

Fig. 6. Validation of the CART analysis. Each patient in the validation group was allocated to one of the six subgroups by following the flow-chart form of the decision tree. The rate of (A) sustained virological response (SVR) and (B) null virological response (NVR) in each subgroup was calculated and plotted. The X-axis represents the rate of SVR or NVR in the model building patients and the Y-axis represents those in the validation patients. The rate of SVR and NVR in each subgroup of patients is closely correlated between the model building and the validation patients (correlation coefficient: $r^2 = 0.98-0.99$).

icantly associated with *IL28B* genotype in the present study (Table 4). The serum level of GGT was significantly associated with NVR when examined independently but was no longer significant when analyzed together with the *IL28B* genotype. These observations indicate that some of the factors that we have previously identified may be associated with virological response to therapy through the underlining link to the *IL28B* genotype.

In conclusion, the present study highlighted the impact of the *IL28B* polymorphism and mutation in the ISDR on the pre-treatment prediction of response to PEG-IFN/RBV therapy. A decision model including these host and viral factors has the potential to

support selection of the optimum treatment strategy for individual patients, which may enable personalized treatment.

Conflict of interest

The authors who have taken part in this study declare that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Financial support

This study was supported by a grant-in-aid from the Ministry of Health, Labor and Welfare, Japan, (H19-kannen-013), (H20-kannen-006).

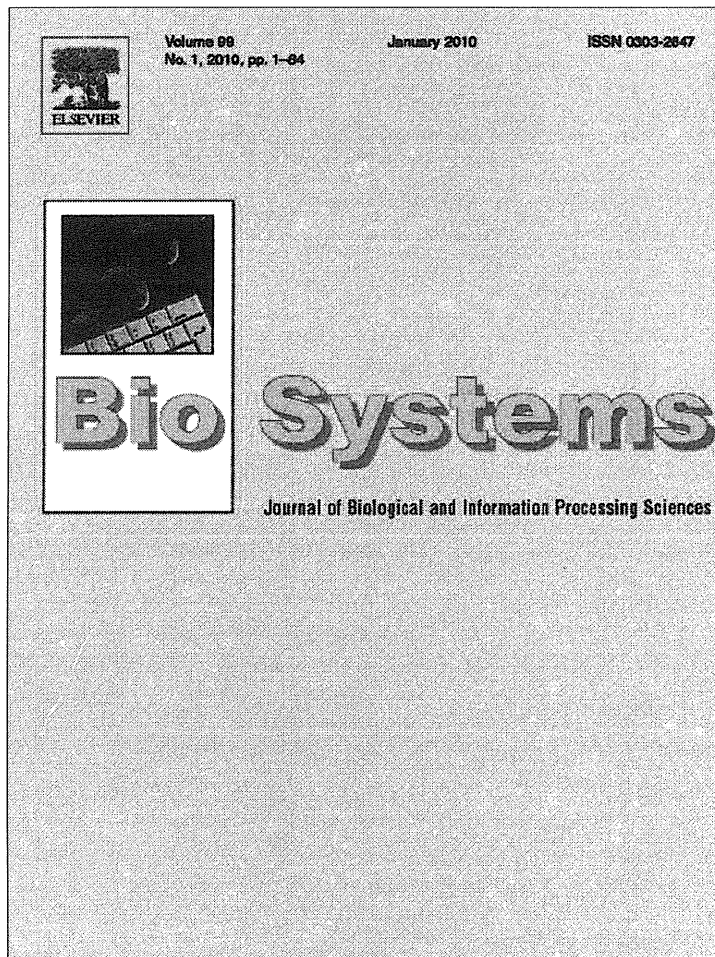
References

- [1] Ray Kim W. Global epidemiology and burden of hepatitis C. *Microbes Infect* 2002;4 (12):1219-1225.
- [2] Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves Jr FL, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002;347 (13):975-982.
- [3] Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001;358 (9286):958-965.
- [4] Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology* 2005;48 (6):372-380.
- [5] Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 2003;38 (3):645-652.
- [6] Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 2009;41:1105-1109.
- [7] Suppiah V, Moldovan M, Ahlenstiel C, Berg T, Weltman M, Abate ML, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 2009;41:1100-1104.
- [8] Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461 (7262):399-401.
- [9] Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* 2010;138 (4):1338-1345.
- [10] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 1995;96 (1):224-230.
- [11] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 1996;334 (2):77-81.
- [12] Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, et al. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* 1997;25 (3):750-753.
- [13] Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, et al. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology* 2008;48 (6):1753-1760.
- [14] Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 2007;46 (3):403-410.

Research Article

- [15] Okanoue T, Itoh Y, Hashimoto H, Yasui K, Minami M, Takehara T, et al. Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study. *J Gastroenterol* 2009;44 (9):952-963.
- [16] Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989;8 (5):539-550.
- [17] LeBlanc M, Crowley J. A review of tree-based prognostic models. *Cancer Treat Res* 1995;75:113-124.
- [18] Garzotto M, Beer TM, Hudson RC, Peters L, Hsieh YC, Barrera E, et al. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 2005;23 (19):4322-4329.
- [19] Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: implications on the future of cancer staging. *Surgery* 2002;132 (4):589-602.
- [20] Leiter U, Buettnner PG, Eigentler TK, Garbe C. Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the German dermatological society. *J Clin Oncol* 2004;22 (18):3660-3667.
- [21] Valera VA, Walter BA, Yokoyama N, Koyama Y, Ijai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol* 2007;14 (1):34-40.
- [22] Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 2005;11 (15):5440-5443.
- [23] Thabane M, Simunovic M, Akhtar-Danesh N, Marshall JK. Development and validation of a risk score for post-infectious irritable bowel syndrome. *Am J Gastroenterol* 2009;104 (9):2267-2274.
- [24] Wu BU, Johannes RS, Sun X, Tabak Y, Conwell DL, Banks PA. The early prediction of mortality in acute pancreatitis: a large population-based study. *Cut* 2008;57 (12):1698-1703.
- [25] Fonarow GC, Adams Jr KF, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *Jama* 2005;293 (5):572-580.
- [26] Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatol Res* 2010;40 (3):251-260.
- [27] Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K. Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. *Anal Biochem* 2007;364 (1):78-85.
- [28] Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goefer T, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 2009;360 (18):1839-1850.
- [29] McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* 2009;360 (18):1827-1838.
- [30] Rossignol JF, Elfert A, El-Gohary Y, Keeffe EB. Improved virologic response in chronic hepatitis C genotype 4 treated with nitazoxanide, peginterferon, and ribavirin. *Gastroenterology* 2009;136 (3):856-862.
- [31] Marcello T, Grakoui A, Barba-Spaeth G, Machlin ES, Kotenko SV, MacDonald MR, et al. Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology* 2006;131 (6):1887-1898.
- [32] Robek MD, Boyd BS, Chisari FV. Lambda interferon inhibits hepatitis B and C virus replication. *J Virol* 2005;79 (6):3851-3854.
- [33] McCarthy JJ, Li JH, Thompson A, Suchindran S, Lao XQ, Patel K, et al. Replicated association between an IL28B Gene Variant and a Sustained Response to Pegylated Interferon and Ribavirin. *Gastroenterology* 2010;138:2307-2314.
- [34] Tanaka Y, Nishida N, Sugiyama M, Tokunaga K, Mizokami M. A-interferons and the single nucleotide polymorphisms: a milestone to tailor-made therapy for chronic hepatitis C. *Hepatol Res* 2010;40:449-460.
- [35] Backus LI, Boothroyd DB, Phillips BR, Mole LA. Predictors of response of US veterans to treatment for the hepatitis C virus. *Hepatology* 2007;46 (1):37-47.
- [36] Mori N, Imamura M, Kawakami Y, Saneto H, Kawaoka T, Takaki S, et al. Randomized trial of high-dose interferon-alpha-2b combined with ribavirin in patients with chronic hepatitis C: correlation between amino acid substitutions in the core/NS5A region and virological response to interferon therapy. *J Med Virol* 2009;81 (4):640-649.
- [37] Hung CH, Lee CM, Lu SN, Lee JF, Wang JH, Tung HD, et al. Mutations in the NS5A and E2-PePHD region of hepatitis C virus type 1b and correlation with the response to combination therapy with interferon and ribavirin. *J Viral Hepat* 2003;10 (2):87-94.
- [38] Yen YH, Hung CH, Hu TH, Chen CH, Wu CM, Wang JH, et al. Mutations in the interferon sensitivity-determining region (nonstructural 5A amino acid 2209-2248) in patients with hepatitis C-1b infection and correlating response to combined therapy of pegylated interferon and ribavirin. *Aliment Pharmacol Ther* 2008;27 (1):72-79.
- [39] Zeuzem S, Lee JH, Roth WK. Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. *Hepatology* 1997;25 (3):740-744.
- [40] Squadrito G, Leone F, Sartori M, Nalpas B, Berthelot P, Raimondo G, et al. Mutations in the nonstructural 5A region of hepatitis C virus and response of chronic hepatitis C to interferon alfa. *Gastroenterology* 1997;113 (2):567-572.
- [41] Sarrazin C, Berg T, Lee JH, Teuber G, Dietrich CF, Roth WK, et al. Improved correlation between multiple mutations within the NS5A region and virological response in European patients chronically infected with hepatitis C virus type 1b undergoing combination therapy. *J Hepatol* 1999;30 (6):1004-1013.
- [42] Murphy MD, Rosen HR, Marousek GI, Chou S. Analysis of sequence configurations of the ISDR, PKR-binding domain, and V3 region as predictors of response to induction interferon-alpha and ribavirin therapy in chronic hepatitis C infection. *Dig Dis Sci* 2002;47 (6):1195-1205.
- [43] Pascu M, Martus P, Hohne M, Wiedenmann B, Hopf U, Schreier E, et al. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut* 2004;53 (9):1345-1351.
- [44] Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 2009;461 (7265):798-801.
- [45] Kurosaki M, Enomoto N, Marumo F, Sato C. Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C. *Virology* 1994;205 (1):161-169.
- [46] Enomoto N, Kurosaki M, Tanaka Y, Marumo F, Sato C. Fluctuation of hepatitis C virus quasispecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. *J Gen Virol* 1994;75 (Pt 6):1361-1369.
- [47] Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, et al. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. *Hepatology* 1904;51 (6):1904-1911.
- [48] Serfaty L, Andreani T, Giral P, Carbonell N, Chazouilleres O, Poupon R. Hepatitis C virus induced hypobetalipoproteinemia: a possible mechanism for steatosis in chronic hepatitis C. *J Hepatol* 2001;34 (3):428-434.

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

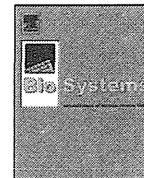
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

BioSystems

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

Reproducibility and usability of chronic virus infection model using agent-based simulation; comparing with a mathematical model

Jun Itakura^{a,*}, Masayuki Kurosaki^a, Yoshie Itakura^a, Sinya Maekawa^b, Yasuhiro Asahina^a, Namiki Izumi^a, Nobuyuki Enomoto^b

^a Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan

^b First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimogatou, Chuo-shi, Yamanashi 409-3898, Japan

ARTICLE INFO

Article history:

Received 30 June 2009

Received in revised form 27 August 2009

Accepted 6 September 2009

Keywords:

Agent-based model

Virus infectious disease

ABSTRACT

We created agent-based models that visually simulate conditions of chronic viral infections using two software. The results from two models were consistent, when they have same parameters during the actual simulation. The simulation results comprise a transient phase and an equilibrium phase, and unlike the mathematical model, virus count transit smoothly to the equilibrium phase without overshooting which correlates with actual biology in vivo of certain viruses. We investigated the effects caused by varying all the parameters included in concept; increasing virus lifespan, uninfected cell lifespan, uninfected cell regeneration rate, virus production count from infected cells, and infection rate had positive effects to the virus count during the equilibrium period, whereas increasing the latent period, the lifespan-shortening ratio for infected cells, and the cell cycle speed had negative effects. Virus count at the start did not influence the equilibrium conditions, but it influenced the infection development rate. The space size had no intrinsic effect on the equilibrium period, but virus count maximized when the virus moving speed was twice the space size. These agent-based simulation models reproducibly provide a visual representation of the disease, and enable a simulation that encompasses parameters those are difficult to account for in a mathematical model.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

All viruses need hosts as a basis for their life. When a virus enters the host body, it invades cells and uses both its own enzymes and those of the host cells to replicate. Host cells infected by viruses launch a self-defense system known as the innate immune system (See and Wark, 2008; Naniche, 2009), which inhibits viral replication and uses the human leukocyte antigen system and cytokines to elicit an immune response. Immune cells that have received signals from host cells activate other immune cells, neutralize viruses in the serum by means of antibodies, and prevent the virus from replicating and proliferating by destroying or curing host cells. Viral infection is a disorder based on the interactions between viruses and cells.

The power relationship between these agents changes along with the progression of the disease. In the very early stages of infection, as the host defense mechanisms are immature, the virus has the ability to overwhelm the host cells, actively replicate, and proliferate. Subsequently, as the capacity of the immune system improves, the speed of viral proliferation drops and the virus count reaches a peak. Infected host cells begin to be disrupted by the immune system or virus particles, and symptoms appear as a result. If the immune system is stronger than the virus, then the viral counts decline, and, in transient viral disorders, the virus is finally eliminated and the host recovers. In chronic viral disorders, however, the power relationship between the virus and host cells reaches equilibrium, and a long-term power balance is maintained with the virus count reaching a plateau.

Mathematical models have been proposed to study the dynamics of such viral disorders, and are regarded as being of value in understanding this phenomenon (Ho et al., 1995; Nowak et al., 1996; Neumann et al., 1998). However, these models are difficult to understand for clinicians, and their applicability is somewhat limited in everyday practice. In clinical research, measurements of viral dynamics in patients for short duration have been made for human

Abbreviations: HIV, human immunodeficiency virus; HBV, hepatitis B virus; HCV, hepatitis C virus.

* Corresponding author. Tel.: +81 422 32 3111; fax: +81 422 32 9551.

E-mail address: jitakura@musashino.jrc.or.jp (J. Itakura).

0303-2647/\$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2009.09.001

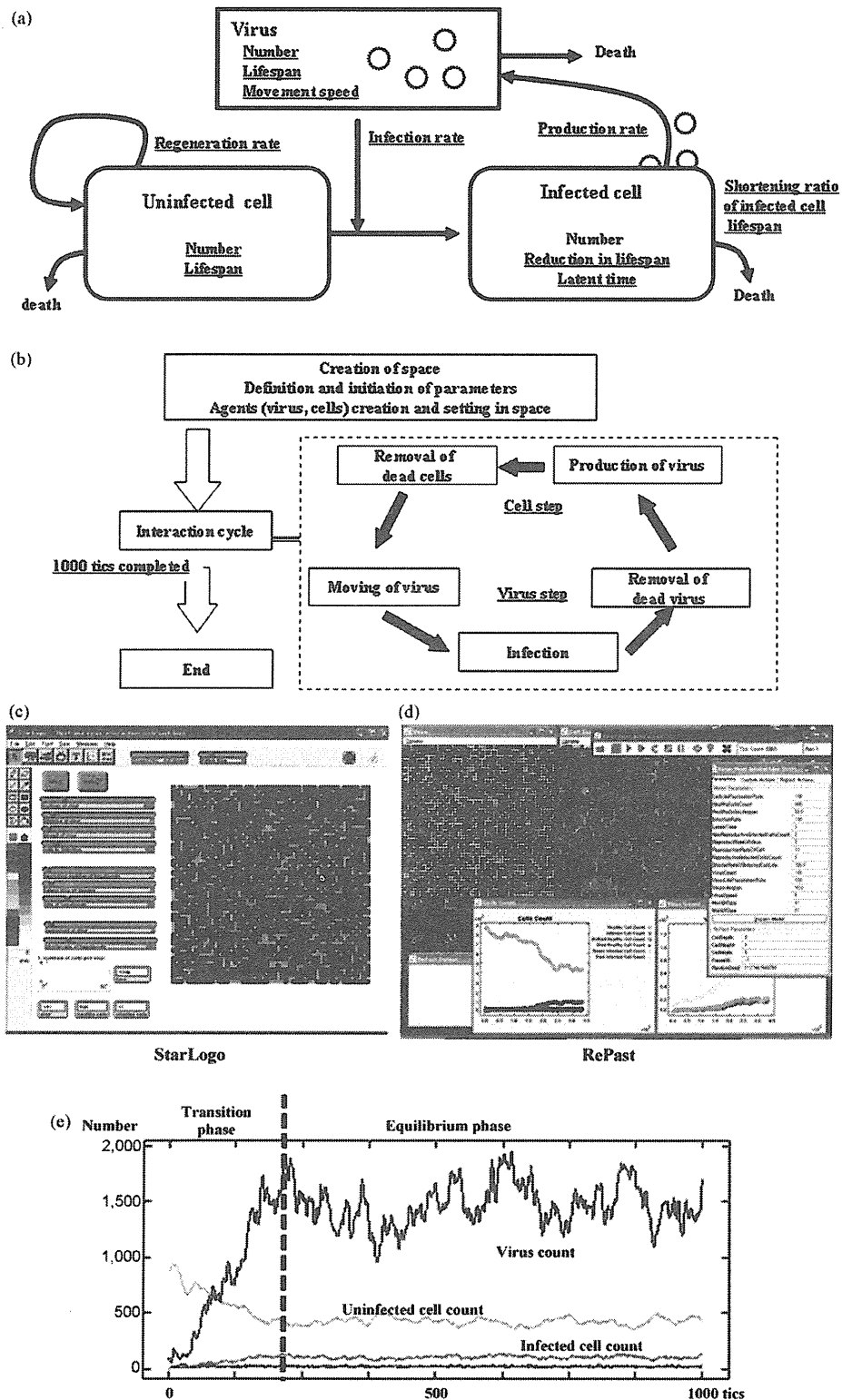


Fig. 1. Simulation design and an example of simulation results. (a) Model concept. Viruses, uninfected cells, and infected cells were treated as agents, and parameters were set for each of these and for interactions between agents (underlined). (b) Flowchart of the program. After preparing the simulation, we entered the interaction cycle, in which virus steps (such as movement) and cell steps were repeated. One cycle was counted as 1 tic, and the simulation concluded after 1000 tics. (c and d) Simulation screen using (c) StarLogo and (d) RePast. Yellow circles are viruses, green squares are uninfected cells, and orange and red indicate infected cells, with orange indicating the latent period. In StarLogo, all the agents are shown on the same screen, but in RePast, viruses and cells are shown in separate windows. (e) Example of a simulation chart in StarLogo. After the start of simulation the virus count and infected cell count increase while the uninfected cell count decreases, with equilibrium state reached after a certain number of tics.

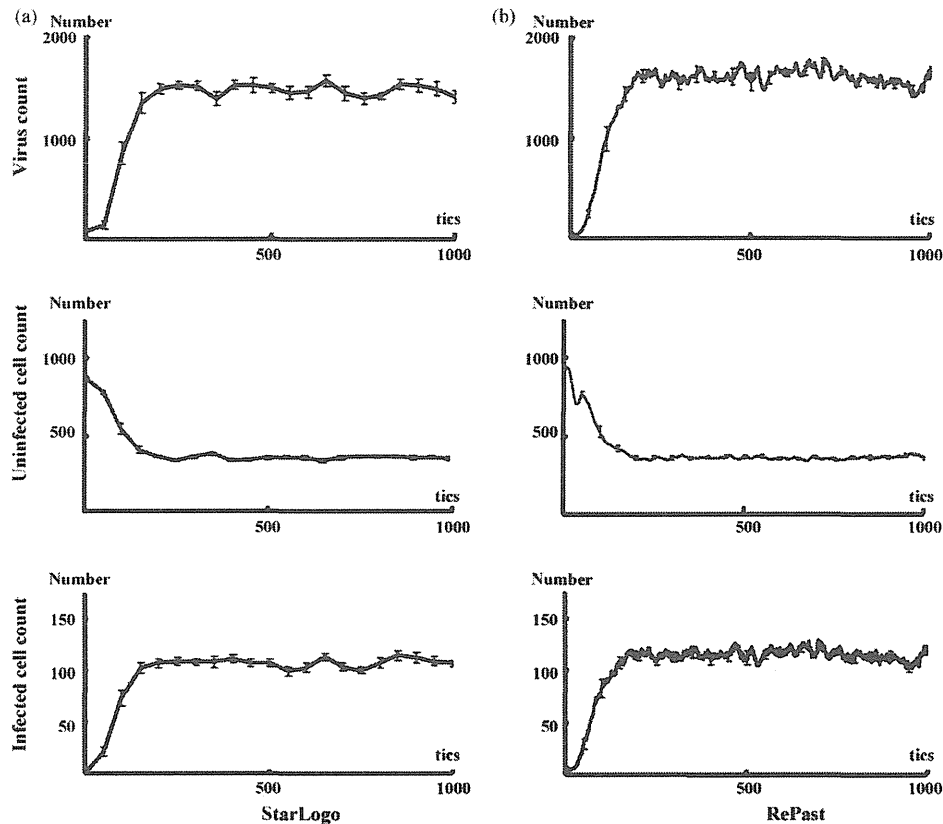


Fig. 2. Comparison of simulation results in (a) StarLogo and (b) RePast. The results were consistent when the parameters were made consistent. (Virus count [average \pm SD]: StarLogo 1458.03 ± 173.1 , RePast 1462.71 ± 178.8 , $p=0.94$. Uninfected cell count: 364.24 ± 30.4 , 368.11 ± 33.4 , $p=0.83$. Infected cell count: 105.73 ± 13.0 , 107.74 ± 13.0 , $p=0.24$. Unpaired Student's *t*-test.) Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; and virus reproduction rate, 5/cells/tic. The following parameter settings were taken from actual measurements: virus lifespan, 4.5 tics; uninfected cell lifespan, 49.8 tics; and infected cell lifespan, 6.7 tics.

immunodeficiency virus (HIV) (Ho et al., 1995), hepatitis B virus (HBV) (Nowak et al., 1996) and hepatitis C virus (HCV) (Neumann et al., 1998), and research is also underway on a range of models based on animal experiments and cell culture systems. As chronic viral disorders persist over long periods of time complete follow-up of viral dynamics is difficult. Furthermore, limitations of items that can be measured, such as the difficulty of measuring whole numbers of host cells, make it extremely difficult to investigate their consistency in mathematical models.

The recent ascend of dynamic-models owes much to advances in computers. Computer performance has improved markedly in recent years, not only in terms of their calculating capacity but also with regard to image displays, and models that offer a visual representation of viral disorders are now being reported (Gilbert and Banks, 2002; Duca et al., 2007; Shapiro et al., 2008; Castiglione et al., 2007). One advantage of such visual models is that by providing a visual representation, they make understanding the disease status easy. Another benefit is that they enable parameters to be identified that are hidden as background noise in mathematical models. However, these models have some problems; it is difficult to prove the reproducibility of the simulation results derived from different languages or libraries, difficult to prove the validity of the model's concepts, and difficult to prove that the simulation results accurately reflect the reality. In this study, we created agent-based computer models that visually simulate the conditions of chronic viral infections using two software. The reproducibility of two agent-based computer models and the differences between agent-based models and the mathematical model were analyzed.

This agent-based model enabled us to investigate how each parameter included in the concept affects the conditions of chronic viral infections.

2. Methods

2.1. Selection of Software

In this study, we used two different types of softwares: StarLogo version 2.0 (<http://education.mit.edu/starlogo/>) supplied by MIT Media Laboratory and Recursive Porous Agent Simulation Toolkit (RePast-3.0, <http://repast.sourceforge.net/>) supplied by the Argonne National Laboratory. StarLogo uses *Logo*, one of the simplest programming languages, and has a fixed graphical user interface. RePast is a library that uses Java, another programming language, which also has a fixed graphical user interface.

Logo is an assembly language, and StarLogo carries out processing sequentially. Java is an object-oriented language, and RePast has a faster processing speed than StarLogo. In addition, StarLogo has a number of stipulations to simplify simulations, such as parameters can only be set up to five decimal places and the simulation space is also fixed as 51×51 square grids. RePast, on the other hand, has fewer such restrictions. Thus, it offers a higher degree of freedom in program settings than StarLogo. Taking simulation space as an example, in spite of the restrictions imposed by the underlying operating system's image display system, any number of grids can be set and a hexagonal grid could also be chosen rather than a square one. However, users must stipulate and set all parameters themselves. This means that they must first declare the shape of the grid and the number of grids they will use to fill the simulation space. Java is also more difficult to learn than Logo, and debugging and correcting the program is also more difficult. Thus, it is difficult to judge whether or not the results agree with the planned simulation.

In effect, these two different types of softwares are polar opposites. It is simple to start a simulation in StarLogo, but producing results takes time and it is difficult to carry out more complex simulations. In RePast it is difficult to compose the program and judge whether or not the planned study has actually been achieved, but the

simulation itself takes only a short time to complete and there are lesser restrictions in the construction of a simulation model.

2.2. Concept for Modeling

We applied the basic virus–host interaction mathematical model to the agent-based simulation system with slight modifications. The mathematical model was used to describe the dynamics of HIV (Ho et al., 1995), HBV (Nowak et al., 1996), and HCV (Neumann et al., 1998) and the only agents involved were host cells and viruses, without the inclusion of immune cells. The effects of the immune system are expressed by varying parameters such as lifespan of host cells and viruses.

Fig. 1a illustrates the study concept. Viruses have the ability to infect healthy host cells (uninfected cells) and the infected cells produce new viruses. Both cells and viruses have definite lifespans, and the lifespan of infected cells is usually shorter than that of uninfected cells. Uninfected cells automatically regenerate within the space, whereas infected cells only arise due to infection of uninfected cells. Viruses also lack the ability to regenerate themselves and are only produced from infected cells.

2.3. Parameter Settings

In the present study, as the StarLogo settings are circumscribed, we limited the simulation space to 51×51 square grids. However, we made an exception here while investigating the effects of size of space on the simulation results. The numbers of viruses, uninfected cells, and infected cells could only be set before the start of the simulation. As described in the later, our simulation ran in cycles, with 1 cycle defined as 1 tic.

In mathematical simulation models, the death rate is required as a parameter. However, in our program we set lifespans for viruses and uninfected cells. These lifespans were not uniform, but were set to have a deviation of about 10%. The lifespan of cells was shortened by infection with ratio decided beforehand.

The infection ratio was meaningful only when an infected cell and a virus coincidentally occupied the same grid, and this was used to calculate the probability of the infection occurring in that situation. The virus production rate was set as the number of viruses produced by an infected cell during 1 tic. Infected cells could be set as a parameter indicating the latent period between the time of virus infection and the time of virus replication.

In order to emulate the tissue repair capacity, we set uninfected cell regeneration rate such that grids without any cells had a specified probability of producing uninfected cells on top of themselves. As a result, the more the cell count declined within a space the more regenerated uninfected cells were produced, whereas the number of regenerated cells declined as cell count increased.

The number of grids through which a virus could move in 1 tic was set as the speed of movement, and the direction of movement was set within a range of 90° toward the top of the simulation space. The program used a circulatory method of movement; when a virus arrived at the top of the space, it was translocated to the bottom, and moved again toward the top. Cells were fixed on the grid.

2.4. Simulation Flowchart

Fig. 1b shows a flowchart of the program. First, the simulation space was produced, after which each parameter was defined and the initial settings were made. Next the agents – viruses and uninfected and infected cells – were produced. The simulation cycle was as follows. Viruses moved to a new grid, and if an uninfected cell was present, this was infected with a probability based on the infection rate. The lifespan of the virus decreased, and viruses that had completed their lifespan and those that had caused an infection were removed from the space. Infected cells produced new viruses, the lifespans of both uninfected and infected cells decreased. Then, cells that had completed their lifespan were eliminated and a new cycle began. The program was set such that the simulation ended after this cycle had repeated 1000 times. This meant that one simulation was complete after 1000 tics.

2.5. Data Collection

The RePast model was programmed such that data for each tic was saved automatically as a text file at the end of the simulation. This text file could be opened by a database software. The StarLogo model was programmed to stop the simulation and collect data after every 50 tics.

2.6. Mathematical Model

In order to compare the results of this agent-based simulation, we used a viral infection mathematical model, which we improved as follows.

$$\frac{dT}{dt} = s[2601 - (T + I)] - dT - bVT \quad (1)$$

$$\frac{dI}{dt} = bVT - dI \quad (2)$$

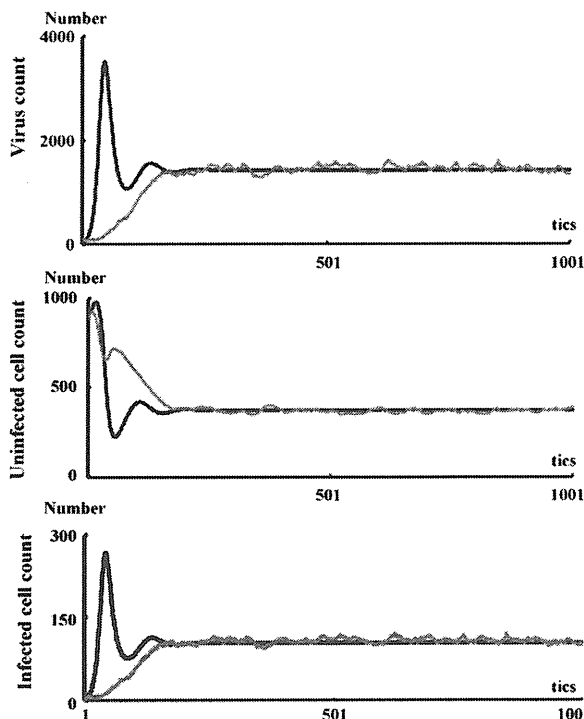


Fig. 3. Comparison of results of agent-based simulation and mathematical simulation. Both sets of results were consistent for the equilibrium phase, but differed in the shift in transition phase. Black line: mathematical model; grey line: results of simulation in RePast. Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; virus reproduction rate, 5/cells/tic; virus lifespan, 10 tics; uninfected cell lifespan, 50 tics; and cell lifespan-shortening ratio as a result of infection, 69%.

$$\frac{dV}{dt} = pI - cV \quad (3)$$

where, T is the uninfected cell count, I is the infected cell count, and V is the virus count. Uninfected cells are supplied to the space with a probability $s[2601 - (T + I)]$, as the number of grids in this agent-based simulation model was 2601 (51×51). The death rate of uninfected cells is d , the death rate of infected cells is δ , and the death rate of viruses is c . The infection rate is indicated by β . Viruses are released from infected cells at a probability p .

2.7. Statistical Analysis

Statistical analyses were performed by statistical tests using the program StatView 5.0 (SAS Institute Inc.). All tests of significance were two-tailed, with p values of <0.05 considered to be significant.

3. Results

3.1. Reproducibility of Chronic Viral Infection Disease Models Using Agent-based Simulation Methods

We constructed the chronic viral infection model with StarLogo library. Fig. 1c shows the simulation screen, and Fig. 1e shows one sample result. Immediately after the start of the simulation, the virus count temporarily dropped in accordance with the onset of an infection. Subsequently, the virus count started to increase with an increase in the infected cells and a decrease in the uninfected cells. After a certain number of tics (around 300 in this example), although the virus count, infected cell count, and uninfected cell count had some fluctuation, an equilibrium state was reached. We use the following descriptive terms in this paper: the transient phase is the period during which virus growth peaks, and the equilibrium phase is the period during which an equilibrium state is

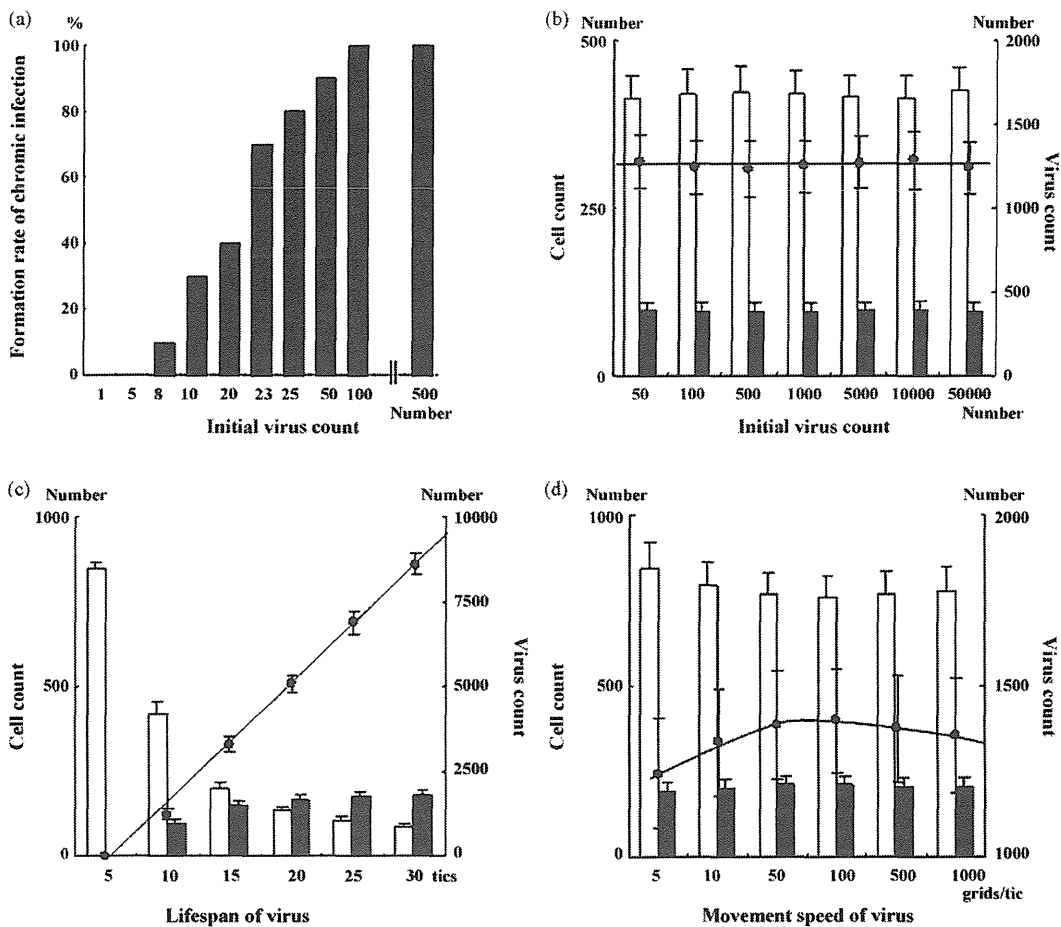


Fig. 4. Effects of changes in viral parameters. (a) The higher the initial virus count, the greater is the increase in the rate of formation of chronic infection, but (b) there was no effect on the conditions in the equilibrium phase. (c) Extending the virus lifespan increased the virus count. (d) Increasing the speed of virus movement to 100 grids/tic increased the virus count, but increasing it to 500 grids/tic had the opposite effect, with a slight declining trend. (a) Black bars: number of infections produced; (b–d) black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

established. When the simulation was performed multiple times, the features described above were maintained, and the average values for virus, infected cell, and uninfected cell counts during the equilibrium state were all consistent.

Fig. 1d shows the simulation screen of the RePast. When we attempted setting all the initial parameters to the same values as those in the StarLogo, the results were not consistent. When we recalculated the parameters from the simulation results, in RePast, the parameters were largely maintained at the levels of the settings, but in StarLogo, the lifespans of both cell types became shorter than the settings while the simulation was in progress. We made the results of both simulations consistent by using the same parameters during the actual simulation (Fig. 2a and b).

3.2. Comparison Between Agent-based Simulation Models and Mathematical Simulation Model

We investigated whether the results of a chronic viral infection disease model produced by RePast would be consistent with the results of a mathematical model. For the mathematical model, we carried out an approximate integration using a four-dimensional Runge–Kutta method to ensure that the uninfected cell count and infected cell count would be in the same class. Parameters were always fixed as constant between simulations. The simulation results were consistent for the equilibrium

phase, but transitions in virus count during the transient phase varied, with a shift to equilibrium state following two overshoots in the mathematical model, but a monotonic increase following a logistic curve in the agent-based model (Fig. 3). In the mathematical model, when the equilibrium condition was calculated with $dT/dt = dI/dt = dV/dt = 0$, the equilibrium-phase virus count, uninfected cell count, and infected cell count were very similar to those of the agent-based model (virus count: mathematical model 371.8/space, agent-based model 371.1 ± 32.4 /space [average \pm SD]; uninfected cell count: mathematical model 1605/space, agent-based model 1454 ± 194 /space; infected cell count: mathematical model 115.9/space, agent-based model 108.3 ± 14.2 /space).

3.3. Usability of the Models; Effect of Changing Parameters

We investigated the changes in the equilibrium phase brought about by changing each parameter. All the investigations below were carried out by using RePast, and we used the average values from ten simulations.

3.4. Viral Parameters

The lower the virus counts at the beginning of the simulation, the lower the probability of a chronic infection (Fig. 4a). However, the initial virus count did not have any effect on the equilibrium

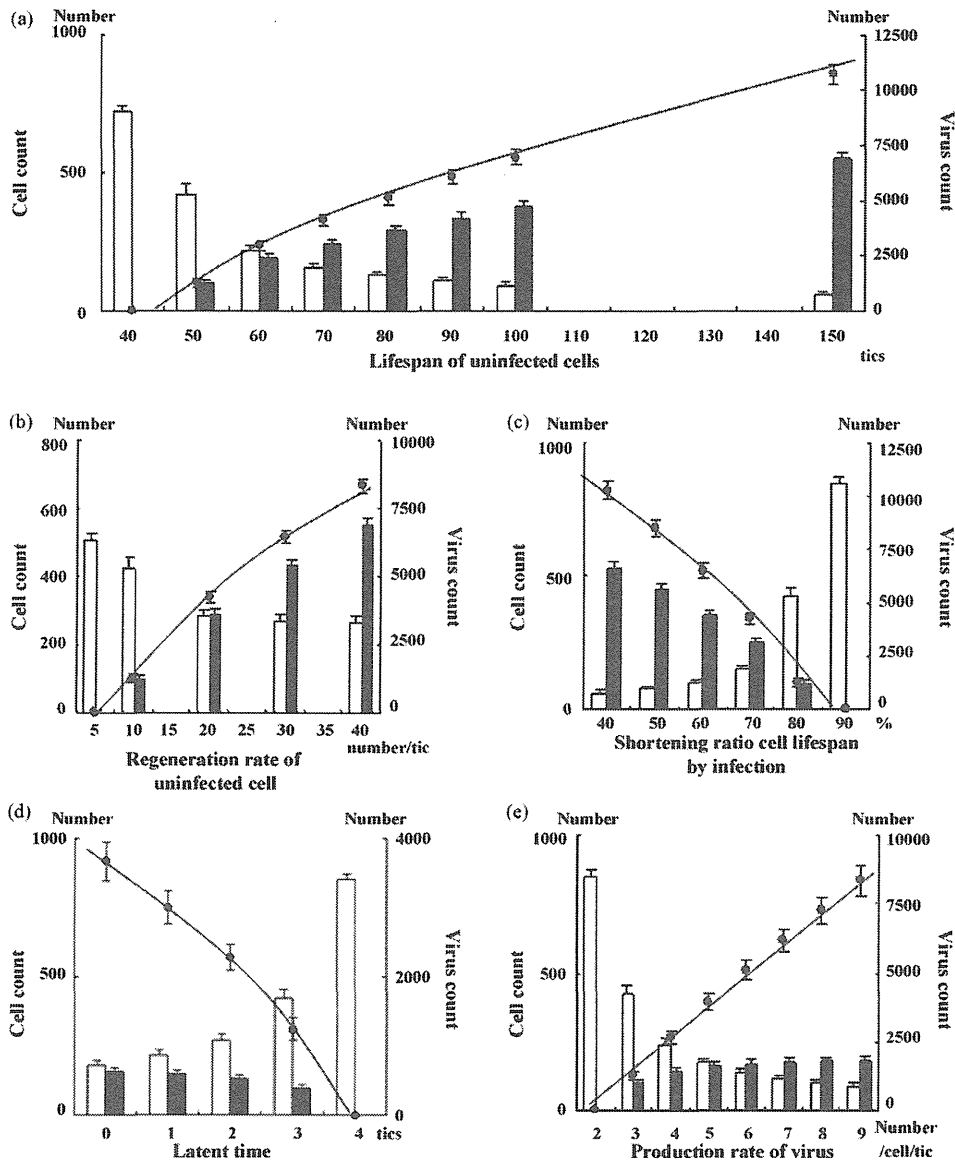


Fig. 5. Effects of changes in cell parameters. (a) Extending the uninfected cell lifespan and (b) increasing the uninfected cell regeneration rate increased the virus count. (c) Raising the lifespan-shortening ratio as a result of infection shortened the lifespan of infected cells, thereby decreasing the virus count. (d) Extending the latent period shortened the period of virus production from infected cells, thereby decreasing the virus count. (e) Increasing the virus production count resulted in a linear increase in equilibrium-phase virus count. Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

phase itself (Fig. 4b). Extending the lifespan of viruses resulted in a linear increase in equilibrium-phase virus count (Fig. 4c). Although the infected cell count increased, the rate of increase gradually declined. Changing the speed of viral movement resulted in the equilibrium-phase virus count to eventually decline after 100 grids/tic was reached, allowing movement over an area twice the size of the simulation space (Fig. 4d).

3.5. Uninfected Cell Parameters

Extending the lifespan of uninfected cells led to an increased virus count during the equilibrium phase (Fig. 5a). Increasing the uninfected cell regeneration rate also contributed to increased equilibrium-phase virus count (Fig. 5b). In both the cases, the

increases in virus count and infected cell count were not linear, but showed a tendency for the rate of increase to decline gradually.

3.6. Infected Cell Parameters

We carried out an investigation of the effects of variation in the lifespan-shortening ratio on the virus count on the assumption that cell lifespan is shortened by infection. When this ratio was increased, the virus count decreased (Fig. 5c). An extended latent period was also related to a decreased virus count (Fig. 5d). However, the virus production from infected cells led to a linear increase in the virus count (Fig. 5e).

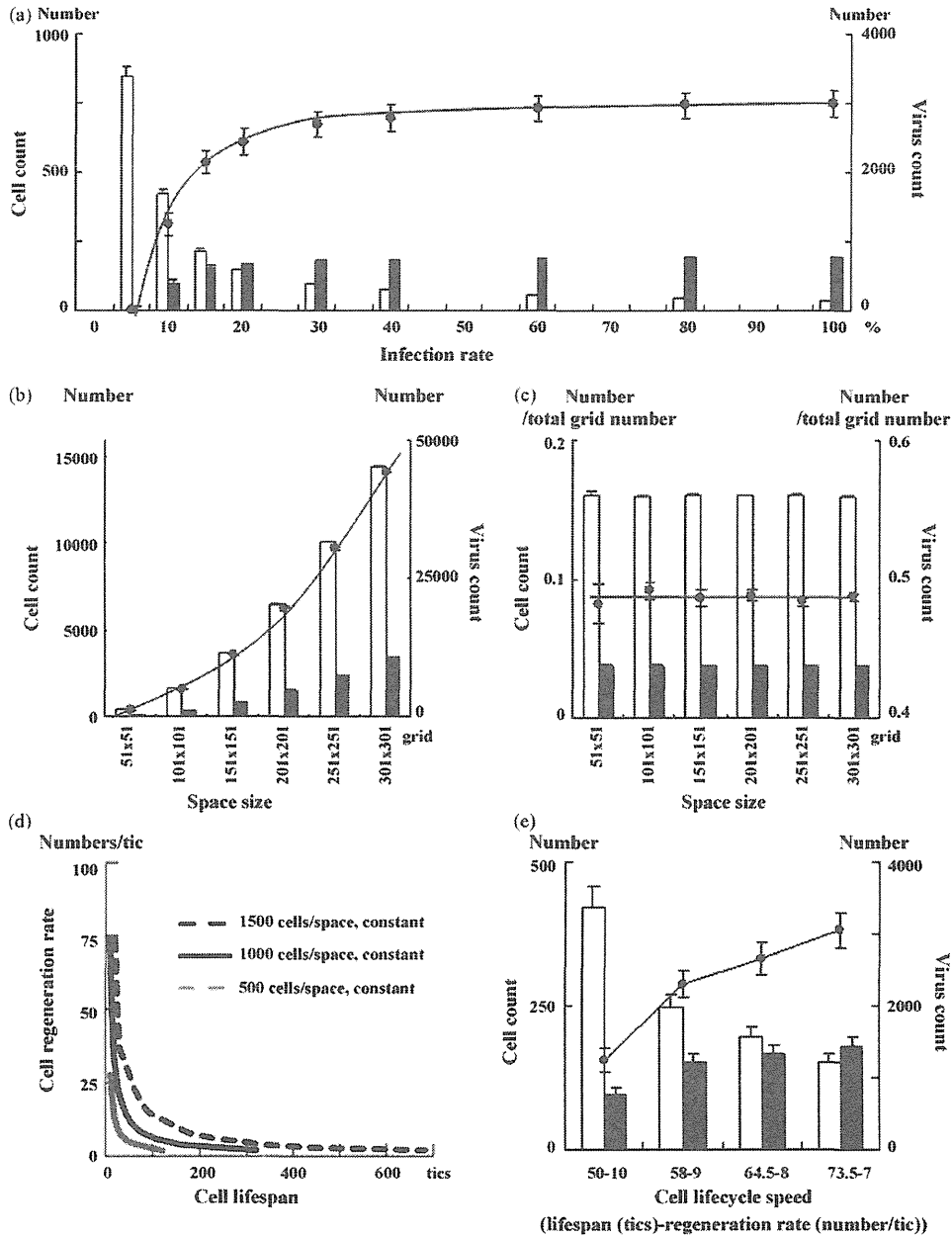


Fig. 6. (a) Increasing the infection rate increased the virus count in equilibrium periods, but the virus count did not change at infection rates of 30% or more. (b) The size of the simulation space increased not only virus count but also the cell count; however, (c) when virus and cell counts were divided by the total number of grids in the space, they were constant for all space sizes. (d) Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible to change only the cell cycle speed without altering the uninfected cell count. (e) When the cell cycle speed was reduced, the virus count increased toward the right of the graph. This may be because the effect of extending the lifespan of cells exceeds that of reducing their regeneration rate. (a–c and e) Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

3.7. Infection Rate and Space Size

Increasing the infection rate caused an increase in the virus count, but the change was minimal at an infection rate of 30% or more. The same results were seen for infected cell count, but a decrease in uninfected cell count resulted in a tendency for the infection rate to decrease by up to 60% (Fig. 6a).

The larger the space, higher the increase in both virus and cell counts (Fig. 6b). This increase was proportional to space size, how-

ever, when virus and cell counts were divided by the total number of grids in the space they were all constant (Fig. 6c).

3.8. Cell Cycle Speeds

Running a simulation with the initial virus count set to zero enables only the equilibrium condition for uninfected cells to be simulated. Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible

to change the cell cycle speed without altering the uninfected cell count (Fig. 6d). We used this technique to investigate how changing the cell cycle speed affected the equilibrium phase. Fig. 6e shows the results. Cell lifespan increases while the cell cycle speed declines. The equilibrium virus count increased in accordance with slower cell cycle speeds.

4. Discussion

In this study, we investigated the models using two agent-based simulation methods to program a simple virus–host chronic infection model. The same model written in two different programming language systems displayed the same results. The transient phase was unlike that seen in a mathematical simulation with no overshoot in virus count, but rather a smooth transition to the equilibrium phase. The virus count at the start of the simulation only had effect on the rate of infection development. Increases in virus lifespan, uninfected cell lifespan, uninfected cell regeneration rate, virus production count from infected cells, and infection rate all led to increased equilibrium-phase virus count. Rises in the infected cell lifespan-shortening ratio, latent period, and cell cycle speed decreased the equilibrium-phase virus count. The size of the space itself had no innate effect on the equilibrium phase, but a speed of movement of the virus that was twice the size of the space produced the maximum virus count.

Reproducibility is the basis for all scientific study, but there are many problems to prove it in computer simulations, such as programming bugs. As agent-based simulation deals with numerous agents individually, it requires vast amounts of calculations. Accumulation of very small change of values leads to large differences of results. In this study, we investigated two programs based on two programming languages to confirm the reproducibility of our simulation results in different programming languages. The results of two simulations were consistent, but in StarLogo, the lifespan parameters had a tendency to be lower than when they were set while simulations were actually in progress. This may be because the number of digits used in calculations was different between the two programs. RePast performs calculations to at least eight decimal places. In StarLogo, the library settings only enable settings to be made up to five decimal places. It is probable that these small differences accumulate during repeated calculations and are reflected in the simulation. Ultimately, we confirmed that the differences in results obtained by using different libraries and programming languages were not innate and by making the parameters consistent during simulation, consistent results were obtained.

Mathematical models using formulae for HIV therapy was published in 1994, the method has since been applied to HBV and HCV (Ho et al., 1995; Nowak et al., 1996; Neumann et al., 1998), and they were thought to be good reflections of the reality. In the mathematical model, viruses and cells are conceived as individuals in the concept itself, but both of them are perceived *en masse* when calculations are performed. However a feature of the agent-based simulation is that it deals with individual viruses and cells as separate agents. By moving each agent individually, it probes the factors influencing overall shifts from the micro viewpoint. When the space is viewed as a whole, it is possible to watch on the screen the collective movement of groups of agents. Recently, models that provide a visual representation of Epstein-Barr virus and HIV infection have been reported, both of which are useful for an instinctive and intuitive understanding (Duca et al., 2007; Shapiro et al., 2008; Castiglione et al., 2007).

In agent-based simulation model, virus count transit smoothly to the equilibrium phase. On the other hand, virus counts overshoot during transient phase in mathematical model. We think this difference is derived from technicality of different model-

ing. The difference in concepts between mathematical models and agent-based models is the space. The mathematical model has no space in concept, but agents move across the space in the agent-based model. In agent-based models, the densities of virus and cells change overtime especially in the transition phase because of the limited space. These changes of the densities of virus and cells lead to the dynamic change of the encounter rate of viruses and cells. The mathematical model does not make such concept of the density; the encounter rate is constant. This may be the reason for the difference between two models in the transition phase. Since no overshoot of virus counts in transient phase had been reported from *in vivo* studies of hepatitis C virus and simian immunodeficiency virus (Dahari et al., 2005; Nowak et al., 1997), agent-based model correlates with actual biology *in vivo* at least for these viruses. The increase of initial virus count at the start of simulation correlates with higher encounter rate of viruses and cells which make the linear increasing of infection forming rate. Mathematical model can only express the infection formation rate as “infected or not”.

The importance of viral passing speed in the agent-based model is also explained by the “space”. Although the virus actually moves through the blood stream in our body and virus could not decide their moving speeds by themselves, there is most appropriate speed for virus to meet the cells on the simulation space by the highest probability. The effect of cell cycle speed should be mentioned by another affection of the space. A fast cell cycle speed means that the lifespan of uninfected cells is short. Then fast cell cycle speed leads to the short lifespan of infected cells. A higher regeneration rate for uninfected cells results in a higher rate of infection among uninfected cells by viruses, but in situations where viruses and cells are dispersed around the space this is ineffective in increasing the infection rate, as the latter depends on the probability that they will encounter one another. As a result, the infected cell count decreases during the equilibrium phase, as does the virus count.

In this study, we confirmed the reproducibility and usability of agent-based models in expressing the interaction between viruses and cells. A feature of this simulation system is that it uses the concept of space as actual space, which means that the existence of the space becomes an additional controlling factor on the simulation results. This is a concept that is absent from mathematical models. The reality is that we have a spatial existence, and an advantage of the agent-based simulation system is the fact that it accounts for the space. Another feature of the simulation system is that it enables the condition to be perceived in visual terms, making it easy to understand. However it may be affected by computer performance and by the limitations of programming languages or the program itself, this system may offer a powerful tool for the future analysis of real virus–host interaction disease.

Conflict of interest

No conflicts of interest exist for all authors.

References

- Castiglione, F., Pappalardo, F., Bernaschi, M., Motta, S., 2007. Optimization of HAART with genetic algorithms and agent-based models of HIV infection. *Bioinformatics* 23, 3350–3355, doi:10.1093/bioinformatics/btm408.
- Dahari, H., Major, M., Zhang, X., Mihalik, K., Rice, C.M., Perelson, A.S., Feinstone, S.M., Neumann, A.U., 2005. Mathematical modeling of primary hepatitis c infection: noncytolytic clearance and early blockage of virion production. *Gastroenterology* 128, 1056–1066, doi:10.1053/j.gastro.2005.01.049.
- Duca, K.A., Shapiro, M., Delgado-Eckert, E., Hadinoto, V., Jarrah, A.S., Laubenbacher, R., Lee, K., Luzuriaga, K., Polys, N.F., Thorley-Lawson, D.A., 2007. A virtual look at Epstein-Barr virus infection: biological interpretations. *PLoS Pathog.* 3, 1388–1400, doi:10.1371/journal.ppat.0030137.
- Gilbert, N., Bankes, S., 2002. Platforms and methods for agent-based modelling. *Proc. Natl. Acad. Sci. U.S.A.* 99 (Suppl. 3), 7197–7198.

- Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123–126, doi:10.1038/373123a0.
- Naniche, D., 2009. Human immunology of measles virus infection. *Curr. Top. Microbiol. Immunol.* 330, 151–171.
- Neumann, A.U., Lam, N.P., Dahari, H., Gretch, D.R., Wiley, T.E., Layden, T.J., Perelson, A.S., 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* 282, 103–107, doi:10.1126/science.282.5386.103.
- Nowak, M.A., Bonhoeffer, S., Hill, A.M., Boehme, R., Thomas, H.C., McDade, H., 1996. Viral dynamics in hepatitis B virus infection. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4398–4402.
- Nowak, M.A., Lloyd, A.L., Vasquez, G.M., Wiltout, T.A., Wahl, L.M., Biscoberger, N., Williams, J., Kinter, A., Fauci, A.S., Hirsch, V.M., Lifson, J.D., 1997. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *J. Virol.* 71, 7518–7525.
- Shapiro, M., Duca, K.A., Lee, K., Delgado-Eckert, E., Hawkins, J., Jarrah, A.S., Laubacher, R., Polys, N.F., Hadinoto, V., Thorley-Lawson, D.A., 2008. A virtual look at Epstein-Barr virus infection: simulation mechanism. *J. Theor. Biol.* 252, 633–648, doi:10.1016/j.jtbi.2008.01.032.
- See, H., Wark, P., 2008. Innate immune response to viral infection of the lungs. *Paediatr. Respir. Rev.* 9, 243–250, doi:10.1016/j.prrv.2008.04.001.

Sequences in the Interferon Sensitivity-Determining Region and Core Region of Hepatitis C Virus Impact Pretreatment Prediction of Response to PEG-Interferon Plus Ribavirin: Data Mining Analysis

Masayuki Kurosaki,¹ Naoya Sakamoto,² Mamabu Iwasaki,³ Minoru Sakamoto,⁴ Yoshiyuki Suzuki,⁵ Naoki Hiramatsu,⁶ Fuminaka Sugauchi,⁷ Akihiro Tamori,⁸ Mina Nakagawa,² and Namiki Izumi, MD^{1*}

¹Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan

²Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan

³Department of Computer and Information Science, Seikei University, Tokyo, Japan

⁴First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan

⁵Department of Hepatology, Toranomon Hospital, Tokyo, Japan

⁶Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka, Japan

⁷Department of Gastroenterology and Metabolism, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan

⁸Department of Hepatology, Osaka City University Medical School, Osaka, Japan

The aim of the present study was to clarify the significance of viral factors for pretreatment prediction of sustained virological response to pegylated-interferon (PEG-IFN) plus ribavirin (RBV) therapy for chronic hepatitis C using data mining analysis. Substitutions in the IFN sensitivity-determining region (ISDR) and at position 70 of the HCV core region (Core70) were determined in 505 patients with genotype 1b chronic hepatitis C treated with PEG-IFN plus RBV. Data mining analysis was used to build a predictive model of sustained virological response in patients selected randomly ($n = 304$). The reproducibility of the model was validated in the remaining 201 patients. Substitutions in ISDR (odds ratio = 9.92, $P < 0.0001$) and Core70 (odds ratio = 1.92, $P = 0.01$) predicted sustained virological response independent of other covariates. The decision-tree model revealed that the rate of sustained virological response was highest (83%) in patients with two or more substitutions in ISDR. The overall rate of sustained virological response was 44% in patients with a low number of substitutions in ISDR (0–1) but was 83% in selected subgroups of younger patients (<60 years), wild-type sequence at Core70, and higher level of low-density lipoprotein cholesterol (LDL-C) (≥ 120 mg/dl). Reproducibility of the model was validated ($r^2 = 0.94$, $P < 0.001$). In conclusion, substitutions in ISDR and Core70 of

HCV are significant predictors of response to PEG-IFN plus RBV therapy. A decision-tree model that includes these viral factors as predictors could identify patients with a high probability of sustained virological response. *J. Med. Virol.* 83:445–452, 2011.

© 2011 Wiley-Liss, Inc.

KEY WORDS: data mining; decision-tree model; ISDR; core region; PEG-interferon

INTRODUCTION

The combination of pegylated-interferon (PEG-IFN) plus ribavirin (RBV) is currently the most effective therapy for chronic hepatitis C, but the rate of sustained virological response after 48 weeks of therapy is about 50% in patients with HCV genotype 1b and a high HCV

Grant sponsor: Ministry of Health, Labor and Welfare, Japan. The authors report no conflicts of interest.

*Correspondence to: Namiki Izumi, MD, Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. E-mail: nizumi@musashino.jrc.or.jp

Accepted 26 October 2010

DOI 10.1002/jmv.22005

Published online in Wiley Online Library (wileyonlinelibrary.com).

RNA titer [Manns et al., 2001; Fried et al., 2002]. The most reliable means to predict sustained virological response is to monitor the viral response during the early weeks of treatment. The early virological response, defined as undetectable HCV RNA at week 12, is associated with a high rate of sustained virological response [Davis et al., 2003; Lee and Ferenci, 2008]. The rapid virological response, defined as undetectable HCV RNA at week 4 of therapy, is even more predictive of sustained virological response than the early virological response [Jensen et al., 2006; Yu et al., 2008; Izumi et al., 2010]. However, there is no established means that predicts the virological response before commencing treatment. Recent reports have revealed that single nucleotide polymorphisms located near the *IL28B* gene show a strong association with the response to PEG-IFN plus RBV therapy [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Kurosaki et al., 2010c]. These findings indicate that the host factor is an important determinant of the treatment response. On the other hand, the present study's authors have reported that a stretch of 40 amino acids in the NS5A region of HCV, designated as the interferon sensitivity-determining region (ISDR), has a close association with the virological response to interferon mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997]. More recently, amino acid substitutions at positions 70 and 91 of the core region have been reported to be associated with response to PEG-IFN plus RBV combination therapy [Akuta et al., 2005, 2007a]. The impact of these HCV substitutions on treatment response is yet to be validated.

Decision-tree analysis is a core component of data mining analysis that can be used to build predictive models [Breiman et al., 1980]. This method has been used to define prognostic factors in various diseases such as prostate cancer [Garzotto et al., 2005], diabetes [Miyaki et al., 2002], melanoma [Averbook et al., 2002; Leiter et al., 2004], colorectal carcinoma [Zlobec et al., 2005; Valera et al., 2007], and liver failure [Baquerizo et al., 2003]. The major advantage of decision-tree analysis over logistic regression analysis is that the results of analysis are easy to understand. The simple allocation of patients into subgroups by following the flowchart form could define the predicted possibility of outcome [LeBlanc and Crowley, 1995].

Decision-tree analysis was used for the prediction of early virological response (undetectable HCV RNA within 12 weeks of therapy) to PEG-IFN and RBV combination therapy in chronic hepatitis C [Kurosaki et al., 2010a], and more recently for the pretreatment prediction of sustained virological response [Kurosaki et al., 2010b]. In the latter model, simple and noninvasive standard tests were used as parameters; specialized tests such as viral mutations and host genetics, or invasive tests such as liver histology, were not included because the aim of that model was for use in general medical practice, especially in some countries or areas where resources are limited. Thus, the impact of viral mutations or liver histology was not considered in that model.

J. Med. Virol. DOI 10.1002/jmv

The present study examined whether including viral substitutions in ISDR and the core region of HCV in the decision-tree model could improve its predictive accuracy over the previous model to identify chronic hepatitis C patients who are likely to respond to PEG-IFN plus RBV therapy.

MATERIALS AND METHODS

Patients

This multicenter retrospective cohort study included 505 chronic hepatitis C patients who were treated with PEG-IFN alpha-2b and RBV at Musashino Red Cross Hospital, Toranomon Hospital, Tokyo Medical and Dental University, Osaka University, Nagoya City University Graduate School of Medical Sciences, Yamanashi University, Osaka City University, and their related hospitals. The inclusion criteria were: (1) genotype 1b, (2) HCV RNA titer higher than 100 kIU/ml by quantitative PCR (Cobas Amplicor HCV Monitor v 2.0, Roche Diagnostic Systems, Pleasanton, CA), (3) no coinfection with hepatitis B virus or human immunodeficiency virus, (4) no other causes of liver disease, (5) patients having undergone liver biopsy prior to IFN treatment, (6) number of substitutions in ISDR having been determined, (7) substitutions in the amino acid positions 70 and 91 of the core region having been determined, and (8) completion of at least 12 weeks of therapy. Patients were treated with PEG-IFN alpha-2b (1.5 μ g/kg) weekly plus RBV. The daily dose of RBV was adjusted by weight: 600 mg for <60 kg, 800 mg for 60–80 kg, and 1,000 mg for >80 kg. For the analysis, patients were assigned randomly to either the model building (304 patients) or validation (201 patients) groups. There were no significant differences in the clinical backgrounds between these two groups (Table I). Informed consent was obtained from each patient. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional review committees of all concerned hospitals.

Laboratory Tests

Hematological tests, blood chemistry, and HCV RNA titer were analyzed before therapy and at least once every month during therapy. Sequences of ISDR and the core region of HCV were determined by direct sequencing after amplification by reverse transcription and polymerase chain reaction as reported previously. At position 70 of the core region (Core70), arginine was defined as the wild type, and glutamine or histidine was defined as the mutant type. At position 91 of the core region, leucine was defined as the wild type and methionine was defined as the mutant type, as described previously [Akuta et al., 2005]. Fibrosis and activity were scored according to the METAVIR scoring system [Bedossa and Poynard, 1996]. Fibrosis was staged on a scale of 0–4: F0 (no fibrosis), F1 (mild fibrosis), F2 (moderate fibrosis), F3 (severe fibrosis), and F4 (cirrhosis). Activity of necroinflammation was graded on a scale of

TABLE I. Comparison of Pretreatment Factors Between Model Building and Validation Patients

	Model (n = 304)	Validation (n = 201)	P-value
Age (years)	55.6 (9.4)	56.0 (12.2)	0.80
Male (%)	53 (%)	55 (%)	0.13
Body mass index (kg/m ²)	23.1 (3.1)	23.1 (4.0)	0.99
Albumin (g/dl)	4.0 (0.3)	4.0 (0.3)	0.47
Creatinine (mg/dl)	0.72 (0.15)	0.72 (0.14)	0.62
AST (IU/L)	63.3 (45.6)	58.9 (46.4)	0.91
ALT (IU/L)	78.7 (58.6)	74.5 (67.5)	0.68
GGT (IU/L)	53.2 (49.1)	57.4 (63.5)	0.43
Total cholesterol (mg/dl)	170.9 (32.6)	169.4 (34.1)	0.33
Triglyceride (mg/dl)	107.0 (44.7)	105.7 (48.0)	0.90
LDL-C (mg/dl)	95.5 (28.0)	96.4 (28.8)	0.34
White blood cell count (/μl)	4,902 (1,489)	4,906 (1,319)	0.86
Hemoglobin (g/dl)	14.1 (1.3)	14.3 (1.4)	0.09
Platelets (10 ⁹ /L)	164 (56)	172 (55)	0.68
HCV RNA (10 ³ IU/ml)	1,859 (1,468)	2,021 (1,393)	0.09
ISDR mutations: >2 (%)	15 (%)	20 (%)	0.11
Core70: mutant (%)	36 (%)	29 (%)	0.22
Core91: mutant (%)	40 (%)	36 (%)	0.20
Fibrosis: F2-4 (%)	49 (%)	48 (%)	0.36
Activity: A2-3 (%)	42 (%)	34 (%)	0.10

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region. Data expressed as mean (SD).

0-3: A0 (no activity), A1 (mild activity), A2 (moderate activity), and A3 (severe activity). Sustained virological response was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems) at week 24 after the completion of therapy.

Statistical Analysis

A database of pretreatment variables included hematological tests (hemoglobin level, white blood cell count, and platelet count), blood chemistry tests (serum levels of creatinine, albumin, aspartate aminotransferase, alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), total cholesterol, triglyceride, and low-density lipoprotein cholesterol (LDL-C)), viral factors (HCV RNA titer, number of substitutions in ISDR, substitutions in the amino acid positions 70 and 91 of the core region), histological findings (stage of fibrosis and grade of activity) and patient characteristics (age, sex, and body mass index). Based on this database, decision-tree analysis was used to define a predictive model for sustained virological response.

Student's *t*-test was used for the univariable comparison of quantitative variables and Fisher's exact test was used for the comparison of qualitative variables. For the multivariable analysis for factors associated with sustained virological response, logistic regression models with backward selection were used to identify independent predictors of sustained virological response. Variables that showed significant association with sustained virological response by univariable analysis were included in the multivariable analysis. IBM-SPSS software v.15.0 (SPSS, Inc., Chicago, IL) was used for these analyses. For the decision-tree analysis [Segal and

Bloch, 1989], the data mining software IBM SPSS Modeler 13 (IBM SPSS, Inc.) was used, as reported previously [Kurosaki et al., 2010a,b]. In brief, the software searched for the optimal split variables to build a decision-tree structure. The entire study population was first evaluated to determine the variables and cut-off points for the most significant division into two subgroups having different probabilities of sustained virological response. Thereafter, analysis was repeated on all subgroups in the same way until either no additional significant variable was detected or the sample size was below 20.

RESULTS

Generation of the Decision-Tree Model

The decision-tree analysis selected five predictive variables to produce six subgroups of patients (Fig. 1). The number of substitutions in ISDR was selected as the best predictor of sustained virological response. The possibility of achieving sustained virological response was 83% for patients with two or more substitutions in ISDR compared with 44% for patients with a single or no substitution. Among patients with a single or no substitution in ISDR, age, with an optimal cut-off of 60 years, was selected as the variable of second split. Patients younger than 60 had the higher probability of sustained virological response (55%) compared with those older than 60 years (31%). Among younger patients, amino acid substitution at Core70 was selected as the third variable of split—wild-type sequence being the predictor of favorable response compared with the mutant type (65% vs. 36%). Among patients with wild-type Core70, the level of serum LDL-C was selected as the fourth variable of split, with an optimal cutoff of

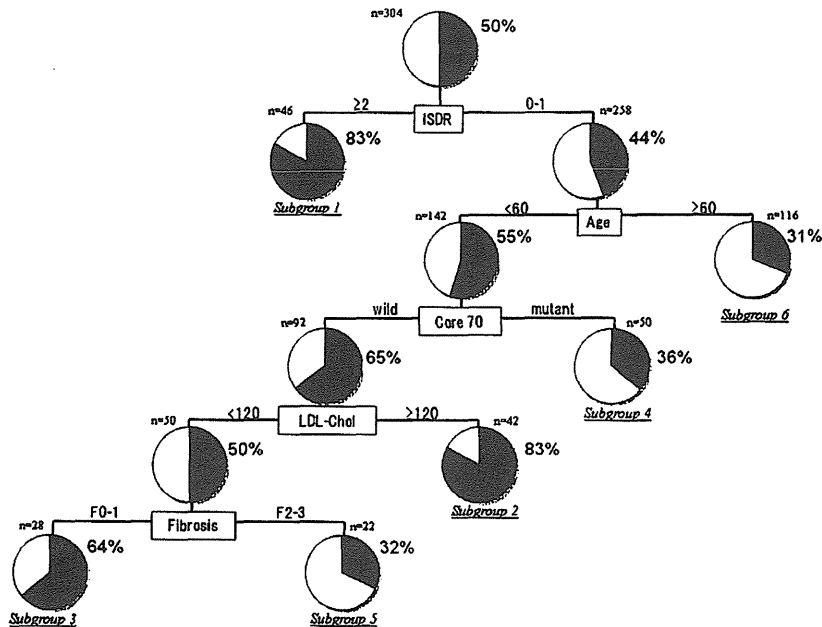


Fig. 1. Decision-tree model. Boxes indicate the factors used for splitting and the cutoff value for the split. Pie charts indicate the rate of sustained virological response for each group of patients after splitting. Terminal subgroups of patients discriminated by the analysis are numbered from 1 to 7. The rate of sustained virological response was >80% in subgroups 1 and 2, 64% in subgroup 3, and 31–36% in subgroups 4, 5, and 6. LDL-C represents low-density lipoprotein cholesterol and Core70 represents amino acid substitution at position 70 of the core region.

120 mg/dl. Patients with higher LDL-C level had the higher probability of sustained virological response (83% vs. 50%). The stage of fibrosis was selected as the final variable of split, with significant fibrosis (F2–4) being the predictor of lower sustained virological response probability (64% vs. 32%).

Among the six subgroups derived by this decision tree, the subgroup of patients with two or more substitutions in ISDR (subgroup 1) or with a single or no substitution in ISDR but younger than 60 years of age, having the wild-type Core70 and high serum level of LDL-C (≥ 120 mg/dl) (subgroup 2) showed the highest probability of sustained virological response (83%).

Validation of the Decision-Tree Model

The decision-tree model was validated using a validation dataset of 201 cases that were not included the model-building dataset. Each patient in the validation set was allocated to subgroups 1–6 using the flowchart form of the decision tree. The rates of sustained virological response were 75% for subgroup 1, 73% for subgroup 2, 65% for subgroup 3, 41% for subgroup 4, 46% for subgroup 5, and 33% for subgroup 6. The rates of sustained virological response for each subgroup of patients were correlated closely between the model building dataset and the validation dataset ($r^2 = 0.94$) (Fig. 2).

J. Med. Virol. DOI 10.1002/jmv

The six subgroups were reconstructed into three groups according to their rate of sustained virological response: the high-probability group consisted of subgroups 1 and 2, the intermediate-probability group consisted of subgroup 3, and the low-probability group consisted of subgroups 4, 5, and 6. The rate of sustained virological response in the high-probability group was high on a consistent basis: 83% for model-building patients and 74% for validation patients. The rate of sustained virological response in the intermediate-probability group was 64% for model building patients and 65% for internal validation patients. The rate of sustained virological response in the low-probability group was low on a consistent basis: 32% for model-building patients and 36% for internal validation patients (Fig. 3). Thirty percent of the patients were classified into the high-probability group and 10% of the patients were classified into intermediate-probability group, which means that about 40% of patients with higher than average probability of achieving sustained virological response were identified.

Effect of Dose Reductions of PEG-IFN and RBV

The possible effect of drug reductions was analyzed in the three groups of patients divided by decision tree (low-, intermediate-, and high-probability groups)

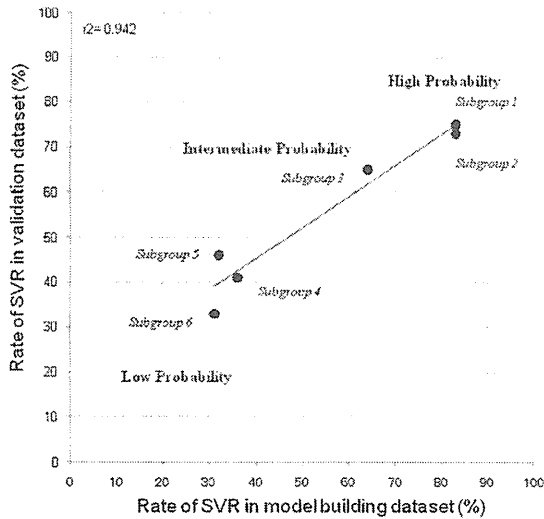


Fig. 2. Validation of the decision-tree analysis: Subgroup-stratified comparison of the rate of sustained virological response. Each patient in the validation set was allocated to subgroups 1–6 by following the flowchart form of the decision tree, and the rates of sustained virological response were then calculated and plotted for each subgroup. The x-axis represents the rate of sustained virological response in the model-building datasets and the y-axis represents the rate of sustained virological response in the validation datasets. The rates of achieving sustained virological response in each subgroup of patients correlated closely between the model-building dataset and the validation dataset (correlation coefficient: $r^2 = 0.94$).

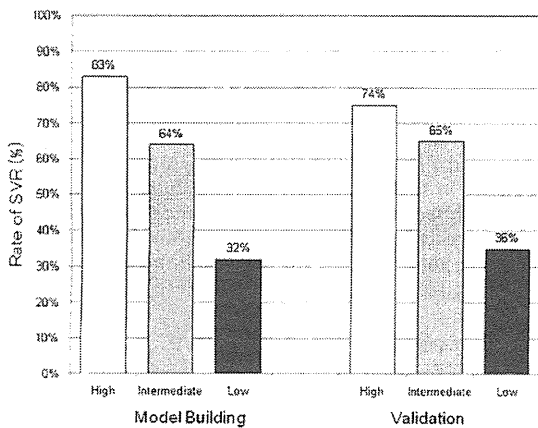


Fig. 3. Comparison of sustained virological response rates between groups divided by the decision tree. The rate of sustained virological response was compared between three groups of patients as divided by the decision-tree analysis. Black, gray, and white boxes indicate the low-probability group (subgroup 4, 5, and 6), intermediate-probability group (subgroup 3), and high-probability group (subgroup 1 and 2), respectively. The rate of sustained virological response showed significant difference between the three groups.

(Fig. 4). Patients were stratified according to the cumulative drug exposure with PEG-IFN and RBV: the good adherence group consisted of patients who took $\geq 80\%$ planned doses of both PEG-IFN and RBV; the poor adherence group consisted of patients who took $< 80\%$ of planned doses of both PEG-IFN and RBV. Even after adjustment for drug adherence, the three groups of patients divided by decision-tree analysis still had low, intermediate, and high probability of achieving sustained virological response, respectively, indicating that this model predicts sustained virological response independent of drug exposure.

Multivariable Logistic Regression Analysis

Age, sex, serum levels of creatinine, ALT, GGT, LDL-C, hemoglobin, platelet count, HCV RNA titer, ISDR substitution, substitution at Core70, substitution at Core91, histological stage of fibrosis, and grade of activity were found to be associated with sustained virological response by standard univariable analysis. Multivariable analysis including these factors showed that age, sex, LDL-C levels, GGT levels, platelet count, ISDR substitution, and substitution at Core70 showed independent associations with sustained virological response (Table II). Substitution in ISDR had the highest odds ratio, at 9.92. Fibrosis, which was selected as a significant predictor of response in the decision-tree analysis, was not found to be an independent predictor of response in standard multivariable analysis, indicating that the decision-tree analysis could identify significant predictors that would apply specifically to selected patients.

DISCUSSION

The present study revealed that viral factors such as substitutions in ISDR and Core70 are significant and independent predictors of sustained virological response to PEG-IFN plus RBV in chronic hepatitis C. In a decision-tree model for the pretreatment prediction of sustained virological response, the number of substitutions in ISDR was the best predictor of sustained virological response, followed by younger age, wild-type sequence at Core70, higher level of LDL-C, and absent fibrosis. This decision-tree model could identify patients with high probability of sustained virological response (83%) among difficult-to-treat genotype 1b chronic hepatitis C patients. Using this model, rapid estimates of the response before treatment can be made by allocating patients to specific subgroups with a defined rate of response simply by following the flowchart form. Because more potent therapy, such as a combination of protease inhibitor, PEG-IFN, and RBV, is under clinical trial and may become available in the near future [Hezode et al., 2009; McHutchison et al., 2009], pretreatment prediction of the likelihood of sustained virological response may be useful for both patients and physicians to support clinical decisions whether to start current standard therapy or to wait for emerging new therapies.

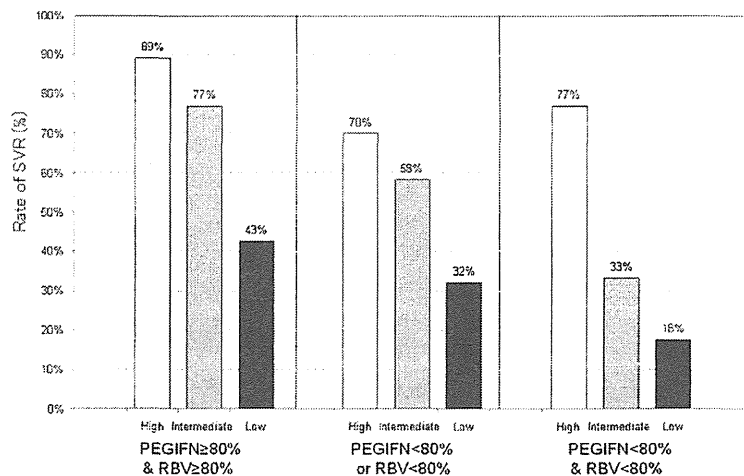


Fig. 4. Comparison of the rate of sustained virological response between the decision-tree groups stratified by drug adherence. The three groups of patients divided by the decision tree (black, gray, and white boxes indicating the low-, intermediate-, and high-probability groups, respectively) were further stratified according to cumulative drug exposure to PEG-IFN and RBV.

Two or more substitutions in ISDR had a strong impact on sustained virological response, because this factor was selected as a top variable in decision-tree analysis and had the highest odds ratio in multivariable analysis. Moreover, even among patients with unfavorable ISDR (0 or 1 mutation), younger patients (<60 years) with the wild-type sequence at Core70 and high level of LDL-C (≥ 120 mg/dl) had a high rate of sustained virological response. The sustained virological response rate of these two subgroups of patients was 83% in the model-building patients and 75% in the validation patients. Thus, patients with high possibility of sustained virological response could be extracted by the combined analysis of ISDR and Core70. These patients may be the best-suited candidates for treatment with the current combination therapy. Conversely, the following patients with 0–1 mutation in ISDR had a low probability of sustained virological response (32–35%): (1) older (>60 years); or (2) younger (<60 years) patients but having mutant-type sequence at Core70; or (3) younger (<60 years) patients having a wild-type sequence at Core70, but having a low level of LDL-C (<120 mg/dl) and advanced fibrosis. These patients may

be advised to wait for a more effective therapy. Decision may be made on a case-by-case basis, taking into account the potential risk of disease progression while waiting.

In a previous decision-tree model using simple and noninvasive standard tests that are available readily worldwide [Kurosaki et al., 2010b], the rate of sustained virological response was at most 65–76% among those in the high-probability group. That model focused on use by general physicians in routine general practice, especially where specialized resources, such as liver biopsy or determination of viral sequences, are not available. In that model, younger age, male sex, higher platelet counts, lower alpha-fetoprotein (AFP) levels, and lower GGT levels were identified as favorable predictive parameters. Higher AFP levels and lower platelet counts that are hallmarks of advanced fibrosis [Shiratori and Omata, 2000; Akuta et al., 2007b] were associated with low probability of sustained virological response in that model. On the other hand, the present analysis aimed to clarify the significance of viral factors for pretreatment prediction of sustained virological response, and to build an advanced model that may be used by specialist physicians engaged in the

TABLE II. Multivariable Logistic Regression Analysis for Factors Associated With SVR

Parameter	Odds	95% CI	P-value	
Age (years)	<60 vs. ≥ 60	2.28	1.31–3.94	0.003
Sex	Male vs. female	3.36	1.87–5.99	<0.0001
GGT (IU/L)	<40 vs. ≥ 40	2.65	1.45–4.85	0.002
LDL-C (mg/dl)	≥ 120 vs. <120	1.79	0.91–3.53	0.094
Platelets (10 ⁹ /L)	≥ 120 vs. <120	2.69	1.22–5.90	0.014
ISDR mutations	≥ 2 vs. 0–1	9.92	3.71–26.54	<0.0001
Core70	Wild vs. mutant	1.92	1.07–3.47	0.030

GGT, gamma-glutamyltransferase; LDL-C, low-density-lipoprotein-cholesterol; ISDR, interferon sensitivity-determining region.

treatment of hepatitis. In the present model, stage of fibrosis was selected as a predictive factor, but at lower level of significance than HCV mutations. The predicted rate of sustained virological response in the high-probability group of the present model is higher than that in the previous model (75–83% vs. 65–76%). These results indicate that substitutions in ISDR and Core70 were important pretreatment predictors of sustained virological response. Determination of these viral factors is not available readily in clinical practice, but is of value for improving the accuracy of pretreatment prediction of sustained virological response.

Substitutions in ISDR and Core70 have been reported previously to be associated with efficacy of IFN therapy. The association between the number of substitutions in ISDR and response to therapy was demonstrated originally in patients treated with IFN mono-therapy [Enomoto et al., 1995, 1996; Kurosaki et al., 1997], but recent studies have reported a positive correlation with PEG-IFN and RBV combination therapy as well [Munoz de Rueda et al., 2008; Shirakawa et al., 2008; Ikeda et al., 2009]. Another important viral factor relevant to treatment response is amino acid substitution in Core70. The sequence of this amino acid was reported originally to be associated with nonresponse to therapy [Akuta et al., 2005], but subsequent studies confirmed the positive correlation of a wild-type Core70 with sustained virological response [Akuta et al., 2009]. The multiple logistic regression analysis showed that ISDR and Core70 were independent factors associated with sustained virological response along with host factors. How these important viral factors and other host factors can be combined to predict response to PEG-IFN plus RBV is an important clinical question. Decision-tree modeling can make the response probability apparent by combining all these factors. Some factors that may be associated with treatment outcome, such as levels of ferritin or homocysteine, were not included. This may be a potential limitation of the present study.

It is of interest that a recent study by Li et al. [2010] has shown that a high serum level of LDL-C is linked to the *IL28B* major allele (CC in rs12979860). In that study, a high serum level of LDL-C was associated with sustained virological response, but it was no longer significant when analyzed together with the *IL28B* genotype in multivariate analysis. Thus, the association between treatment response and LDL cholesterol levels in the present study may reflect the underlining link of LDL cholesterol levels to the *IL28B* genotype. Recent reports indicate that the *IL28B* genotype and HCV substitutions are correlated closely [Akuta et al., 2010; Kurosaki et al., 2010c]. Still, Core70 [Akuta et al., 2010] or ISDR [Kurosaki et al., 2010c] were predictors of response to therapy independent of *IL28B* genotype. Future study is needed to elucidate the possible mechanisms underlying the association between HCV sequences and host genetic factors, and also the role of host and viral factors for the prediction of treatment response.

In conclusion, a data mining analysis emphasized the impact of substitutions in ISDR and Core70 on pretreatment prediction of sustained virological response to PEG-IFN plus RBV therapy. A decision-tree model that includes substitutions in ISDR and Core70 of HCV could identify patients with high probability of sustained virological response, and could thereby improve the predictive accuracy over predictions that are based on standard tests.

REFERENCES

- Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. 2005. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. *Intervirology* 48:372–380.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007a. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: Amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J Hepatol* 46:403–410.
- Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Arase Y, Ikeda K, Kumada H. 2007b. Predictors of viral kinetics to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b. *J Med Virol* 79:1686–1695.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. 2009. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: Amino acid substitutions in HCV core region as predictor of sustained virological response. *J Med Virol* 81:452–458.
- Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Chayama K, Nakamura Y, Kumada H. 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. *Hepatology* 52:421–429.
- Averbuch BJ, Fu P, Rao JS, Mansour EG. 2002. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: Implications on the future of cancer staging. *Surgery* 132:589–602.
- Baquerizo A, Anselmo D, Shackleton C, Chen TW, Cao C, Weaver M, Gornbein J, Geevarghese S, Nissen N, Farmer D, Demetriou A, Busuttill RW. 2003. Phosphorus and an early predictive factor in patients with acute liver failure. *Transplantation* 75:2007–2014.
- Bedossa P, Poynard T. 1996. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 24:289–293.
- Breiman LJH, Friedman RA, Olshen CJ, Stone CM. 1980. Classification and regression trees. CA: Wadsworth.
- Davis GL, Wong JB, McHutchison JG, Manns MP, Harvey J, Albrecht J. 2003. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. *Hepatology* 38:645–652.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Izumi N, Marumo F, Sato C. 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest* 96:224–230.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. 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* 334:77–81.
- Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncalves FL, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J. 2002. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 347:975–982.

- Garzotto M, Beer TM, Hudson RG, Peters L, Hsieh YC, Barrera E, Klein T, Mori M. 2005. Improved detection of prostate cancer using classification and regression tree analysis. *J Clin Oncol* 23:4322-4329.
- Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461:399-401.
- Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, Bronowicki JP, Bourliere M, Gharakhanian S, Bengtsson L, McNair L, George S, Kieffer T, Kwong A, Kauffman RS, Alam J, Pawlowsky JM, Zeuzem S. 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med* 360:1839-1850.
- Ikeda H, Suzuki M, Okuse C, Yamada N, Okamoto M, Kobayashi M, Nagase Y, Takahashi H, Matsunaga K, Matsumoto N, Itoh F, Yotsuyanagi H, Koitabashi Y, Yasuda K, Iino S. 2009. Short-term prolongation of pegylated interferon and ribavirin therapy for genotype 1b chronic hepatitis C patients with early viral response. *Hepatol Res* 39:753-759.
- Izumi N, Nishiguchi S, Hino K, Suzuki F, Kumada Y, Itoh Y, Asahina Y, Tamori A, Hiramatsu N, Hayashi N, Kudo M. 2010. Management of hepatitis C; Report of the consensus meeting at the 45th annual meeting of the Japan society of hepatology (2009). *Hepatol Res* 40:347-368.
- Jensen DM, Morgan TR, Marcellin P, Pockros PJ, Reddy KR, Hadziyannis SJ, Ferenci P, Acrrill AM, Willems B. 2006. Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. *Hepatology* 43:954-960.
- Kurosaki M, Enomoto N, Murakami T, Sakuma I, Asahina Y, Yamamoto C, Ikeda H, Tozuka S, Izumi N, Marumo F, Sato C, Ogura Y. 1997. Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* 25:750-753.
- Kurosaki M, Matsunaga K, Hirayama I, Tanaka T, Sato M, Yasui Y, Tamaki N, Hosokawa T, Ueda K, Tsuchiya K, Nakanishi H, Ikeda H, Itakura J, Takahashi Y, Asahina Y, Higak M, Enomoto N, Izumi N. 2010a. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatol Res* 40:251-260.
- Kurosaki M, Sakamoto N, Iwasaki M, Sakamoto M, Suzuki Y, Hiramatsu N, Sugauchi F, Yatsushashi H, Izumi N. 2010b. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. *J Gastroenterol* DOI: 10.1007/s00535-010-0322-5.
- Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, Honda M, Sugiyama M, Matsuura K, Sugauchi F, Asahina Y, Nakagawa M, Watanabe M, Sakamoto M, Maekawa S, Sakai A, Kaneko S, Ito K, Masaki N, Tokunaga K, Izumi N, Mizokami M. 2010c. Pretreatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. *J Hepatol* DOI: 10.1016/j.jhep.2010.07.037.
- LeBlanc M, Crowley J. 1995. A review of tree-based prognostic models. *Cancer Treat Res* 75:113-124.
- Lee SS, Ferenci P. 2008. Optimizing outcomes in patients with hepatitis C virus genotype 1 or 4. *Antivir Ther* 13:9-16.
- Leiter U, Buettner PG, Eigentler TK, Garbe C. 2004. Prognostic factors of thin cutaneous melanoma: An analysis of the central malignant melanoma registry of the German Dermatological Society. *J Clin Oncol* 22:3660-3667.
- Li JH, Lao XQ, Tillmann HL, Rowell J, Patel K, Thompson A, Suchindran S, Muir AJ, Guyton JR, Gardner SD, McHutchison JG, McCarthy JJ. 2010. Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. *Hepatology* 51:1904-1911.
- Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. *Lancet* 358:958-965.
- McHutchison JG, Everson GT, Jacobson IM, Sulkowski M, Kauffman R, McNair L, Alam J, Muir AJ. 2009. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* 360:1827-1838.
- Miyaki K, Takei I, Watanabe K, Nakashima H, Omae K. 2002. Novel statistical classification model of type 2 diabetes mellitus patients for tailor-made prevention using data mining algorithm. *J Epidemiol* 12:243-248.
- Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, Gila A, Quiles R, Leon J, Ruiz-Extremera A, Salmeron J. 2008. Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82:6644-6653.
- Segal MR, Bloch DA. 1989. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 8:539-550.
- Shirakawa H, Matsumoto A, Joshita S, Komatsu M, Tanaka N, Umemura T, Ichijo T, Yoshizawa K, Kiyosawa K, Tanaka E. 2008. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology* 48:1753-1760.
- Shiratori Y, Omata M. 2000. Predictors of the efficacy of interferon therapy for patients with chronic hepatitis C before and during therapy: How does this modify the treatment course? *J Gastroenterol Hepatol* 15:E141-E151.
- Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, Bassendine M, Spengler U, Dore GJ, Powell E, Riordan S, Sheridan D, Smedile A, Fragomeli V, Muller T, Bahlo M, Stewart GJ, Booth DR, George J. 2009. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 41:1100-1104.
- Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. 2009. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 41:1105-1109.
- Valera VA, Walter BA, Yokoyama N, Koyama Y, Iiai T, Okamoto H, Hatakeyama K. 2007. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. *Ann Surg Oncol* 14:34-40.
- Yu ML, Dai CY, Huang JF, Chiu CF, Yang YH, Hou NJ, Lee LP, Hsieh MY, Lin ZY, Chen SC, Wang LY, Chang WY, Chuang WL. 2008. Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: A randomized trial. *Hepatology* 47:1884-1893.
- Zlobec I, Steele R, Nigam N, Compton CC. 2005. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. *Clin Cancer Res* 11:5440-5443.