a normal subject [16,17]. From 2000 candidate genes, we omitted low frequency-tag genes based on the SAGE data. Genes previously identified as predictive host factors for IFN efficacy [18–20] were given a high priority. For most of the genes, each cDNA was designed approximately 500–600 bp and within approximately 1 kb region from the 3′-poly(A) tail and all cDNAs for microarray probe were cloned into the pGEM vector (Promega, Madison, WI). We also selected and cloned external control genes (approximately 0.5–1 kb) into the pGEM vector to establish the dynamic range of the microarray. These genes were firefly luciferase (FL; negative control), *Renilla* luciferase (RL) [21], baculovirus envelope gp64 (GP), and lambda phage DNA (LD). All clones for capture probe on microarray were sequenced and validated by comparison with the GenBank sequence. The aminosilane surfaces of SuperAmine glass slides (TeleChem International, Sunnyvale, CA) were stamped with triplicate spots of cDNA probe corresponding to each of the remaining 295 genes. The average spot size was 150 μm and separated each other with a distance (500 μm) as shown in Fig. 2B.

Reference RNA preparation. Extracted total RNAs from four hepatocellular carcinoma cell lines (HepG2, Hep3B, Huh7, and IMY-4 [22]) purified through RNeasy column (Qiagen, Hilden, Germany) were mixed as a reference source. In order to find a mixing ratio of four cell derived RNAs and provide reliable reference source, we measured the copy number of certain genes in each cell line by real-time PCR method. Using real-time PCR with the PRISM 7900HT system (Applied Biosystems, Foster City, CA), we measured the copy number of several genes, including the IFN-α/β receptor, double-strand RNA-activated protein kinase (PKR), 2′,5′-oligoadenylate synthetase (2,5-AS), interferon regulatory factor-1 (IRF-1), interferon regulatory factor-3 (IRF-3), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The primer sets were as follows: IFN-α/β receptor (forward: GTGACCTCACAGATGAGTGG; reverse: CCCTCTGA CTGTTCTTCAATGA; and probe: CACCGTCCTAGAAGGA TTCAGCGG), PKR (forward: CCTGTCCTCTGGTTCTTTTG; reverse: TGTCAGGAAGGTCAAATCTG; and probe: CTACGTGT GAGTCCCAAAGCAAC), 2,5-AS (forward: CTCAGAAATAC CCCAGCCAAATC; reverse: GTGGTGAGAGGACTGAGGAA; and probe: CCAGGTCAGCGTCAGATCGGCCTC), IRF-1 (forward: GCAAGGCCAAGAGGAAGTCA; reverse: TCATCAGG CAGAGTGGAGCT; and probe: TTCCAGCCCTGATACCTTC TCTGATGG), IRF-3 (forward: AAGGAAGGAGGCGTGTTTGA; reverse: ATTTCTACCAAGGCCCTGAGG; and probe: CGTCCGC TTCCTTCCGTGAAGGTAAT), GAPDH (forward: GAAGGTGA AGGTCGGAGT; reverse: GAAGATGGTGATGGGATTTC; and probe: CAAGCTTCCCGTTCTCAGCC). The RNA mixture was amplified using MessageAmp aRNA kit (Ambion, Austin, TX). The resulting aRNA was used as the reference aRNA. Moreover, we cloned genes (RL, GP, and LD; ~1000 bp) into pCRII TOPO vector (Invitrogen, Carlsbad, CA) as scanning range markers as well as 48 genes (in the same TOPO vector) of inflammatory genes to spike into reference aRNA. Each cloning region was designed to be larger than the size of capture probes on the microarray. Then three external control RNAs and additional spike RNAs of 48 genes were synthesized by Megascript T7 kit (Ambion, Austin, TX). Three external control RNAs were mixed as spike control mixture in both target sample and reference aRNAs. Other 48 spike RNAs were added to the reference aRNA.

Sample RNA preparation, labeling, hybridization, and scanning. Total RNA of liver biopsy samples was isolated by Isogen (Nippon Gene, Tokyo, Japan) extraction according to manufacturer's instruction. The total RNA quality was confirmed with a Bioanalyser 2100 microcapillary-electrophoretic (MEP) analyzer (Agilent Technologies, Palo Alto, CA). The 28S/18S ratio of the total RNA was >1.0. Then total RNA (1–2 μg) was transcribed and amplified to produce amplified sample RNA (aRNA) using the MessageAmp aRNA kit (Ambion, Austin, TX) according to manufacturer's instructions. Next, an external control RNA mixture (LD, GP, and RL) was added to both the sample and reference aRNAs. These mixed sample and reference aRNAs were labeled using SuperScript II kit with random hexamer (TaKaRa, Kyoto, Japan) with Cy3-dUTP and Cy5-dUTP (Perkin–Elmer, Boston, MA), respectively. Competitive hybridization of Cy3-labeled sample and Cy5-labeled reference cDNAs on the microarray was carried out according to Brown and coworkers [23]. Slides were scanned three times with ScanArray 5000 (Perkin–Elmer, Boston, MA). Each scan was carried out using the external spike level around 30,000. The data were converted from tif image data to signal using ImaGene software (BioDiscovery, El Segundo, CA) for further statistical analysis. Three file data of each three spot data of each gene were merged to establish the single representative data for each gene (Patent pending: PCT/JP03/06677). The Cy3 (patient sample)/Cy5 (reference sample) ratio of each mRNA signal was calculated for further analysis.

Patients. Liver biopsy samples from five patients receiving IFN-α monotherapy and 10 patients receiving a combination therapy (a mixture of IFN-α, IFN-β, and IFN-α/β) were obtained during 1992–2000 from Kyushu University Hospital and Nagasaki National Medical Center, respectively. Biopsy samples were stored at −80 °C. Informed consent was obtained from all patients in accordance with the Helsinki protocol. The samples were classified into three groups: sustained responders (SR) had an absence of serum HCV RNA both during the therapy and 6 months after the completion of therapy, non-responders (NR) were positive for serum HCV RNA during the

therapy, and transient responders (TR) had an absence of serum HCV RNA during the therapy or at the end of IFN treatment, but has serum HCV RNA after cessation of the therapy. Because RNA degradation may have occurred during storage, and because this can be a major source of variation in microarray data [24], we verified the quality of the extracted RNA by assessing the ribosomal RNA 28S/18S ratio.

Statistical data analysis. The merged data were subjected for hierarchical clustering to noise reduction and normalization (patent pending, PCT/JP03/06677) using the reference control and then analyzed with Genomic Profiler software (Mitsui Knowledge Industry, Tokyo, Japan). In addition, we developed a novel algorithm to calculate the Mahalanobis distance (MD) for the data from 15 patients using a standard space database (SSDB) (Eqs. (1)–(5) and Fig. 1). To establish the SSDB, we searched a gene set representing the differences between the SR/TR and NR groups. The necessary genes for the SSDB and MD calculations were selected using MATLAB (MathWorks, Natick, MA). We have calculated a graded scale utilizing variance–covariance to valuate dispersion and correlation of the standard group as a training set to establish the center of the gravity of SSDB. Once the SSDB was established, new test sample data were applied to the equations to calculate the MD. We utilize sMD as a predictive value. The sMD was presented from the center of gravity of SSDB (0:zero) along its scale to theoretically ∞. This method is one of the pattern recognition analysis dealing with correlation of multi-parameters [25].

$$d_{xi} = \frac{D_{xi} - \overline{D_x}}{\sigma_x} \qquad \text{(auto scale)}, \quad (1)$$

$$S_{xx'} = \frac{\sum_{i=1}^{n} (d_{xi} - \overline{d_x})(d_{x'i} - \overline{d_{x'}})}{n-1} \qquad \text{(variance–covariance)}, \quad (2)$$

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1(k-1)} & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2(k-1)} & S_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{(k-1)1} & S_{(k-1)2} & \cdots & S_{(k-1)(k-1)} & S_{(k-1)k} \\ S_{k1} & S_{k2} & \cdots & S_{k(k-1)} & S_{kk} \end{bmatrix} \qquad \text{(variance–covariance matrix)}, \quad (3)$$

$$\mathrm{MD}^2 = [d_1 \ \ \ldots \ \ d_k] S^{-1} \begin{bmatrix} d_1 \\ \vdots \\ d_k \end{bmatrix} \qquad \text{(Mahalanobis distance)}, \quad (4)$$

$$\mathrm{sMD} = \frac{\mathrm{MD}^2}{k} \qquad \text{(scaled Mahalanobis distance)}. \quad (5)$$

Results and discussion

The low-density cDNA microarray

High-density microarray data were so hard to handle its huge data for analysis and difficult to understand their meaning. One approach is to minimize gene set for collection of mRNA profiling data to each category of research field. Chang et al. [26] have described the selection of data from high-density microarrays for prediction of "docetaxel" therapy efficacy of breast cancer. Specifically, they omitted low level signals of genes from the high-density microarray data at first. We followed a similar approach to select genes on our microarray, also omitting low frequency tag genes from the SAGE data. This ensured a steady state signal amongst the target samples. Based on this selection, we chose 295 genes for a low-density microarray system. The DNA sequence of each cloned gene fragment (500–600 bp) was validated by comparison with the published sequences in the GenBank database. To provide replicate data, the cDNAs were spotted in triplicate on the aminosilane-coated slide glass [8]. Scanning electron microscopy confirmed that the spots were round, smooth, and homogeneous without any doughnut features (Fig. 2). To obtain stable signals, we used three independent internal RNA references, including RL for the high expression range, GP for the middle range, and LD for the low range. The signals from the microarray were adjusted so that the ScanArray would give reliable signal of 30,000, which should be within the linear and stable range of the scanner (maximum signal = 65,535). We also carried out noise reduction and normalization of data using the artificial external spike genes as well as some housekeeping genes. Validation of the low-density cDNA microarray system was carried out using the RNA from HepG2 (data not shown).

Adjustment of reference RNA

Selection of an appropriate control reference is another important factor for accuracy in microarray analysis. One method has been to use laser microdissection to select normal tissue from the same patient as a reference. Although this is useful for single patients, it cannot be applied to comparison of multiple patients' samples because the baseline expression of specific genes can vary from patient to patient. Therefore, conditions including the duration of disease, the medications used, sampling conditions, storage conditions, and life style differences can cause variability in the microarray data. To eliminate this problem, we have used an artificial reference RNA isolated from cell lines as a reference. When we screened the signal levels of both reference and patient samples, we found 48 genes out of 295 genes in the microarray that were expressed in the patients but not or background level in the RNA mixture from the four cell lines. Typical data from the four cell reference mixture are shown in Fig. 3A. The IFN-receptor and some other well-known IFN-inducible factors are indicated. The graph shows that the levels of these mRNAs are in the low signal range. This includes the IFN-α/β receptor, even though it has been proposed as a possible marker for the prediction of IFN efficacy [18,19]. The problem in this case appears to be high variability in IFN-α/β receptor gene expression between different reference RNA preparations. To avoid this problem, we produced a large single preparation of reference RNA for future analyses. In addition, we have produced 48 synthetic RNAs, which we added to the reference RNA mixture. These synthetic RNAs were designed to be larger than the size of the capture probes on the microarray. The design and purity of some these synthetic

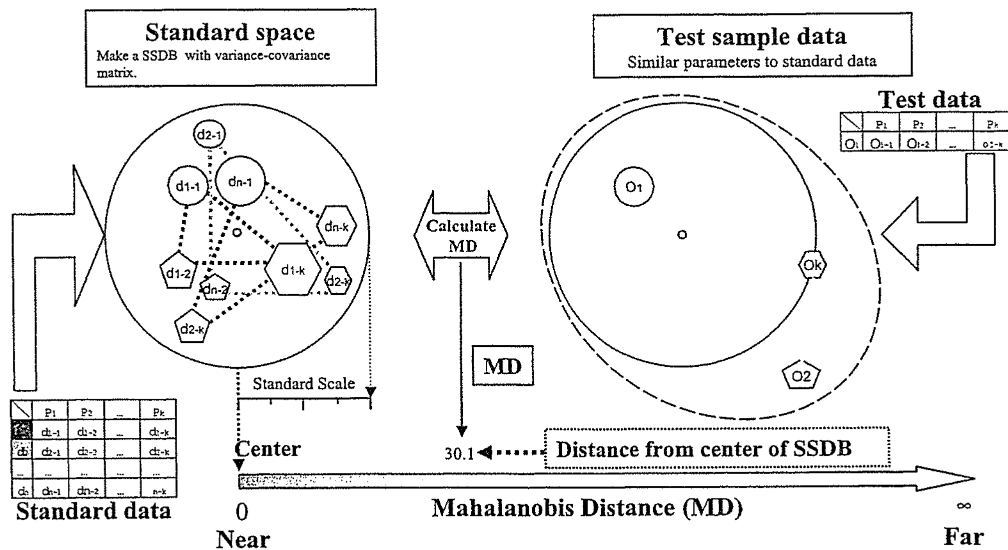


Fig. 1. Pattern recognition for establishment of SSDB and MD calculations. The standard space database (SSDB) was established based on a training data set. The parameter (1 to k) represents each data point (signal level for each gene) in the patient sample. A second parameter (1 to n) represents each patient. Both parameters were utilized along with the distinguishing genes for each group that were used in a variance–covariance matrix calculation to create the SSDB. Next, the test sample data were calculated to obtain the Mahalanobis distance (MD), where the MD represents the distance of new test sample data from the center of gravity in the SSDB. In theory, MD can be from 0 to ∞. A high MD value means that its distance is far away from the SSDB center of gravity.

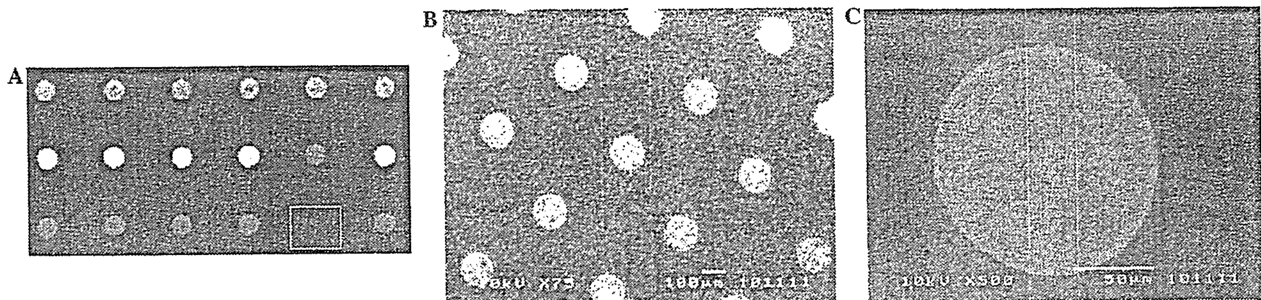


Fig. 2. Spot geometry on the cDNA microarray. (A) A typical fluorescent image was presented. The negative control (firefly luciferase) gene was spotted inside of the white square. Scanning electron microscopic images at (B) low power (75×) and (C) higher power (500×) are shown.

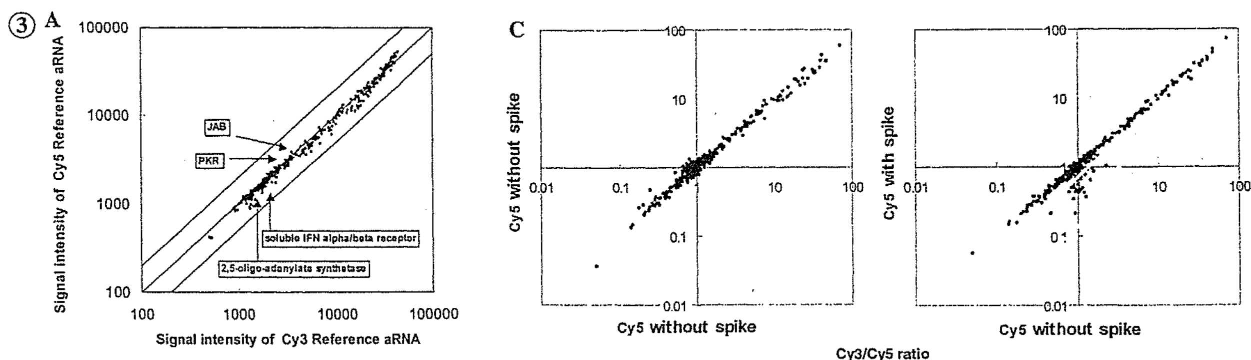
RNAs is shown in Fig. 3B. The level of the synthetic reference RNAs is shown as a scatter plot in Fig. 3C. These results show that the level of the synthetic RNA corresponds to the range of pseudo-inflammatory conditions. Thus, we spiked these 48 synthetic RNAs to the average signal level of patients which was surveyed at first. Without this synthetic reference RNA, it would be difficult to analyze and categorize the microarray data and use it to predict the efficacy of IFN treatment. Thus, the signal below the negative cut-off level will be treated as zero for further ratio calculation leads to nonsense.

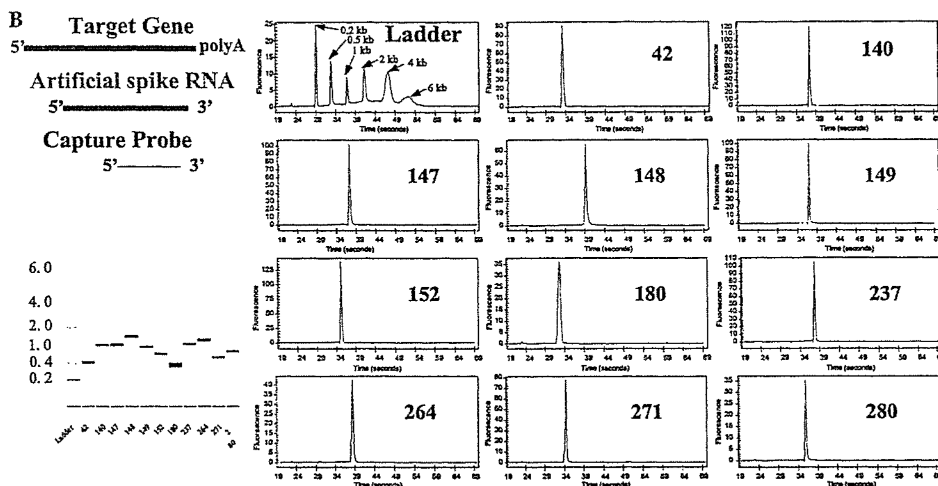
Statistical analysis

A variety of normalization techniques have been used for the analysis of DNA microarray data [9,27–30]. Many of these techniques rely on the expression of housekeeping genes. However, it is difficult to find suitable candidates, and it would be difficult to integrate a large set of housekeeping genes onto the low-density cDNA microarray [30–32]. For these reasons, we have added synthetic non-human genes as external controls. We have also utilized some type of housekeeping genes for normalization of the microarray data (patent pending, PCT/JP03/06677). Furthermore, to minimize variability in the calculations due to variability in the fluorescence measurements, we used six data files (three scans of each Cy3/Cy5 wavelength) to merge into a single representative data for each gene expression analysis.

Hierarchical clustering by the classical method

Hierarchical clustering of the merged data was carried out using Euclid distance and Ward method with


③ A
Signal intensity of Cy5 Reference aRNA
Signal intensity of Cy3 Reference aRNA
JAB
PKR
soluble IFN alpha/beta receptor
2,5-oligo-adenylate synthetase
C
Cy5 without spike
Cy5 with spike
Cy3/Cy5 ratio


B
Target Gene
5' polyA
Artificial spike RNA
5' 3'
Capture Probe
5' 3'
Ladder
42
140
147
148
149
152
180
237
264
271
280

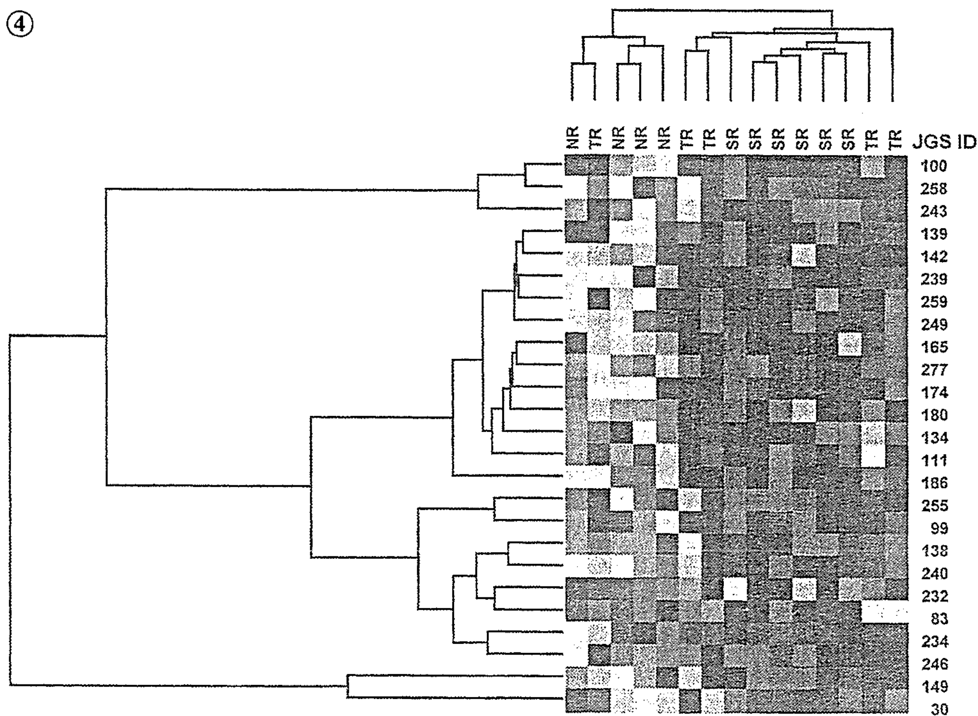

④
JGS ID
NR TR NR NR NR TR TR SR SR SR SR SR SR TR TR
100
258
243
139
142
239
259
249
165
277
174
180
134
111
186
255
99
138
240
232
83
234
246
149
30

Fig. 3. Establishment of the artificial reference RNA. (A) Scatter plot of the mRNA signal level in the four cell RNA mixture. The RNA mixture from four cell lines was covalently modified with Cy3 and Cy5 dyes. The signal levels for typical inflammatory genes, including 2,5-oligoadenylate synthetase (2,5-AS), IFN-α/β receptor, IFN-α receptor, and PKR, are indicated with arrows. (B) Design and purity of the synthetic RNA. The general design of the synthetic RNA (positional relation), including size and position of the probe and the artificial RNA, is shown on the left. Thus, each reference RNA was designed to be longer than the captured probe on the microarray, but not to exceed the size of the target RNA. The purity of the 11 of the synthetic RNA samples is shown. The corresponding gene numbers on our cDNA microarray are shown in each panel and are as follows (GenBank accession number in parentheses): 42, gamma-G globin (X55656); 140, T cell activation antigen (CD27) (M63928); 147, (2′,5′)-oligoadenylate synthetase (D00068); 148, p68 kinase (M35663); 152, CIS3 (AB006967); 180, calcium-binding protein in macrophages (MRP-14), also known as macrophage migration inhibitory factor-related protein (X06233); 237, interleukin 2 receptor β chain (p70-75) (M26062); 264, interferon-induced protein 44 (IFI44) (NM_006417); 271, interleukin 4 (M13982); and 280, hepatocyte growth factor (X16323). (C) Synthetic spiked RNA signal level. The panel on the left represents the scatter plot without any synthetic RNA added to the reference RNA, while the panel on the right shows the reference with added synthetic RNAs. The plot shows the ratio of sample Cy3/Cy5 rather than real signal level. The spots in the white rectangle represent the level of the added synthetic RNAs.

Fig. 4. Hierarchical clustering. cDNA microarray data of 15 patients' samples were analyzed with Genomic Profiler software (MKI, Japan). For clustering, normalization, filtering, and T test were essential. Because of interest in predicting clinical outcomes of IFN treatment, we tried to classify the data into two groups, including non-responders (NR) and transient responders (TR)/sustained responders (SR). The accuracy of this prediction was >93%. The corresponding microarray number and according GenBank accession number of the genes responsible for clustering are shown on the right and include: 100, cytoplasmic dynein light chain 1 (U32944); 258, thymosin β-10 (M92383); 243, stathmin (X53305); 139, homeobox 1.4 protein (M74297); 142, cAMP-dependent protein kinase regulatory subunit RI-bet (M65066); 239, alternatively spliced interferon receptor (IFNAR2) (L42243); 259, eukaryotic translation initiation factor 2, subunit 1 α, 35 kDa (BC002513); 249, brain-derived neurotrophic factor precursor (BDNF) (M61176); 165, interleukin 2 (X01586); 277, natural killer cell stimulatory factor (NKSF) (M65290); 174, IFN-responsive transcription factor subunit (M87503); 180, calcium-binding protein in macrophages (MRP-14) also known as macrophage migration inhibitory factor-related protein (X06233); 134, lunatic fringe U94354); 111, protein tyrosine kinase (Syk) (L28824); 186, leukocyte-associated molecule-1 α subunit (LFA-1 α subunit) (Y00796); 255, FLICE-like inhibitory protein short form (U97075); 99, CDK4-inhibitor (p16-INK4) (L27211); 138, α 7B integrin (X74295); 240, interferon-stimulated T-cell α chemoattractant precursor(AF030514); 232, Charcot–Leyden crystal protein (L01664); 83, NADH:ubiquinone oxidoreductase MLRQ subunit (U94586); 234, apoptotic cysteine protease Mch4 (Mch4) (U60519); 246, metallothionein-III (M93311); 149, interferon regulatory factor 1 (X14454); and 30, heat shock 70 kDa protein 1A (BC002453).

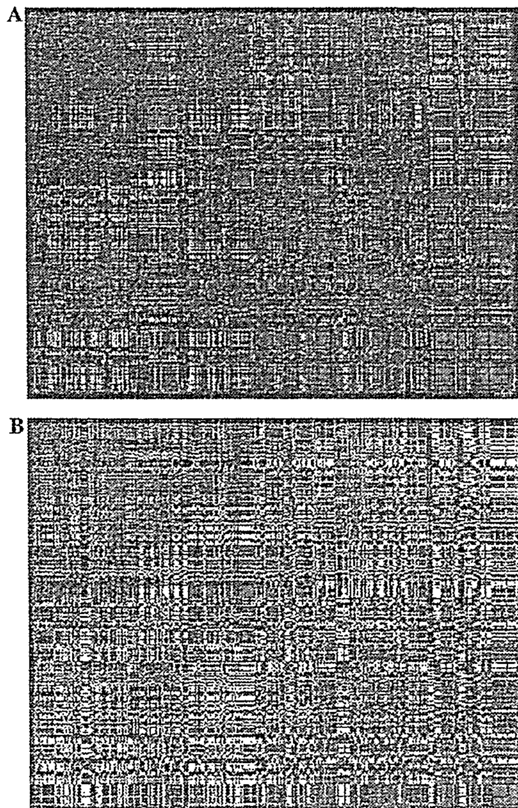


Fig. 5. Typical expression pattern: correlation matrix of 295 gene signals. (A) This expression pattern represented the established standard data as SSDB. Each axis represents genes in consideration. (B) The expression pattern shows an example of NR data as new test sample data. Color and brightness are adjusted according to Eqs. (1)–(3).

Patient No	Prediction	Clinical outcome	sMD (MD/k)*1	
10	NR	NR	6.31	so far from SSDB center of gravity
9	NR	NR	5.97	
8	NR	NR	4.38	
15	NR	NR	3.53	
4	SR or TR	SR	1.96	threshold =2.0
12	SR or TR	TR	1.90	
13	SR or TR	SR	1.44	
1	SR or TR	SR	1.36	
14	SR or TR	TR	1.09	
6	SR or TR	TR	0.98	
3	SR or TR	SR	0.73	
7	SR or TR	SR	0.67	
11	SR or TR	TR	0.38	
5	SR or TR	TR	0.24	
2	SR or TR	SR	0.24	SSDB center of gravity

		Clinical Outcome			
		SR or TR	NR	Total	%
Prediction	SR or TR	11	0	11	100%
	NR	0	4	4	100%
	Total	11	4	15	
	%	100%	100%		

$P < 0.0005$

Fig. 6. Mahalanobis distance (MD) and classification for efficacy prediction. The resulting MD data were arranged along the scale based on numbers of sMD value. The first column shows the patient number, the second column shows the prediction from our microarray analysis, and the third column shows the actual clinical outcome. Blue represents the NR group, purple represents TR, and red represents the SR group. We have set a threshold at the sMD level of 2.0 to set two groups from the sMD calculation. The hit ratio of prediction to clinical outcome was also shown.

Genomic Profiler software. Because of our interest in predicting IFN efficacy in hepatitis C, we compared the NR group against SR/TR group [33–35]. We chose genes with a 5% discriminated expression pattern (T test) between two groups ([SR vs. NR] $\cap$ [(SR + TR) vs. NR]). The hierarchical tree for these genes is shown in Fig. 4. Appropriate groups were assigned to all but one case, a TR case that was classified into the NR group.

Development of a new algorithm for hierarchical clustering

Although the classical method dealing with multi-parameters allowed satisfactory classification of the patients into two groups, this method is not useful for practical purposes. Thus, the entire data set is necessary for interpretation of the results from the classical analysis even for the analysis of a single test sample. In general, clinical diagnosis requires the common scale to compare the analytical data from samples. But a classical classification presents only a relative scale among samples for comparison.

For these reasons, we have developed a new algorithm. This method is based on the calculation of MD. The concept of the MD calculation is outlined in Fig. 1. This is one of the pattern recognition analysis and determines how close or how far from the standard group of interest. Thus, it deals with multi-parameters leading to single parameter, Mahalanobis distance (MD) as a scale, from the center of gravity of SSDB established by a training set shown in Fig. 1. Prior to the MD calculation, it was necessary to establish a SSDB with a training data set randomly selected but with clear and known clinical outcome. The standard expression pattern obtained from the SR/TR group, which was the established data source for SSDB, is shown in Fig. 5A and the new test sample pattern of the NR group is shown in Fig. 5B. The red color represents the higher expression profile and green depicts the lower expression profile with an interrelating style at the same time. Then, the pattern recognition algorithm Eqs. (1)–(3) was applied to compare the two groups. Among these expression patterns, we selected the stably and differentially expressed gene set data. Then, every selected gene expression pattern was correlated to each other like connecting a network. Thus, based on Eqs. (1)–(3), the elements that distinguish the groups shown were selected to create the SSDB. The SSDB was created based on variance and covariance. Once we established the SSDB and the center of gravity of the SSDB, we calculated a MD value for each new test sample (Eq. (4)). We utilized sMD value (divided the MD value by the number of parameters) to reduce MD value (Eq. (5)) and simplify understanding of the results[25]. The classification of IFN efficacy prediction to hepatitis C patients is shown in Fig. 6 and clearly shows that this analysis generates the NR and SR/TR groups and they are predicted with 100% accuracy ($p < 0.005$). The sample size (15 cases) was too small for statistical validation. Further detailed analyses and validation are ongoing in our laboratory and will be reported elsewhere.

Although the MD method is popular in even biological system publications [36,37], an application of MD method to microarray data is not so popular yet. The current studies are not the first published report where MD for analysis of gene expression data [38]. However, that report focused on the differential expression of a causative gene in conjunction with a standard hierarchical clustering algorithm. In our studies, we have attempted to scale the test sample position as a simple understandable parameter with a new pattern recognition algorithm. Once the SSDB scale has been established for a project, the MD can be easily used to classify new data according to the NR and TR/SR groups on an absolute scale (Fig. 6). This system will be acceptable for clinicians as a simple system for understanding the microarray data.

Conclusions

Besides the technical issues, there are many factors that control the variability within a microarray system, including individual differences between patients, the duration of the disease of each patient, different therapeutic protocols, and complications with other diseases. Some of these factors interact with each other, while others are independent. Therefore, an algorithm that allows some variability in the measurements is needed for prediction of therapeutic outcomes. The algorithm presented here appears to satisfy this requirement and it simplifies handling of large data sets. Finally, this algorithm should be generally applicable to the prediction of therapeutic outcome of diseases.

Acknowledgments

We thank Professor Kazunari Akiyoshi and Dr. Akihiko Watanabe (Tokyo Medical and Dental University) for scanning electron microscopy and Professor Kouji Matsushima (University of Tokyo) for helpful discussions.

References

[1] M. Sanchez-Carbayo, Use of high-throughput DNA microarrays to identify biomarkers for bladder cancer, Clin. Chem. 49 (2003) 23–31.

[2] P.F. Macgregor, J.A. Squire, Application of microarrays to the analysis of gene expression in cancer, Clin. Chem. 48 (2002) 1170–1177.

[3] C.H. Chung, P.S. Bernard, C.M. Perou, Molecular portraits and the family tree of cancer, Nat. Genet. 32 (Suppl.) (2002) 533–540.

[4] N.L. Harris, H. Stein, S.E. Coupland, M. Hummel, R.D. Favera, L. Pasqualucci, W.C. Chan, New approaches to lymphoma diagnosis, Hematology (Am Soc Hematol Educ Program) (2001) 194–220.
[5] L.T. Lam, O.K. Pickeral, A.C. Peng, A. Rosenwald, E.M. Hurt, J.M. Giltnane, L.M. Averett, H. Zhao, R.E. Davis, M. Sathyamoorthy, L.M. Wahl, E.D. Harris, J.A. Mikovits, A.P. Monks, M.G. Hollingshead, E.A. Sausville, L.M. Staudt, Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol, Genome Biol 2 (2001) research0041.
[6] D.G. Beer, S.L. Kardia, C.C. Huang, T.J. Giordano, A.M. Levin, D.E. Misek, L. Lin, G. Chen, T.G. Gharib, D.G. Thomas, M.L. Lizyness, R. Kuick, S. Hayasaka, J.M. Taylor, M.D. Iannettoni, M.B. Orringer, S. Hanash, Gene-expression profiles predict survival of patients with lung adenocarcinoma, Nat. Med. 8 (2002) 816–824.
[7] E. Huang, S.H. Cheng, H. Dressman, J. Pittman, M.H. Tsou, C.F. Horng, A. Bild, E.S. Iversen, M. Liao, C.M. Chen, M. West, J.R. Nevins, A.T. Huang, Gene expression predictors of breast cancer outcomes, Lancet 361 (2003) 1590–1596.
[8] M.L. Lee, F.C. Kuo, G.A. Whitmore, J. Sklar, Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations, Proc. Natl. Acad. Sci. USA 97 (2000) 9834–9839.
[9] G.C. Tseng, M.K. Oh, L. Rohlin, J.C. Liao, W.H. Wong, Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects, Nucleic Acids Res. 29 (2001) 2549–2557.
[10] F. Diehl, S. Grahlmann, M. Beier, J.D. Hoheisel, Manufacturing DNA microarrays of high spot homogeneity and reduced background signal, Nucleic Acids Res. 29 (2001) E38.
[11] L. Luo, R.C. Salunga, H. Guo, A. Bittner, K.C. Joy, J.E. Galindo, H. Xiao, K.E. Rogers, J.S. Wan, M.R. Jackson, M.G. Erlander, Gene expression profiles of laser-captured adjacent neuronal subtypes, Nat. Med. 5 (1999) 117–122.
[12] O. Kitahara, Y. Furukawa, T. Tanaka, C. Kihara, K. Ono, R. Yanagawa, M.E. Nita, T. Takagi, Y. Nakamura, T. Tsunoda, Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia, Cancer Res. 61 (2001) 3544–3549.
[13] L. Assersohn, L. Gangi, Y. Zhao, M. Dowsett, R. Simon, T.J. Powles, E.T. Liu, The feasibility of using fine needle aspiration from primary breast cancers for cDNA microarray analyses, Clin. Cancer Res. 8 (2002) 794–801.
[14] D.A. Wigle, I. Jurisica, N. Radulovich, M. Pintilie, J. Rossant, N. Liu, C. Lu, J. Woodgett, I. Seiden, M. Johnston, S. Keshavjee, G. Darling, T. Winton, B.J. Breitkreutz, P. Jorgenson, M. Tyers, F.A. Shepherd, M.S. Tsao, Molecular profiling of non-small cell lung cancer and correlation with disease-free survival, Cancer Res. 62 (2002) 3005–3008.
[15] T.C. Van Der Pouw Kraan, F.A. Van Gaalen, T.W. Huizinga, E. Pieterman, F.C. Breedveld, C.L. Verweij, Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair, Genes Immun. 4 (2003) 187–196.
[16] T. Yamashita, S. Hashimoto, S. Kaneko, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, K. Matsushima, Comprehensive gene expression profile of a normal human liver, Biochem. Biophys. Res. Commun. 269 (2000) 110–116.
[17] T. Yamashita, S. Kaneko, S. Hashimoto, T. Sato, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, K. Matsushima, Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma, Biochem. Biophys. Res. Commun. 282 (2001) 647–654.
[18] H. Yatsuhashi, K. Yamasaki, T. Aritomi, P. Maria, D. Carmen, O. Inoue, M. Koga, M. Yano, Quantitative analysis of interferon alpha/beta receptor mRNA in the liver of patients with chronic hepatitis C: correlation with serum hepatitis C virus-RNA levels and response to treatment with interferon, J. Gastroenterol. Hepatol. 12 (1997) 460–467.
[19] E. Mizukoshi, S. Kaneko, M. Yanagi, H. Ohno, K. Kaji, S. Terasaki, A. Shimoda, E. Matsushita, K. Kobayashi, Expression of interferon alpha/beta receptor in the liver of chronic hepatitis C patients, J. Med. Virol. 56 (1998) 217–223.
[20] F.L. Dumoulin, U. Wennrich, H.D. Nischalke, L. Leifeld, H.P. Fischer, T. Sauerbruch, U. Spengler, Intrahepatic mRNA levels of interferon gamma and tumor necrosis factor alpha and response to antiviral treatment of chronic hepatitis C, J. Hum. Virol. 4 (2001) 195–199.
[21] H.F. Kawai, S. Kaneko, M. Honda, Y. Shirota, K. Kobayashi, Alpha-fetoprotein-producing hepatoma cell lines share common expression profiles of genes in various categories demonstrated by cDNA microarray analysis, Hepatology 33 (2001) 676–691.
[22] T. Ito, K. Yasui, J. Mukaigawa, A. Katsume, M. Kohara, K. Mitamura, Acquisition of susceptibility to hepatitis C virus replication in HepG2 cells by fusion with primary human hepatocytes: establishment of a quantitative assay for hepatitis C virus infectivity in a cell culture system, Hepatology 34 (2001) 566–572.
[23] A.B. Khodursky, B.J. Peter, N.R. Cozzarelli, D. Botstein, P.O. Brown, C. Yanofsky, DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 97 (2000) 12170–12175.
[24] M. Ellis, N. Davis, A. Coop, M. Liu, L. Schumaker, R.Y. Lee, R. Srikanchana, C.G. Russell, B. Singh, W.R. Miller, V. Stearns, M. Pennanen, T. Tsangaris, A. Gallagher, A. Liu, A. Zwart, D.F. Hayes, M.E. Lippman, Y. Wang, R. Clarke, Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses, Clin. Cancer Res. 8 (2002) 1155–1166.
[25] G. Taguchi, J. Rajesh, The Mahalanobis–Taguchi strategy, ed., Wiley, New York, 2002.
[26] J.C. Chang, E.C. Wooten, A. Tsimelzon, S.G. Hilsenbeck, M.C. Gutierrez, R. Elledge, S. Mohsin, C.K. Osborne, G.C. Chamness, D.C. Allred, P. O'Connell, Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer, Lancet 362 (2003) 362–369.
[27] C. Workman, L.J. Jensen, H. Jarmer, R. Berka, L. Gautier, H.B. Nielser, H.H. Saxild, C. Nielsen, S. Brunak, S. Knudsen, A new non-linear normalization method for reducing variability in DNA microarray experiments, Genome Biol 3 (2002) research0048.
[28] J.H. Kim, D.M. Shin, Y.S. Lee, Effect of local background intensities in the normalization of cDNA microarray data with a skewed expression profiles, Exp. Mol. Med. 34 (2002) 224–232.
[29] Y.H. Yang, S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, T.P. Speed, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res. 30 (2002) e15.
[30] X. Wang, S. Ghosh, S.W. Guo, Quantitative quality control in microarray image processing and data acquisition, Nucleic Acids Res. 29 (2001) E75.
[31] E.M. Glare, M. Divjak, M.J. Bailey, E.H. Walters, β-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels, Thorax 57 (2002) 765–770.
[32] P.D. Lee, R. Sladek, C.M. Greenwood, T.J. Hudson, Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies, Genome Res. 12 (2002) 292–297.
[33] S. Nishiguchi, T. Kuroki, S. Nakatani, H. Morimoto, T. Takeda, S. Nakajima, S. Shiomi, S. Seki, K. Kobayashi, S. Otani, Randomised trial of effects of interferon-alpha on incidence of

hepatocellular carcinoma in chronic active hepatitis C with cirrhosis, Lancet 346 (1995) 1051–1055.
[34] K. Kuwana, T. Ichida, T. Kamimura, S. Ohkoshi, N. Ogata, T. Harada, K. Endoh, H. Asakura, Risk factors and the effect of interferon therapy in the development of hepatocellular carcinoma: a multivariate analysis in 343 patients, J. Gastroenterol. Hepatol. 12 (1997) 149–155.
[35] M.R. Brunetto, F. Oliveri, M. Koehler, F. Zahm, F. Bonino, Effect of interferon-alpha on progression of cirrhosis to hepatocellular carcinoma: a retrospective cohort study. International Interferon-alpha Hepatocellular Carcinoma Study group, Lancet 351 (1998) 1535–1539.
[36] K.R. Coombes, H.A. Fritsche Jr., C. Clarke, J.N. Chen, K.A. Baggerly, J.S. Morris, L.C. Xiao, M.C. Hung, H.M. Kuerer, Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization, Clin. Chem. 49 (2003) 1615–1623.
[37] O. Samek, H.H. Telle, D.C. Beddows, Laser-induced breakdown spectroscopy: a tool for real-time, in vitro and in vivo identification of carious teeth, BMC Oral Health 1 (2001) 1.
[38] A. Chilingaryan, N. Gevorgyan, A. Vardanyan, D. Jones, A. Szabo, Multivariate approach for selecting sets of differentially expressed genes, Math. Biosci. 176 (2002) 59–69.

CLINICAL GASTROENTEROLOGY AND HEPATOLOGY 2005;3:1253–1259

Pretreatment Prediction of Interferon-Alfa Efficacy in Chronic Hepatitis C Patients

KAZUHIRO HAYASHIDA,* AKITO DAIBA,† AKITO SAKAI,§ TAKESHI TANAKA,¶ KYOSUKE KAJI,§ NIRO INABA,† SATOSHI ANDO,† NAOKI KAJIYAMA,† HIROSHI TERASAKI,† AKI ABE,† MASANORI OGASAWARA,† MICHINORI KOHARA,‖ MINE HARADA,* TAKESHI OKANOUE,# SATORU ITO,† and SHUICHI KANEKO§

*Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Higashi-ku, Fukuoka, Japan; †JGS Japan Genome Solutions, Inc., Hachioji, Tokyo, Japan; §Kanazawa University Graduate School of Medical Sciences, Kanazawa, Japan; ¶Tokyo Metropolitan Komagome Hospital, Bunkyo-ku, Tokyo, Japan; ‖Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo, Japan; #Kyoto Prefectural University of Medicine, Graduate School of Medical Science, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto, Japan

Background & Aims: Interferon has been used widely to treat patients with chronic hepatitis C infections. Prediction of interferon efficacy before treatment has been performed mainly by using viral information, such as viral load and genotype. This information has allowed the successful prediction of sustained responders (SR) and non-SRs, which includes transient responders (TR) and nonresponders (NR). In the current study we examined whether liver messenger RNA expression profiles also can be used to predict interferon efficacy. Methods: RNA was isolated from 69 liver biopsy samples from patients receiving interferon monotherapy and was analyzed on a complementary DNA microarray. Of these 69 samples, 31 were used to develop an algorithm for predicting interferon efficacy, and 38 were used to validate the precision of the algorithm. We also applied our methodology to the prediction of the efficacy of interferon/ribavirin combination therapy using an additional 56 biopsy samples. Results: Our microarray analysis combined with the algorithm was 94% successful at predicting SR/TR and NR patients. A validation study confirmed that this algorithm can predict interferon efficacy with 95% accuracy and a *P* value of less than .00001. Similarly, we obtained a 93% prediction efficacy and a *P* value of less than .0001 for patients receiving combination therapy. Conclusions: By using only host data from the complementary DNA microarray we are able to successfully predict SR/TR and NR patients for interferon therapy. Therefore, this technique can help determine the appropriate treatment for hepatitis C patients.

Chronic hepatitis C is one of the major causes of chronic liver disease and can lead to cirrhosis and hepatocellular carcinoma. Interferon is the only effective drug for chronic hepatitis C patients, although better efficacy can be attained with modification of the regimen including the amount of interferon, the duration of treatment, and the use of a combination of pegylated-interferon and ribavirin.

Many studies have identified factors that can help predict the efficacy of interferon therapy such as hepatitis C virus (HCV) genotype[1] and viral loads.[2] Methods based on viral information are able to identify sustained responders (SR). However, this method places transient responders (TR) and nonresponders (NR) in the same category. Follow-up data clearly indicate that interferon treatment of patients in the TR group can lead to a reduction in the probability of tumor development compared with the NR group.[3,4] This suggests that the NR patients should be separated out first and that the TR group should be handled separately as an SR-like group. Furthermore, host factors may help the prediction of NR clinical outcome before treatment. Several candidates have been suggested that may be used to predict this effect including body mass index,[5] γ-glutamyltransferase/alanine transaminase levels,[6] the messenger RNA expression levels of the interferon receptor,[7,8] interferon-γ and tumor necrosis factor-α levels,[9] and the Th1/Th2 ratio[10]; however, there is no definitive evidence that any of these is a single dominant factor. Therefore, additional studies must be performed to identify host factors that can predict the efficacy of interferon therapy because complex changes in these host parameters may reflect variations in hepatic gene expression.

Complementary DNA (cDNA) microarrays can provide an enormous amount of data for identifying clusters

Abbreviations used in this paper: cDNA, complementary DNA; HCV, hepatitis C virus; MD, Mahalanobis distance; NR, nonresponder; SR, sustained responder; SSDB, standard space database; TR, transient responder.

© 2005 by the American Gastroenterological Association
1542-3565/05/$30.00
PII: 10.1053/S1542-3565(05)00412-X

of predictive factors. For example, we previously have used custom-made cDNA microarrays to dissect gene expression patterns and to differentiate between patients infected with HCV and hepatitis B virus.[11,12] Other oligo-DNA chip approaches have proven to be very effective for identifying sets of genes expressed in vitro in response to interferon.[13] However, these approaches have not been useful for determining which treatment regimen should be used for each patient. In the current study we developed a cDNA microarray and a data analysis algorithm that can predict whether a patient will be an NR for interferon therapy based only on host messenger RNA expression and without the use of viral data.

Materials and Methods

Patients and Biopsy Samples

From 1993 to 2001, we collected liver biopsy samples from 99 HCV patients undergoing interferon monotherapy at Kyushu University Hospital. In addition, between 1999 and 2002, an additional 4 samples were collected from patients undergoing interferon monotherapy at Kanazawa University Hospital and 5 samples from Kyoto Prefectural University Hospital as part of a validation study. These patients received the standard 6-month protocol for interferon-alfa treatment. Thus, all patients received more than 468 MU of interferon-alfa monotherapy. Finally, between 2002 and 2003, 56 patients at Kanazawa University Hospital and Tokyo Metropolitan Komagome Hospital were treated with a 6-month regimen of interferon-alfa combined with 600–800 mg/day of ribavirin. Informed consent was obtained from all patients in accordance with the Helsinki protocol. Liver samples were obtained from these patients by biopsy procedure with a 14- or 16-gauge needle. The samples were snap-frozen in liquid nitrogen and stored at −80°C until use for RNA extraction. The viral genotype in pretreatment serum samples from these patients was determined as described previously,[14] the viral RNA copy number was tested using the HCV Amplicore kit (Roche Japan, Tokyo, Japan), and the viral serotype was assayed using an F-HCV-Gr enzyme-linked immunosorbent assay kit (Sysmex, Kobe, Japan). The patients were categorized into 3 groups: SR (patients with an absence of serum HCV RNA both during therapy and 6 months after the completion of therapy), NR (patients persistently positive for serum HCV RNA during therapy), and TR (patients negative for serum HCV RNA at the end of interferon treatment but positive after cessation of therapy).

RNA Extraction, Complementary DNA Microarray, Data Collection, and Data Mining

The total RNA extraction procedure from biopsy samples and the low-density cDNA microarray together with a unique artificial reference RNA (Genomessage; JGS, Tokyo, Japan) used in these studies were described in our previous report.[15]

Results

Selection of Liver Biopsy Samples

RNA degradation is one of the main factors causing variability in data from cDNA microarrays. Because some of the biopsy samples used in this study were stored for more than 8 years, we examined the quality of the extracted RNA by microcapillary electrophoresis. Enough RNA (>2 μg) was obtained from only 69 of the 108 samples from patients treated with interferon monotherapy. We randomly divided these 69 samples into 2 groups of 31 and 38 samples for training and validation of the prediction algorithm, respectively. Based on the 28S/18S ratio, the RNA quality of these 69 samples was good. Of the 69 total patients, 47 were men and 22 were women, and the average age was 49 ± 12 years (range, 21–71 y). Table 1 summarizes the values of alanine transaminase, γ-glutamyltransferase, viral load, and genotypes for the 31 samples used for developing the prediction algorithm. Similarly, qualified RNA extracted from all 56 samples that had been obtained from patients receiving combination therapy were divided randomly into 2 groups of 33 and 23 samples. Of these patients, 46 were men and 10 were women, and the average age was 54 ± 8 years (range, 39–71 y). The characteristics of the group of 33 patients for establishment of SSDB are summarized in Table 2.

Development of the Complementary DNA Microarray

To develop the cDNA microarray for the current study we first performed a serial analysis of gene expression on data from normal and hepatitis B and C patients for approximately 2000 genes. For this serial analysis of gene expression study we analyzed the results from our previous microarray analysis combined with publicly available data.[16,17] During this initial screening phase we tried to choose genes that could distinguish between hepatitis and normal samples. In addition, to focus on genes with meaningful signal levels, we omitted those with a low-frequency expressed tag in serial analysis of gene expression. These approaches are consistent with those of Chang et al[18] who, for statistical calculation, selected only the strong intensity signals from their GeneChip (Affymetrix; Santa Clara, CA) data. Furthermore, we omitted most sequences representing expressed sequence tags in the serial analysis of gene expression data. Finally, we selected genes whose functions have been well established. For example, 26 interferon-related genes were selected for the microarray. We also tried to select genes that have been reported previously to predict interferon efficacy, including interferon-alfa/β receptor,

Table 1. Characteristics of the Core Patients Used for the SSDB and Training

Number	Age	Sex	Genotype	Viral load (*KIU/mL*)	Histology/stage and activity	ALT (*IU/L*)	γ-GT (*IU/L*)	Clinical outcome	Use
1	23	F	1b	4.4	F1A1	90	32	SR	SSDB
2	31	M	2a	23	F1A1	29	11	SR	SSDB
3	34	F	2a	3.5	F1A1	32	199	SR	SSDB
4	40	M	2a	100	F1A2	233	68	SR	SSDB
5	41	M	1b	110	F1A2	182	117	SR	SSDB
6	48	M	2a	2.2	F2A2	189	37	SR	SSDB
7	50	M	2b	3.7	F1A3	267	114	SR	SSDB
8	54	F	2a	2.3	F1A2	41	31	SR	SSDB
9	55	M	2a	2.4	F1A1	301	85	SR	SSDB
10	58	M	1b	50	F1A2	36	59	SR	SSDB
11	60	M	2b	50	F1A1	149	150	SR	SSDB
12	66	M	2a	1.8	F3A2	286	104	SR	SSDB
13	66	M	1b	140	F1A1	88	31	SR	SSDB
14	21	M	1a, 1b	480	F1A1	34	32	NR	Training
15	27	M	1b	520	F1A1	62	39	NR	Training
16	31	F	2a	20	F1A1	63	36	SR	Training
17	35	M	2a	5.9	F1A1	72	34	SR	Training
18	37	F	1b	650	F1A1	219	58	NR	Training
19	37	M	1b	150	F0A1	79	74	TR	Training
20	37	M	2b	250	F1A1	225	29	TR	Training
21	40	M	2a	16	F1A2	211	129	SR	Training
22	42	M	1b	900	F3A2	86	139	TR	Training
23	49	M	1b	540	F1A1	100	30	TR	Training
24	51	F	1b	480	F1A1	80	34	NR	Training
25	52	M	1b	50	F1A2	96	79	SR	Training
26	53	M	1b	520	F4A2	97	90	NR	Training
27	57	M	1b	130	F1A2	61	37	TR	Training
28	57	M	2a	120	F1A2	164	53	SR	Training
29	59	F	1b	230	F3A2	70	38	NR	Training
30	59	M	2b	32	F1A1	162	119	NR	Training
31	62	F	1b	91	F1A2	90	34	NR	Training

ALT, alanine transaminase; γ-GT, γ-glutamyltransferase.

tumor necrosis factor-α.[7–9] In addition, we incorporated clinical information to help select genes; specifically, we included iron transporter–related genes, such as transferrin and the transferrin receptor, because iron depletion has been shown to enhance the efficacy of hepatitis C treatment.[19] Finally, we included some genes (eg, housekeeping genes) as controls for the microarray. Together, 295 genes were selected originally for the low-density cDNA microarray. Furthermore, after we developed the low-density cDNA microarray, interferon-stimulated genes were analyzed systematically by using a different microarray.[20] Another 452 genes that were derived mainly from interferon-, tumor necrosis factor–, and extracellular matrix–related genes were added to a new cDNA microarray that included a total of 747 genes.

Establishment of the Algorithm for Predicting Interferon Efficacy Based on Complementary DNA Microarray Data

As previously described,[15] we used a series of steps to make a reasonable prediction, including establishing a standard space database (SSDB), selecting characteristic parameters to differentiate groups of interest, setting variance–covariance, calculating the variance–covariance matrix, selecting a correlation/gene network, and, finally, calculating the Mahalanobis distance (MD) (the distance from the center of gravity can be determined for a new test sample using the SSDB), leading to a single parameter as a scale from multiple parameters. Thus, the SSDB dataset was selected from the SR patients that had clear clinical outcomes (13 members). This SSDB was trained by expanding it to different datasets, including SR/TR (10 members) and NR (8 members) data, to find genes that are expressed differentially between the 2 groups. The prediction probability of this stage was as follows: SR/TR (10 of 10; 100%), NR (7 of 8; 88%), with a P value of less than .0005.

By using the new prediction algorithm we assessed the accuracy of prediction using the 38 validation samples (31 SR/TR and 7 NR). We calculated the MD and scaled MD from this established dataset for each patient to determine the distance from the established SSDB center of gravity. At this point the calculation does not incor-

Table 2. Characteristics of the Core Patients Used for the SSDB and Training in Combination Therapy

Number	Age	Sex	Naive[a]	Serotype	Viral load (KIU/mL)	Histology/stage and activity	ALT (IU/L)	γ-GT (IU/L)	Clinical outcome	Use
1056	62	M	0	1	595	F4A3	199	87	SR	SSDB
1043	54	M	0	1	77	F2A2	95	80	SR	SSDB
1042	39	M	0	2	850	F1A1	59	89	SR	SSDB
1044	53	M	0	2	300	F3A1	194	147	SR	SSDB
1052	53	M	1	2	440	F1A1	97	80	SR	SSDB
1051	54	M	1	2	600	F3A1	30	22	SR	SSDB
1048	52	M	1	1	580	F1A0	81	37	SR	SSDB
1046	55	M	1	1	510	F1A2	68	49	SR	SSDB
1040	37	M	1	1	360	F1A1	45	90	SR	SSDB
1041	57	M	0	1	250	F4A2	159	93	SR	SSDB
1050	62	M	0	2	690	F1A2	118	96	SR	SSDB
1034	47	F	0	1	820	F1A1	39	43	TR	SSDB
1026	57	M	1	1	550	F3A2	106	14	TR	SSDB
1024	42	M	1	2	570	F2A2	639	83	TR	SSDB
1022	60	M	1	1	610	F1A1	56	209	TR	SSDB
1035	55	F	0	1	360	F1A2	131	42	TR	SSDB
1025	58	F	0	1	850	F1A1	58	35	TR	SSDB
1028	52	M	1	1	650	F1A1	44	17	TR	SSDB
1029	46	M	1	1	850	F1A1	40	30	TR	SSDB
1031	53	M	0	1	690	F2A2	83	52	TR	SSDB
1033	61	M	0	1	850	F2A2	64	46	TR	SSDB
1027	59	M	0	1	630	F3A2	79	59	TR	SSDB
1023	61	M	1	2	300	F3A3	67	61	TR	SSDB
1036	44	M	1	1	850	F1A1	75	54	TR	SSDB
1020	64	F	0	2	850	F1A2	358	76	TR	SSDB
1007	63	M	1	1	850	F3A2	257	132	NR	Training
1009	49	M	1	1	620	F3A1	346	274	NR	Training
1005	58	M	1	1	570	F3A1	87	42	NR	Training
1015	53	M	0	1	850	F1A1	37	65	NR	Training
1014	45	M	1	1	310	F2A2	125	187	NR	Training
1013	57	F	1	1	440	F3A2	57	35	NR	Training
1006	40	M	1	1	>850	F3A1	244	237	NR	Training
1011	57	M	0	1	>850	F2A2	90	48	NR	Training

ALT, alanine transaminase; γ-GT, γ-glutamyltransferase.
[a]0, first treatment; 1, retreatment.

porate any viral information such as genotype or viral load. The prediction probability of the validation stage was as follows: SR/TR (30 of 31; 97%) and NR (6 of 7; 86%). The *P* value, calculated using the χ^2 test, was less than .00001 for prediction accuracy. During the development of the algorithm we found several genes that were expressed differentially between the NR and SR/TR groups. The highlighted 75 genes according to *t*-test values are presented in Table 3. These could be separated into distinct groups such as interferon-, lipid metabolism–, complement-, and oxidoreductase-related genes. Because we used an artificial reference RNA as a control,[15] we were unable to determine whether the up- or down-regulation of these genes was meaningful biologically. *F*-test and *t*-test values for each gene, which represent the comparative expression levels between the SR/TR and NR groups, indicate only that the genes were expressed differentially between the SR/TR and NR groups.

Subsequently, we further analyzed samples from ribavirin combination therapy by using a DNA microarray containing 295 genes. The algorithm developed from the interferon monotherapy samples was ineffective for the samples from the combination therapy. Therefore, we tried to establish a new algorithm based on the data from this 295-gene DNA microarray but failed to obtain accurate prediction even using training samples from the combination therapy group (data not shown). To solve this problem we developed a new algorithm based on a new DNA microarray that included an additional 452 genes. The resulting algorithm allowed prediction of the outcome as follows: training (33 samples) stage: SR/TR (25 of 25; 100%) and NR (7 of 8; 88%); and validation (23 samples) stage: SR/TR (15 of 16; 94%) and NR (5 of 7; 71%), with prediction accuracies of 97% ($P < .0001$) and 87% ($P < .05$), respectively. Thus, the new microarray and algorithm could attain a high accuracy for prediction of treatment outcome using pretreatment liver

Table 3. Genes Differentially Expressed Between SR/TR and NR Groups in Monotherapy

GenBank number	Gene name	*F*-test	*t*-test
U05340	Cell division cycle 20 homolog (*S cerevisiae*)(CDC20)	.020	.169
BC008767	Acyl-coenzyme A oxidase 1, palmitoyl (ACOX1)	.123	.066
AF279437	Interleukin 22	.005	.201
M36807	Glycogen phosphorylase type IV	.007	.250
X03663	Colony-stimulating factor 1 receptor	.254	.096
X02750	Protein C (inactivator of coagulation factors Va and VIIIa)	.020	.172
BC000337	Glucose-6-phosphate dehydrogenase	.146	.066
BC009345	NADH dehydrogenase 1	.082	.233
X00566	Apolipoprotein A-1	.183	.063
BC001188	Transferrin receptor (p90, CD71)	.014	.395
J04026	Thioredoxin	.050	.160
S57235	CD68 antigen	.071	.377
M62403	Insulin-like growth factor binding protein 4	.010	.113
M65128	FK506-binding protein 2 (13 kD)	.050	.023
M29145	Hepatocyte growth factor	.115	.034
M11220	Granulocyte-macrophage colony stimulating factor	.000	.076
M55654	TATA box binding protein	.111	.036
X05360	Cell division cycle 2, G1 to S and G2 to M	.193	.043
M21097	CD19 antigen	.155	.059
J03171	Interferon (α, β, and ω receptor 1	.002	.032
U58196	Interleukin enhancer binding factor 1	.052	.032
Z12020	p53-associated gene; Mdm2, transformed 3T3 cell double minute 2	.014	.121
M93311	Metallothionein-III (growth inhibitory factor [neurotrophic])	.010	.062
X01992	Intereferon γ	.001	.156
Y14736	Immunoglobulin κ (light chain) variable 1D8	.030	.276
M22538	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24 kilodaltons	.087	.172
X00955	Apolipoprotein A-II	.019	.292
U94586	NADH: ubiquinone oxidoreductase MLRQ subunit	.232	.046
X15949	Interferon-regulatory factor-2	.298	.093
M19154	Transforming growth factor β-2	.005	.078
X04688	Interleukin 5 (colony-stimulating factor, eosinophil)	.050	.173
M14505	Cyclin-dependent kinase 4	.116	.080

NOTE. Genes having *F*-test and *t*-test *P*-values <.4 and either value under .1 are listed.
Although *t*-test *P* values were calculated in Welch's method in cases in which the *F*-test *P* value <.05, the Student *t*-test method was applied in cases >.05.

samples. Interestingly, a different set of genes was highlighted in this study, including cyclophilin A and multidrug resistance protein 1 (Table 4).

Discussion

In the current study we developed an algorithm for predicting the efficacy of interferon treatment in hepatitis C patients based only on host microarray data. Once the microarray dataset was normalized, we searched for the most suitable statistical method to differentiate it. We chose a statistical method based on the MD[15] because it allows the maximal flexibility in data dispersion. By using this method we were able to obtain a distinct separation of the NR and the SR/TR groups. In our algorithm we used a gene network system rather than a set of independent differentially expressed genes to generate the categories. Thus, none of the genes listed in Table 3 with *t*-test and *F*-test values alone was able to provide any definitive information for classification, and none of the genes could predict the classification independently. Despite this, we were able to validate our new algorithm using a separate validation group. We were able to predict SR/TR patients accurately in 30 of 31 cases (97%) and NR patients in 6 of 7 cases (86%). The *P* value for these predictions was less than .00001, which is acceptable for prediction purposes and suggests that the dataset from the low-density cDNA microarray can predict NR and non-NR patients with high accuracy without any viral information. Pretreatment classification and identification of NR patients is useful because they have a higher risk than TR or SR patients for developing hepatocellular carcinoma.[3,4] This same approach also allowed prediction of the efficacy of interferon/ribavirin combination therapy with high accuracy. A different gene set was required for the establishment, however, which reflects the different underlying mechanism of the drug action between the 2 treatment regimens.

Genes were selected for the SSDB based on differential expression in our cDNA microarray. Besides interferon-

Table 4. Gene List Highlighted as Differentially Expressed in Combination Therapy

GenBank number	Gene name	*F*-test	*t*-test
X66362	PCTAIRE-3 for serine/threonine protein kinase	.007	.000
U90551	Histone (H2A II; histone 2A-like protein)	.054	.002
Y00285	Insulin-like growth factor II receptor	.027	.002
X03884	CD3 epsilon (T3 epsilon chain [20K] of T-cell receptor)	.045	.007
U12779	MAP kinase-activated protein kinase 2	.143	.008
Z33642	Leukocyte surface protein V7; immunoglobulin superfamily, member 2	.086	.010
U49837	LIM protein (cystein-rich protein 3)	.001	.011
M77349	BIGH3, TGF-β–induced product, TGF-β–induced 68 kilodalton	.141	.013
L16499	Orphan homeobox protein; hematopoietically expressed homeobox	.017	.013
X78817	ρ-GAP hematopoietic protein C1; ρ guanosine triphosphatase–activating protein 4	.070	.016
AF159442	Phospholipid scramblase 3	.000	.017
J04164	Interferon-inducible transmembrane protein 9–27	.199	.018
L41351	Serine protease 8 (prostatin)	.033	.019
U62437	Nicotinic acetylcholine receptor β2 subunit prexursor	.052	.020
X58072	GATA binding protein 3; transacting T-cell–specific transcription factor	.132	.027
X53414	L-alanine: glyoxylate transaminase	.021	.030
Y00052	Cyclophilin A (peptidylprolyl isomerase A; T-cell cyclophilin)	.152	.034
BC004490	Fos	.002	.035
U03397	Tumor necrosis factor–receptor superfamily, member 9	.140	.035
Z47087	Pol V elongation factor-like protein; S-phase kinase-associated protein 1A	.190	.047
M14758	Multiple drug resistance protein 1; P-glycoprotein	.066	.047
U61397	Ubiquitin-homology domain protein PIC1 (sentrin)	.021	.050
U16031	Interleukin-4–induced transcription factor, signal transducer and activator of transcription 6	.194	.050
BC032130	Asialoglycoprotein receptor 1	.032	.057
X05610	Type IV collagen alpha (2)	.046	.059
D23661	Ribosomal protein L37	.035	.066
X69150	Ribosomal protein S18	.013	.068
M15400	Retinoblastoma susceptibility	.040	.104
NM_001012	Ribosomal protein S8	.037	.133
M31627	X-box binding protein-1	.005	.198

NOTE. Both *F*-test and *t*-test values less than .2 are listed.

related genes, the SSDB includes genes related to immune response, stress, metal transport, and lipid metabolism. The inclusion of genes controlled by the interferon signal cascade and related to the immune response is not surprising. In addition, genes associated with lipid metabolism are not unexpected because HCV has a high affinity for lipids.[21,22] Furthermore, lipoprotein receptors were reported as HCV receptor candidates.[23,24] In fact, the involvement of lipid metabolism–related genes is described in depth in a study of HCV clearance in the chimpanzee by GeneChip analysis.[25] Therefore, the lipid metabolism–related genes that we included in our analysis could be targets for future study and therapeutic intervention. Finally, the presence of iron transport–related genes in the SSDB corresponds with the use of blood depletion therapies to reduce liver inflammation in hepatitis patients. It also may be of interest to study how genes in the SSDB, including additional metal-related genes such as metallothioneins, play a role in interferon efficacy.

These findings suggest that the TR patients have an anti-HCV interferon response similar to that of the SR patients. Indeed, it is possible that these TR patients may have become SR patients if interferon treatment was administered for more than 6 months because there is a significant effect of treatment duration in the efficacy of interferon treatment for chronic hepatitis C.[26] This observation is consistent with a study of chimpanzee HCV cases based on oligo-chip data.[25] Furthermore, the fact that we could predict the NR group without any viral information suggests that, in these cases, the host has an unfavorable response to the interferon treatment, which also suggests that, as in the SR group, there is an interaction between the host and the virus. Understanding the host response to interferon in NR patients could provide interesting targets for the development of new treatments for HCV.

In conclusion, we have established a low-density cDNA microarray for predicting interferon efficacy in chronic hepatitis C patients. Based only on host messenger RNA expression profiles from pretreatment biopsy samples, we can categorize patients successfully into SR/TR and NR groups with over 90% accuracy.

References

1. Yoshioka K, Kakumu S, Wakita T, et al. Detection of hepatitis C virus by polymerase chain reaction and response to interferon-

alpha therapy: relationship to genotypes of hepatitis C virus. Hepatology 1992;16:293–299.
2. Lau JY, Davis GL, Kniffen J, et al. Significance of serum hepatitis C virus RNA levels in chronic hepatitis C. Lancet 1993;341: 1501–1504.
3. International Interferon-alpha Hepatocellular Carcinoma Study Group. Effect of interferon-alpha on progression of cirrhosis to hepatocellular carcinoma: a retrospective cohort study. Lancet 1998;351:1535–1539.
4. Kasahara A, Hayashi N, Mochizuki K, et al. Risk factors for hepatocellular carcinoma and its incidence after interferon treatment in patients with chronic hepatitis C. Osaka Liver Disease Study Group. Hepatology 1998;27:1394–1402.
5. Bressler BL, Guindi M, Tomlinson G, et al. High body mass index is an independent risk factor for nonresponse to antiviral treatment in chronic hepatitis C. Hepatology 2003;38:639–644.
6. Mihm S, Monazahian M, Grethe S, et al. Ratio of serum gamma-GT/ALT rather than ISDR variability is predictive for initial virological response to IFN-alpha in chronic HCV infection. J Med Virol 1999;58:227–234.
7. Yatsuhashi H, Yamasaki K, Aritomi T, et al. Quantitative analysis of interferon alpha/beta receptor mRNA in the liver of patients with chronic hepatitis C: correlation with serum hepatitis C virus-RNA levels and response to treatment with interferon. J Gastroenterol Hepatol 1997;12:460–467.
8. Mizukoshi E, Kaneko S, Yanagi M, et al. Expression of interferon alpha/beta receptor in the liver of chronic hepatitis C patients. J Med Virol 1998;56:217–223.
9. Dumoulin FL, Wennrich U, Nischalke HD, et al. Intrahepatic mRNA levels of interferon gamma and tumor necrosis factor alpha and response to antiviral treatment of chronic hepatitis C. J Hum Virol 2001;4:195–199.
10. Sobue S, Nomura T, Ishikawa T, et al. Th1/Th2 cytokine profiles and their relationship to clinical features in patients with chronic hepatitis C virus infection. J Gastroenterol 2001;36:544–551.
11. Honda M, Kaneko S, Kawai H, et al. Differential gene expression between chronic hepatitis B and C hepatic lesion. Gastroenterology 2001;120:955–966.
12. Shirota Y, Kaneko S, Honda M, et al. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. Hepatology 2001;33:832–840.
13. Der SD, Zhou A, Williams BR, et al. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. Proc Natl Acad Sci U S A 1998;95: 15623–15628.
14. Okamoto H, Sugiyama Y, Okada S, et al. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. J Gen Virol 1992;73:673–679.
15. Daiba A, Inaba N, Ando S, et al. A low-density cDNA microarray with a unique reference RNA: pattern recognition analysis for IFN efficacy prediction to HCV as a model. Biochem Biophys Res Commun 2004;315:1088–1096.
16. Yamashita T, Hashimoto S, Kaneko S, et al. Comprehensive gene expression profile of a normal human liver. Biochem Biophys Res Commun 2000;269:110–116.
17. Yamashita T, Kaneko S, Hashimoto S, et al. Serial analysis of gene expression in chronic hepatitis C and hepatocellular carcinoma. Biochem Biophys Res Commun 2001;282:647–654.
18. Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003;362:362–369.
19. Fontana RJ, Israel J, LeClair P, et al. Iron reduction before and during interferon therapy of chronic hepatitis C: results of a multicenter, randomized, controlled trial. Hepatology 2000;31: 730–736.
20. de Veer MJ, Holko M, Frevel M, et al. Functional classification of interferon-stimulated genes identified using microarrays. J Leukoc Biol 2001;69:912–920.
21. Andre P, Komurian-Pradel F, Deforges S, et al. Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. J Virol 2002;76:6919–6928.
22. Kono Y, Hayashida K, Tanaka H, et al. High-density lipoprotein binding rate differs greatly between genotypes 1b and 2a/2b of hepatitis C virus. J Med Virol 2003;70:42–48.
23. Agnello V, Abel G, Elfahal M, et al. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. Proc Natl Acad Sci U S A 1999;96:12766–12771.
24. Scarselli E, Ansuini H, Cerino R, et al. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. EMBO J 2002;21:5017–5025.
25. Su AI, Pezacki JP, Wodicka L, et al. Genomic analysis of the host response to hepatitis C virus infection. Proc Natl Acad Sci U S A 2002;99:15669–15674.
26. Poynard T, Leroy V, Cohard M, et al. Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: effects of dose and duration. Hepatology 1996;24:778–789.

Address requests for reprints to: Shuichi Kaneko, MD, Department of Gastroenterology, Kanazawa University Graduate School of Medical Sciences, 13-1 Takara-machi, Kanazawa 920-8641, Japan. e-mail: skaneko@medf.m.kanazawa-u.ac.jp; fax: (81) 76-234-4250.

Current addresses: Sasebo Kyosai Hospital, 10-17 Shimaji, Sasebo, Nagasaki, Japan (K.H.); Visiting position: Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan (S.I.); and Kawasaki Saiwai Hospital, 39-1 Miyako-machi, Sawai-ku, Kawasaki, Japan (T.T.).

K.H. and A.D. contributed equally to this study.

The authors thank Dr. Hiroshi Yatsuhashi of the National Nagasaki Medical Center, Japan, for initial support and discussions, and Professor Kouji Matsushima of the University of Tokyo for his encouragement and useful discussions.

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Biochemical and Biophysical Research Communications 329 (2005) 370–380

BBRC

www.elsevier.com/locate/ybbrc

Differential gene alteration among hepatoma cell lines demonstrated by cDNA microarray-based comparative genomic hybridization

Kazunori Kawaguchi, Masao Honda, Taro Yamashita, Yukihiro Shirota, Shuichi Kaneko *

Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, Kanazawa, Japan

Received 17 January 2005

Abstract

We assayed chromosomal abnormalities in hepatoma cell lines using the microarray-based comparative genomic hybridization (array-CGH) method and investigated the relationship between genomic copy number alterations and expression profiles in these hepatoma cell lines. We modified a cDNA array-CGH assay to compare genomic DNAs from seven hepatoma cell lines, as well as DNA from two non-hepatoma cell lines and from normal cells. The mRNA expression of each sample was assayed in parallel by cDNA microarray. We identified small amplified or deleted chromosomal regions, as well as alterations in DNA copy number not previously described. We predominantly found alterations of apoptosis-related genes in Hep3B and HepG2, cell adhesion and receptor molecules in HLE, and cytokine-related genes in PLC/PRF/5. About 40% of the genes showing amplification or loss showed altered levels of mRNA ($p < 0.05$). Hierarchical clustering analysis showed that the expression of these genes allows differentiation between α-fetoprotein (AFP)-producing and AFP-negative cell lines. cDNA array-CGH is a sensitive method that can be used to detect alterations in genomic copy number in tumor cells. Differences in DNA copy alterations between AFP-producing and AFP-negative cells may lead to differential gene expression and may be related to the phenotype of these cells.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Microarray-based comparative genomic hybridization; Hepatocellular carcinoma; α-Fetoprotein

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide [1]. Among the factors implicated in the etiology of HCC are infection with hepatitis B virus (HBV) and hepatitis C virus (HCV), dietary aflatoxin, alcohol consumption, and exposure to chemical carcinogens [2–4]. Alterations in genomic DNA copy number are key genetic events in the development and progression of human cancers, including HCC [5–7].

Comparative genomic hybridization (CGH) is a highly specific molecular cytogenetic approach that allows positional identification of gains and losses of DNA sequences throughout the entire genome [8–12]. CGH is based on the use of competitive fluorescence in situ hybridization (FISH) on normal chromosome spreads of differently labeled total genomic DNA from appropriate control and tumor tissue [13]. CGH, however, is unable to detect DNA copy number changes within narrow regions of chromosomes, and alterations of <1 Mb are difficult to detect [14].

Microarray-based CGH (array-CGH) was developed to detect genome-wide alterations in tumor samples by Pinkel et al. using cDNA microarray slides [14–16]. This technique has enabled rapid surveys of known copy number alterations in tumor samples, but resolution can be hampered when only small regions of the genome are amplified [17,18]. Using cDNA clones instead of BAC or PAC clones as probes, however, would make it possible to directly detect amplification and deletion

* Corresponding author. Fax: +81 76 234 4250.
E-mail address: skaneko@medf.m.kanazawa-u.ac.jp (S. Kaneko).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.01.128

of copy numbers of individual genes. Moreover, using cDNAs arrayed on slides would enable parallel measurements of mRNA levels, which may reveal the degree to which variation in gene copy number contributes to variation in gene expression in tumor cells [19,20]. We previously analyzed the alterations of mRNA expression in hepatoma cell lines using an in-house cDNA microarray [21,22], and we showed that α-fetoprotein (AFP)-producing and AFP-negative hepatoma cell lines had different gene expression profiles [23].

In this study, we assayed chromosomal abnormalities in hepatoma cell lines using the cDNA array-CGH method and investigated the relationship between genomic copy number alterations and mRNA expression profiles in these hepatoma cell lines.

Materials and methods

Cell lines and cells. We used three positive control cell lines: the HL-60 promyelocytic leukemia cell line, the IMR-32 neuroblastoma cell line, and the RCF-26 hepatoma cell line, which contain amplified copy numbers of the c-myc, N-myc, and luciferase genes [24–26]. Five AFP-producing hepatoma cell lines: Huh7, Hep3B, HepG2, PLC/PRF/5, and Huh6; two AFP-negative hepatoma cell lines: SK-Hep1 and HLE; two non-hepatoma cell lines: HeLa and KMBC (bile duct cancer cell derived) were used for gene expression profiling and analysis of genomic copy number alterations. As a reference sample in these microarray experiments, we used peripheral blood mononuclear cells (PBMCs) from healthy volunteers and a SV40-T antigen-immortalized normal human hepatocyte cell line (THLE-5b) [27].

cDNA clones and sequence verification. The cDNA microarrays used in this study were made in collaboration with Hitachi Software Engineering (Yokohama, Japan), using an SPBIO2000 robotic arraying machine. The cDNA clones used for making microarrays were selected from UniGene cluster (http://www.ncbi.nlm.nih.gov/UniGene/), and most of them were obtained from the IMAGE Consortium libraries (http://image.llnl.gov/) through Research Genetics (Huntsville, AL). Each microarray contained a total of 1080 cDNA clones, consisting of 930 unique sequence-verified clones and four housekeeping genes. The 930 clones included 141 apoptosis-related genes, 99 cell cycle-related genes, 154 cell–cell interaction-related genes, 198 cytokine and growth factor genes, 123 oncogenes, 81 transcription factor genes, 26 DNA repair-related genes, 93 stress response-related genes, and 87 hematology-related genes. Polymerase chain reaction (PCR) products from the clones were prepared and printed onto glass slides as previously described [28,29].

Copy number analyses by cDNA microarrays. CGH experiments on cDNA microarrays were performed using a modification of a previously described procedure [30]. Nuclei were isolated from cells and dissolved in "hypotonic buffer," consisting of 10 mM Hepes (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM dichlorodiphenyltrichloroethane-1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (DDT), and 0.1 mM phenylmethylsulfonyl fluoride (PMSF). Following digestion in proteinase K in the presence of SDS, genomic DNA was isolated by phenol–chloroform extraction. From each sample, 20 μg of genomic DNA was sonicated for 10 min and digested for 32 h with *Dpn*II (New England Biolabs, Beverly, MA), followed by phenol–chloroform extraction. Each digested hepatoma cell line or positive control DNA was labeled with Cy5-dUTP (Amersham-Pharmacia Biotech, Uppsala, Sweden), and each normal DNA was labeled with Cy3-dUTP (Amersham-Pharmacia) using a Bioprime Labeling Kit (Invitrogen, Carlsbad, CA). Briefly, digested sample DNA was mixed with 15 μg random octamers in a total volume of 41 μl reaction buffer, heated at 100 °C for 5 min, and chilled on ice. Five microliters of 10× dNTPs (i.e., 1.2 mM each of dATP, dGTP, and dCTP, and 0.6 mM dTTP), 3 μl Cy5-dUTP or Cy3-dUTP, and 1 μl Klenow fragment were added to the DNA and then incubated at 37 °C for 2 h. Reactions were stopped by adding 5 μl of 0.5 M EDTA, pH 8.0, and labeled probes were purified on a Microcon 30 column (Millipore, Bedford, MA). Thirty micrograms of human Cot-1 DNA (Invitrogen), 100 μg yeast tRNA (Sigma–Aldrich, St. Louis, MO), and 20 μg poly(A) (Sigma–Aldrich) were added as blocking reagents and then the samples were concentrated to 12 μl. 2.55 μl of 20× SSC and 0.45 μl of 10% SDS were added, and each 15 μl sample was heated at 100 °C for 90 s and then used as a hybridization probe for the DNA-spotted slides. The slides were covered with glass coverslips (22 mm × 22 mm), fixed in a Hybridization Cassette (TeleChem, Sunnyvale, CA), and hybridization was performed at 70 °C for 12 h. The slides were washed in 2× SSC, 0.03% SDS at 70 °C for 5 min, 1× SSC at 70 °C for 5 min, and 0.2× SSC at room temperature for 5 min.

Expression analyses by cDNA microarrays. Total RNA was isolated using a ToTally RNA Kit (Ambion, Austin, TX), and mRNA was isolated from total RNA samples using a MicroPoly(A)Pure Kit (Ambion), according to the manufacturer's instructions. RNA prepared from THLE-5b cells was used as a reference for all cDNA microarray analyses. Fluorescently labeled cDNA probes were made from 2 μg aliquots of mRNA by reverse transcription using oligo(dT) primers and Superscript II reverse transcriptase (Invitrogen). Labeling and hybridization were performed as previously described [23].

Imaging and analyses. Fluorescence intensities generated by the Cy5 and Cy3 probes immobilized on the microarray slides were measured by a laser confocal microscope equipped with a scanning system (ScanArray 5000, GSI Lumonics, Billerica, MA) with appropriate excitation and emission filters. The fluorescence images for Cy5 and Cy3 were scanned separately and stored for image analysis using ImaGene Ver. 3.0 Software (BioDiscovery, Marina del Rey, CA). The signal intensity of each spot was corrected by subtracting background signals in the immediate vicinity, and each Cy5/Cy3 ratio was calculated using global normalization. Clones with a copy number ratio >1.80 were considered to be amplified and those with a ratio <0.55 were considered to be deleted. Over- and underexpression of mRNAs used the same ratios.

Southern hybridization. Genomic DNAs from each cell line and PBMCs were digested overnight with the appropriate restriction enzyme, electrophoresed in 0.8% agarose gels, transferred onto Hybond-N^+ nylon membranes (Amersham-Pharmacia), and hybridized with a ^{32}P-labeled probe previously shown to be amplified or deleted in our array-CGH experiments, according to the standard method [31,32].

Statistical analysis. BRB-Array Tools Ver. 3.1.0 (http://linus.nci.nih.gov/BRB-ArrayTools.html) were used to compare gene amplification in AFP-producing and AFP-negative hepatoma cell lines. Genes showing significant differences were determined by univariate significance test, with a threshold of $p < 0.05$, using a randomized variance model. Hierarchical clustering was performed using centered genes and correlations were determined using average linkage analysis.

Results

Establishment of a sensitive CGH analysis of cDNA arrays

To evaluate the reliability of our CGH method using cDNA microarray slides, we first assayed our positive control cell lines, IMR-32, HL60, and RCF-26, which are known to have amplified copies of c-myc (HL60), N-myc (IMR-32), or luciferase (RCF-26) [24–26]. Using