In the field of vaccine and drug development, it is highly important to refer to the tertiary structure of a target protein. Therefore, we are now designing a tool to handle the tertiary structure data of virus proteins as a new feature. At present, some structural data of the NS5b (RNA polymerase) region of the HCV protein is available through the Protein Data Bank (PDB). We will make a library of the putative structure of the NS5b region by using PDB entries as templates and homology modeling as an algorithm. The structure data will be linked to the relevant sequence and annotated data, and they will be viewable on the WWW using a 3-D structure viewer.

ACKNOWLEDGMENTS

THIS DATABASE IS supported by a Grant-in Aid for the Publication of Scientific Research Results (No. 168111) and a Grant-in Aid for Scientific Research, Japan Society for the Promotion of Science.

REFERENCES

1 Poynard T, Yuen MF, Ratziu V, Lai CL. Viral hepatitis C. *Lancet* 2003; **362**: 2095–100.
2 Afdhal NH. The natural history of hepatitis C. *Semin Liver Dis* 2004; **24** (Suppl 2): 3–8.
3 Choo QL, Richman KH, Han JH *et al.* Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 1991; **88**: 2451–5.
4 Robertson B, Myers G, Howard C *et al.* Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch Virol* 1998; **143**: 2493–503.
5 Martinot-Peignoux M, Marcellin P, Pouteau M *et al.* Pretreatment serum hepatitis C virus RNA levels and hepatitis C virus genotype are the main and independent prognostic factors of sustained response to interferon alpha therapy in chronic hepatitis C. *Hepatology* 1995; **22**: 1050–6.
6 Kao JH, Chen DS. Global control of hepatitis B virus infection. *Lancet Infect Dis* 2002; **2**: 395–403.
7 Custer B, Sullivan SD, Hazlet TK, Iloeje U, Veenstra DL, Kowdley KV. Global epidemiology of hepatitis B virus. *J Clin Gastroenterol* 2004; **38**: S158–68.
8 Hou J, Liu Z, Gu F, Epidemiology and prevention of hepatitis B virus infection. *Int J Med Sci* 2005; **2**: 50–7.
9 Miyakawa Y, Mizokami M. Classifying Hepatitis B virus genotypes. *Intervirology* 2003; **46**: 329–38.
10 Reyes GR, Purdy MA, Kim JP *et al.* Isolation of cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* 1990; **247**: 1336–9.
11 Tam AW, Smith MM, Guerra ME *et al.* Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* 1991; **185**: 120–31.
12 Okubo K, Sugawara H, Gojobori T, Tateno Y. DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Res* 2006; **34**: D6–9.
13 Kolykhalov AA, Agapov EV, Blight KJ, Mihalik K, Feinstone SM, Rice CM. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science* 1997; **277**: 570–4.
14 Kuiken C, Combet C, Bukh J *et al.* A comprehensive system for consistent numbering of HCV sequences, proteins and epitopes. *Hepatology* 2006; **44**: 1355–61.
15 Pearson WR, Lipman DJ. Improved tools for biological sequence analysis. *Proc Natl Acad Sci USA* 1988; **85**: 2444–8.
16 Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–80.
17 Gojobori T, Ishii K, Nei M. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J Mol Evol* 1982; **18**: 414–23.
18 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–25.
19 Kimura M. *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press, 1983; 65–97.
20 Stuyver L, De Gendt S, Van Geyt C *et al.* A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol* 2000; **81**: 67–74.
21 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**: 783–91.
22 Kalinina O, Norder H, Magnius LO. Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation. *J Gen Virol* 2004; **85**: 1853–7.
23 Kuiken C, Yusin K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. *Bioinformatics* 2005; **21**: 379–84.
24 Combet C, Penin F, Geourjon C, Deleage G, Hcvdb. Hepatitis C virus sequences database. *Appl Bioinformatics* 2004; **3**: 237–40.

Early Dynamics of Hepatitis B Virus in Chimeric Mice Carrying Human Hepatocytes Monoinfected or Coinfected with Genotype G

Masaya Sugiyama,[1] Yasuhito Tanaka,[1] Tomoyuki Sakamoto,[1] Isao Maruyama,[2] Takashi Shimada,[2] Satoru Takahashi,[3] Tomoyuki Shirai,[3] Hideaki Kato,[4] Masataka Nagao,[4] Yuzo Miyakawa,[5] and Masashi Mizokami[1]

Of the 8 genotypes of HBV (genotypes A-H), genotype G is unique in that it has an insertion in the core gene and two stop codons in the precore region preventing the synthesis of hepatitis B e antigen. Most individuals with genotype G are coinfected with other genotypes, typically genotype A. Mice with severe combined immunodeficiency disease carrying human hepatocytes were infected with HBV particles propagated in Huh7 cells in culture. Mice monoinfected with genotype G did not raise detectable HBV DNA in serum, although products of the core gene emerged 4 to 8 weeks after inoculation. When they were superinfected with genotype A at week 10, however, HBV DNA of genotype A developed, which was replaced almost completely by that of genotype G within 10 weeks. Such a rapid takeover was also observed in mice initially infected with genotype A or C and superinfected with genotype G. Similar viral dynamics occurred in mice simultaneously coinfected with genotypes G and A. Takeover was markedly enhanced in mice inoculated with a serum passage containing genotype G with a trace of genotype A. Coinfection of mice with genotypes G and A induced abundant cellular steatosis along with increased fibrosis in the liver, which was not detected in mice monoinfected with genotype A or G. ***Conclusion:*** **Genotype G can monoinfect chimeric mice at very low levels, and its replication increases maredly when coinfected with other genotypes. Coinfection with genotype G could enhance fibrosis under immunocompromised states.** (HEPATOLOGY 2007;45:929-937.)

HBV infects an estimated 350 million people worldwide and causes 1 million deaths annually.[1] Eight genotypes of HBV have been classified by the sequence divergence in the entire genome exceeding 8% and have been assigned the names A through H in order of discovery.[2-6] HBV genotypes have distinct geographic distributions and can influence the severity of liver disease and response to antiviral therapies.[7-10] HBV genotypes are further divided into subgenotypes, such as A1/Aa and A2/Ae, B1/Bj and B2/Ba, and C1/Ce and C2/Cs.[11-13] These genotypes may influence clinical outcomes of HBV infection.[14,15]

HBV genotype G (HBV/G) was first described in 2000 among inhabitants of France and the state of Georgia.[5] It has an insertion of 36 base pairs in the core gene and two stop codons in the precore region.[5,16] Despite the inability in encoding hepatitis B e antigen (HBeAg), carriers of HBV/G possess it in serum.[5] They are usually coinfected with HBV of other genotypes, most frequently HBV/A, which is responsible for serum HBeAg.[17] Coinfection with HBV/C, F, and H has also been reported.[18-20] In spite of heavy dependence on other genotypes for replication, HBV/G outgrows them and eventually takes over the great majority of HBV DNA in the circulation.[16,17]

Recently, HBV/G DNA in low levels was reported in a German donor of plasmapheresis who had transmitted it to 2 recipients in look-back studies.[21] Hence, HBV/G would be able to infect recipients by itself. Furthermore,

Abbreviations: ChiM, chimeric mice; HBcrAg, HBV core-related antigen; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; RFLP, restriction fragment length polymorphism; SCID, severe combined immunodeficiency disease.

From the [1]Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; [2]PhoenixBio Co., Ltd, Higashi-Hiroshima, Japan; the Departments of [3]Experimental Pathology and Tumor Biology and [4]Forensic Medical Science, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; and the [5]Miyakawa Memorial Research Foundation, Tokyo, Japan.

Received August 12, 2006; accepted December 1, 2006

Supported in part by a grant-in-aid from the Ministry of Health, Labor, and Welfare of Japan (H16-kanen-3), Scientific Research from the Ministry of Education (18590741), and Toyoaki Foundation.

Masaya Sugiyama and Yasuhito Tanaka contributed equally to this study.

Address reprint requests to: Masashi Mizokami, M.D., Ph.D., Department of Clinical Molecular Informative Medicine, Nagoya, City University Graduate School of Medical Sciences, Kawasumi, Mizuho, Nagoya 467-8601, Japan. E-mail: mizokami@med.nagoya-cu.ac.jp; fax: (81) 52-842-0021.

Published online in Wiley InterScience (www.interscience.wiley.com).

DOI 10.1002/hep.21584

Potential conflict of interest: Nothing to report.

Supplementary material for this article can be found on the HEPATOLOGY website (http://interscience.wiley.com/jpages/0270-9139/suppmat/index.html).

HBV/G has been detected in 25 of the 104 (24%) French patients coinfected with HIV-1 and HBV and was associated with a high risk of fibrosis at an odds ratio of 12.6.[22]

Mice with severe combined immunodeficiency disease (SCID) transgenic for the urokinase-type plasminogen activator gene under control of albumin promoter (uPA/SCID mice) have received human hepatocyte transplants.[23-25] These mice [hereafter referred to as chimeric (ChiM) mice] have been instrumental in experiments with hepatitis viruses *in vivo*[26,27] and offer a rare opportunity in portraying the early kinetics of HBV replication,[28] without having to resort to the ever-endangered species of chimpanzees. In this study, ChiM mice were monoinfected with HBV/G or coinfected with other genotypes, either simultaneously or in sequence, and followed for circulating HBV/G DNA. It is hoped that the emerging dynamics of HBV DNA will further characterize the dependence of HBV/G on other genotypes and unfold the pathogenicity intrinsic to this parasitic genotype.

Patients and Methods

Patients. Sera were obtained from 4 patients with chronic hepatitis B. One HBV DNA clone of subgenotype A2 and two of HBV/C2 were recovered from 3 Japanese patients in our recent study.[28] Because all HBV DNA clones of HBV/A were classified into subgenotype A2, they will be called HBV/A comprehensively in the present study. The other HBV/A and G clones were obtained from a coinfected Caucasian patient in San Francisco who represented patient 1 in our previous study.[17] All the HBV/A or C clones did not have precore or core promoter mutations affecting the expression of HBeAg. The study design conformed to the 1975 Declaration of Helsinki and was approved by the institutional ethics committees. Written informed consent was obtained from each patient.

Plasmid Constructs of HBV DNA and Sequencing. HBV DNA was extracted from 100 μL of serum using the QIAamp DNA blood kit (Qiagen, GmbH, Hilden, Germany). Four primer sets were designed for amplification of 2 fragments (A and B) covering the entire HBV/G genome. PCR with nested primers was performed with TaKaRa LA Taq polymerase (Takara Biochemicals, Kyoto, Japan) for 35 cycles (30 s at 95°C; 30 s at 60°C; 2 min at 72°C). Primer pairs and protocols for plasmid construction were described in supporting information. As reported previously,[28] these fragments were constructed into the pUC19 vector deprived of promoters (Invitrogen Corp., Carlsbad, CA) by digestion with *Hin*dIII and *Eco*RI, resulting in 1.24-fold the HBV genome—just enough to transcribe oversized pregenome and precore messenger RNA. Cloned HBV DNA sequences were confirmed with Prism BigDye (Applied Biosystems, Foster City, CA) using the ABI 3100 automated sequencer. Additionally, HBV DNA spanning the complete genome were amplified in mouse sera, cloned in the pGEM-T Easy Vector, and then sequenced.

Cell Culture and Transfection. Huh7 cells were transfected with plasmids equivalent to 5 μg of HBV DNA constructs with use of the Fugene 6 transfection reagent (Roche Diagnostics, Indianapolis, IN), and harvested after 3 days in culture. Transfection efficiency was monitored by 0.5 μg of coinfecting reporter plasmids expressing secreted alkaline phosphatase to estimate the latter's enzymatic activity in the culture supernatant.

Determination of HBV Markers. Hepatitis B surface antigen (HBsAg) and HBeAg were determined via chemiluminescent enzyme immunoassay using commercial assay kits (Fujirebio Inc., Tokyo, Japan). HBV core-related antigens (HBcrAg) were measured in serum using the chemiluminescent enzyme immunoassay described previously.[29,30]

Detection and Quantification of Serum HBV DNA. HBV DNA sequences spanning the S gene were determined via real-time detection PCR according to the method of Abe et al.[31] It had a sensitivity of 100 copies/ml (equivalent to 20 IU/ml) on the assay curve obtained with a calibrated World Health Organization standard serum containing HBV of genotype A (kindly provided by Dr. Hiroshi Yoshizawa of Hiroshima University) when 100 μl of the test sample was used. However, in assays for HBV DNA in mouse sera, in which only 10 μl of sample is used, the sensitivity decreased to 1,000 copies/ml (200 IU/ml). For real-time detection PCR specific for HBV/G, 10 μl of DNA sample was amplified in a 25-μl mixture containing 2 × SYBR Green PCR Master Mix (Applied Biosystems) and 2 primers specific for HBV/G: a forward primer (HBVG1620F: ACG TTA CAT GGA AAC CGC CA) and reverse primer (HBVHKR2: AGC CAA AAA GGC CAT ATG GCA) covering the 36–base pair insertion characteristic of this genotype.[5,16] Amplification and detection were performed in the ABI Prism 7700 Sequence Detection System (Applied Biosystems) with an initial activation of UNG at 50°C for 2 minutes, followed by incubation at 95°C for 10 minutes and subsequently, 40 three-step cycles (30 s at 95°C; 30 s at 60°C; 1 min at 72°C) were performed. The standard was prepared on serial dilutions of a known amount of the cloned HBV plasmid of HBV/G. The specificity of 2 primers (HBVG1620F and HBVHKR2) was confirmed in every PCR run via dissociation curve analysis (ABI Prism 7700 dissociation curve software; Applied Biosystems). The

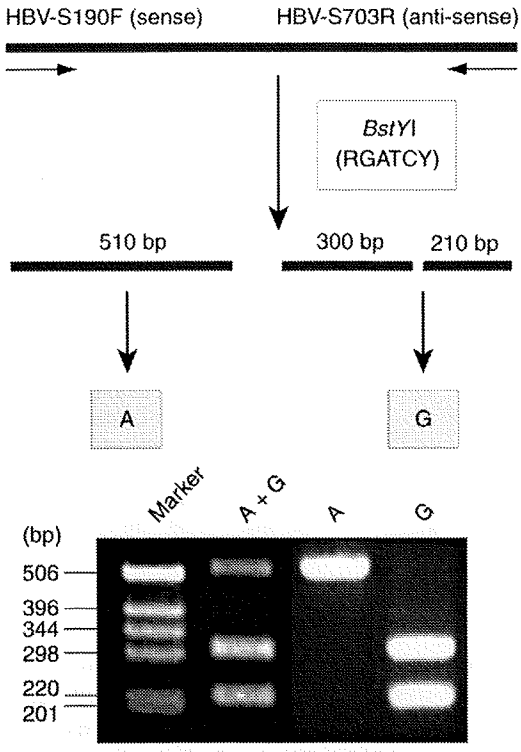

Fig. 1. PCR-RFLP for distinguishing between genotypes A and G. Products of PCR, when digested with *Bst*YI, split into 2 fragments for genotype G (upper panel). Infection with genotype A or G, or coinfection with these genotypes, can be determined by analyzing the patterns of electrophoresis of digests (lower panel).

sensitivity of detecting HBV/G via real-time detection PCR was 1000 copies/ml (200 IU/ml).

PCR Restriction Fragment Length Polymorphism for Distinguishing HBV DNA of Genotype G from Others. A novel method for specific determination of HBV/G DNA in the presence of other genotypes has been developed. It involved single-cycle PCR followed by restriction fragment length polymorphism (RFLP) with an endonuclease having the restriction site specific for HBV/G (Fig. 1). PCR was performed with a forward primer (HBV-S190F: GCT CGT GTT ACA GGC GGG) and reverse primer (HBV-S703R: GAA CCA CTG AAC AAA TGG CAC TAG TA) within the S region. To distinguish HBV/G from other genotypes such as HBV/A and C, a portion (5 μl) of amplification products of 510 base pairs was digested with 5 U *Bst*YI (restriction site: RGATCY) at 60°C for 2 hours. Digests were run on electrophoresis in 3.0% agarose gel, stained with ethidium bromide and examined for their sizes under the ultraviolet light. The results were supported by another method (Supplementary Fig. 1).

Inoculation of Chimeric Mice with the Liver Repopulated for Human Hepatocytes. SCID mice transgenic for the urokinase-type plasminogen activator gene with the liver repopulated for human hepatocytes (chimeric mice) were purchased from Phoenix Bio Co., Ltd. (Hiroshima, Japan). Human serum albumin was measured via ELISA using commercial assay kits (Eiken Chemical Co. Ltd, Tokyo, Japan). They were inoculated with HBV recovered from culture supernatants of Huh7 cells transfected with plasmids constructed with 1.24-fold the HBV genome of genotype HBV/A, C, or G after the method previously reported.[28]

Histopathological Examination. Liver tissues were fixed in buffered formalin, embedded in paraffin, and stained with hematoxylin-eosin or Masson's trichrome. The fibrosis stage was evaluated by an expert pathologist (S. T.) who was blinded to the nature of inocula.

Results

ChiM Mice Monoinfected with HBV/G. Two ChiM mice (ChiM92-3 and ChiM184-4) received an inoculum containing approximately 10^5 copies of HBV/G (G_US1646 strain) and were followed for 12 and 24 weeks, respectively (Fig. 2A,B). HBV DNA remained in undetectable levels ($<10^3$ copies/ml) in them both, but they developed low levels of HBcrAg (1 kU/ml) 4 and 8 weeks after inoculation, respectively. Despite absence of detectable HBV DNA in the circulation, therefore, these mice had contracted infection with HBV/G in very low levels. Intrahepatic cccDNA (covalently closed circular DNA) was detected via PCR specific for it,[32] and HBV/G DNA was detected in hepatocytes via PCR with type-specific primers[33]; they attested to infection with HBV/G in them (data not shown).

Superinfection With HBV/A on Mice Infected with HBV/G. Two chimeric mice with occult infection with HBV/G received 10^5 copies of HBV/A2 of different strains (A2_JPN to ChiM93-4 and A2_USA to ChiM172-3) 10 weeks after initial inoculation with HBV/G (Fig. 3A,B). They both developed HBV DNA in serum in titers $>10^6$ copies/ml at week 17, 7 weeks after superinfection with HBV/A, accompanied by HBcrAg and HBsAg; HBeAg appeared soon thereafter at week 22. HBV DNA and antigens increased, peaked at week 26, and then decreased in exactly the same patterns. HBV/G DNA, which was determined via PCR with type-specific primers, developed 12 weeks after the inoculation at week 22. It increased rapidly, and after the peak, took the same time course as total HBV DNA in serum—it had replaced HBV/A in the two chimeric mice.

Genotypes of HBV in ChiM93-4 were determined at the appearance of HBV DNA (week 17), at peak (week 26), and at the end of observation (week 38) via PCR-

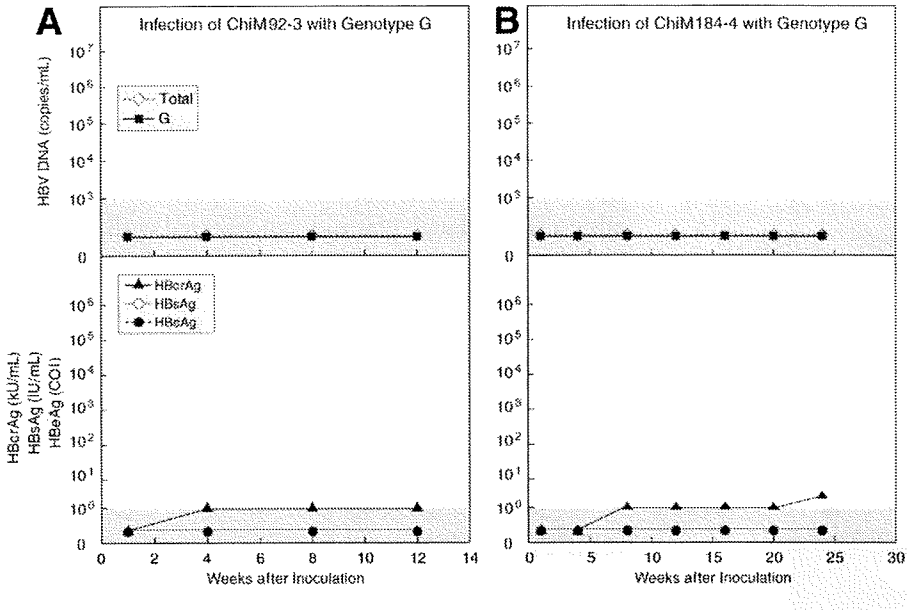

Fig. 2. (A) ChiM92-3 and (B) ChiM184-4 mice monoinfected with HBV/G. Profiles of total HBV DNA and HBV DNA of genotype G, determined via PCR with type-specific primers, are shown in the upper panels, and those of HBV antigens HBVcrAg, HBsAg, and HBeAg are shown in the lower panels. The shaded areas represent HBV DNA titers below the detection limit ($<10^3$ copies/ml).

RFLP (Fig. 3A). HBV/A accounted for all HBV DNA at week 17. At weeks 26 and 38, however, the vast majority of HBV DNA were of HBV/G with a trace of HBV/A. Thus, HBV/G needed coinfection with HBV/A for active replication, and took it over very swiftly.

Superinfection with HVB/G on Mice Infected with HBV/A. The chronological order of superinfection was reversed in ChiM92-9 and ChiM124-11 mice (Fig. 4A,B). The mice received 10^5 copies of HBV/A strains A2_JPN and A2_USA, respectively, and were superinfected with HBV/G (10^5 copies of G_US1646 strain) 10 weeks thereafter, when HBV/A DNA was elevated to $>5 \times 10^7$ copies/ml in both strains. Profiles of HBV DNA and antigens in these mice were quite similar but differed from those with A-on-G superinfection (Fig. 3A,B). HBV/A DNA was detected at week 1 in both groups and increased by approximately 2 logs within the next 3 weeks. HBV/G DNA developed within 3 weeks after superinfection with it, much sooner than the 12 weeks in ChiM mice superinfected with 2 genotypes in the reverse order. Three HBV antigens (HBcrAg, HBsAg, and HBeAg) waxed and waned in profiles similar to that of HBV DNA. HBV DNA levels decreased after they had peaked in ChiM92-9 as in A-on-G mice (Fig. 3A,B) by a margin close to log 2; the decrease was less prominent in ChiM124-11 by merely 1 log.

Composition of different genotypes in serum HBV DNA was followed in ChiM92-9 (Fig. 4A). Rapid replacement of HBV/A with HBV/G was obvious in G-on-A superinfection as in A-on-G superinfection (Fig. 3A). The takeover by HBV/G was not complete as in

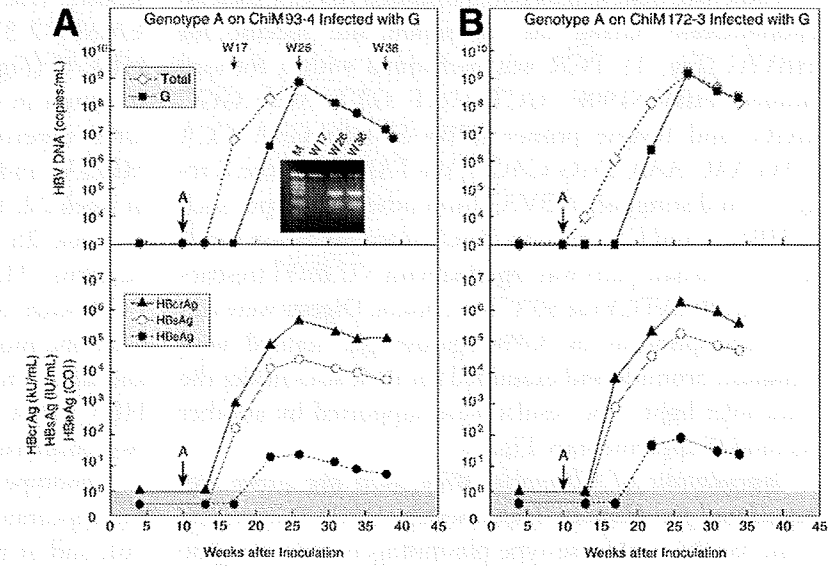

Fig. 3. Superinfection with HBV/A on (A) ChiM93-4 and (B) ChiM172-3 mice infected with HBV/G. Patterns of PCR-RFLP at different time points are shown in the insert in the upper panel of (A). Inoculation with genotype A is indicated by large arrows.

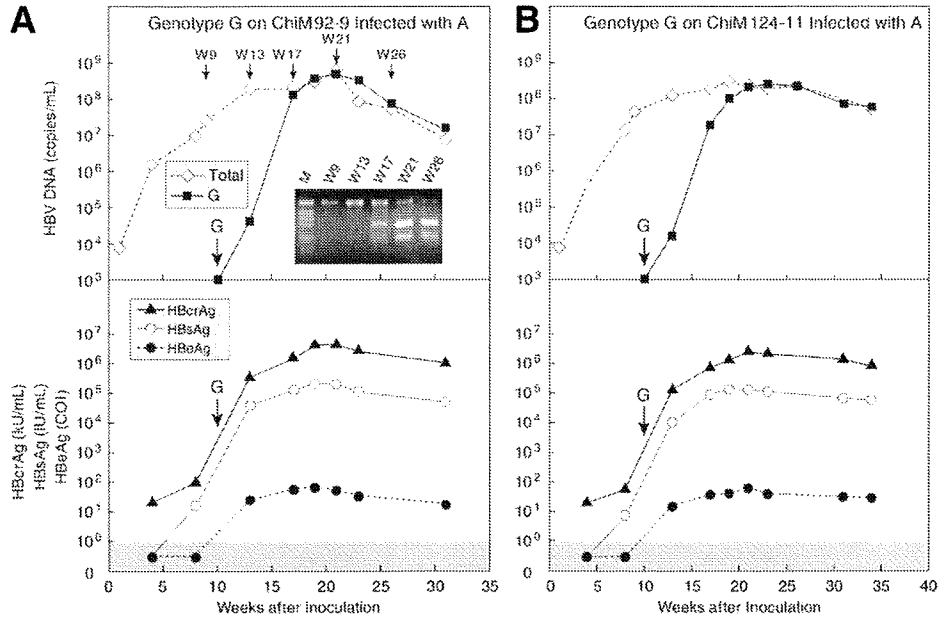

Fig. 4. Superinfection with HBV/G on (A) ChiM92-9 and (B) ChiM124-11 mice infected with HBV/A. Patterns of PCR-RFLP at different time points are shown in the insert in the upper panel of (A). Inoculation with genotype G is indicated by large arrows.

A-on-G superinfection, and HBV/A remained at very low levels throughout the weeks of observation.

Superinfection with HBV/G on Mice Infected with HBV/C. Similar superinfection with HBV/G was performed on ChiM mice that had been infected with HBV/C2 (Fig. 5A,B). Thus, C_22 and C_AT strains (10^5 copies) of HBV/C were injected intravenously into ChiM91-21 and ChiM95-11, respectively. They were superinfected with HBV/G (10^5 copies of G_US1646 strain) at week 10, when HBV DNA stabilized at approximately 10^9 copies/ml. HBV/G appeared in serum 3 weeks thereafter, at week 13 in both groups, and increased exponentially until weeks 21-23. The time required for an increase in HBV DNA level by 10-fold (log time) was 3.3 weeks in both groups, which was twice as long as the 1.6 weeks in mice with A-on-G and G-on-A superinfections (Figs. 3, 4). Likewise, the takeover of HBV/C by HBV/G in these mice was not as rapid or extensive as in superinfection with HBV/G on HBV/A (Figs. 3A, 4A, 5A). HBV antigens took time courses similar to that of HBV DNA, and they never waned after they had stabilized; however, mice were followed until 26 and 34 weeks.

Simultaneous Coinfection of Mice with HBV/A and HBV/G. Two ChiM mice (ChiM93-10 and ChiM93-

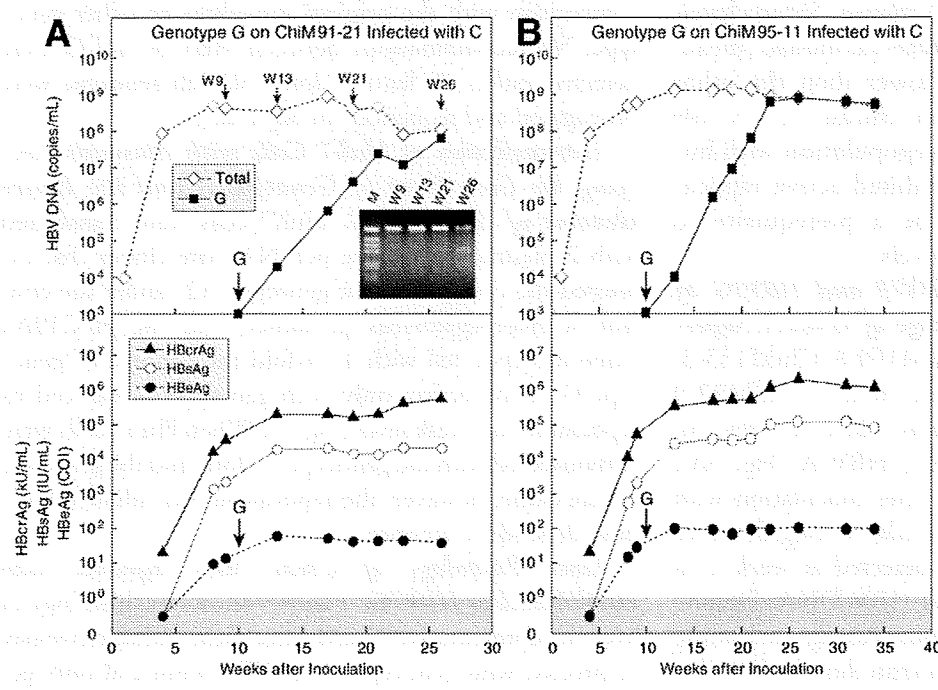

Fig. 5. Superinfection with HBV/G on (A) ChiM91-21 and (B) ChiM95-11 mice infected with HBV/C. Patterns of PCR-RFLP at different time points are shown in the insert the in upper panel of (A). Inoculation with genotype G is indicated by large arrows.

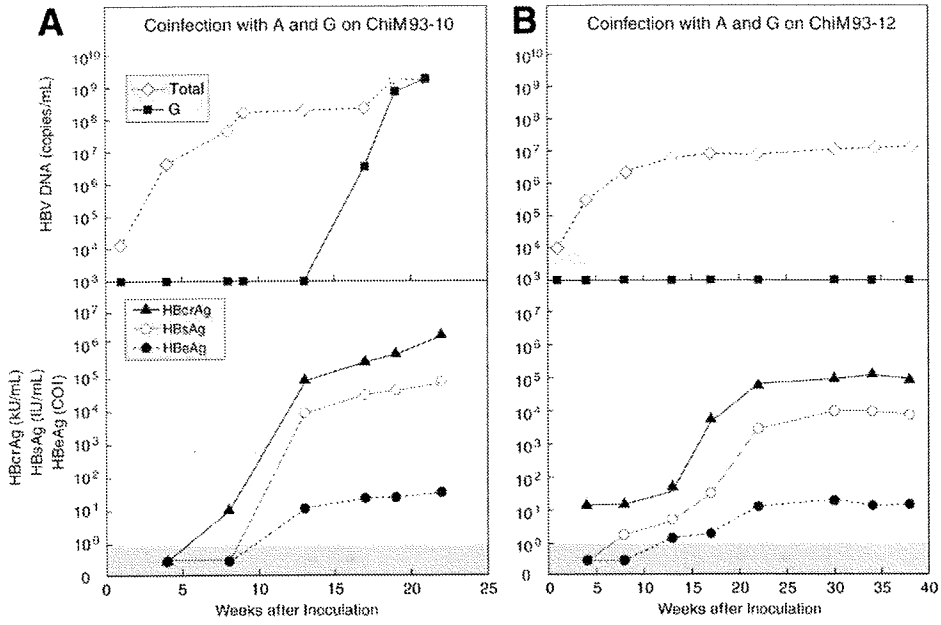

Fig. 6. (A) ChiM93-10 and (B) ChiM93-12 mice simultaneously infected with genotypes A and G.

12) received simultaneous inoculation with 10^5 copies each of HBV/A (A2_JPN strain) and HBV/G (G_US1646 strain). The ChiM93-10 mouse developed HBV/G DNA 17 weeks after inoculation, 9 weeks since HBV/A DNA had increased to $>10^7$ copies/ml (Fig. 6A). HBV/G DNA increased to the level of total HBV DNA at week 21, thereby indicating that by then, HBV/G had taken over HBV/A almost completely.

For the reasons unknown, infection with HBV/G was not established in the ChiM93-12 mouse simultaneously coinfected with HBV/A (Fig. 6B), although it was infected with HBV/A in levels by some 2 logs lower (10^7 copies/ml) than the ChiM93-10 mouse. Serum levels of human albumin in the ChiM93-12 mouse (mean, 2.1×10^6 ng/ml) were much lower than the other chimeric mice used in this study (mean, 4.7×10^6 ng/ml). Thus, a lower extent of repopulation with human hepatocytes may have prohibited active replication of HBV/A. This would be a prerequisite to infection with HBV/G at high levels.

Coinfection of Mice with HBV/A and HBV/G by Inoculation with a Mouse Passage of G-on-A Superinfection. Three ChiM mice (ChiM169-8, ChiM133-3, and ChiM133-6) received serum from a ChiM92-9 mouse with G-on-A superinfection taken at week 26, when HBV/G had almost replaced HBV/A (Fig. 3A). Profiles of HBV/A and HBV/G, after inoculation with 10^5 copies of HBV DNA, were similar among the mice (Fig. 7A-C). HBV/G DNA was detected at week 1 in levels comparable to those of total HBV DNA. Despite receiving the inoculation with a mouse passage containing HBV/G, in copies by 5 logs greater than those of HBV/A, HBV/G DNA decreased thereafter and stayed >1 log lower than total HBV DNA until week 7. Since week 4, HBV/G started to increase and replaced HBV/A almost completely until weeks 10-12, and continued to do so through weeks 19-22 of the observation (Fig. 7A).

Cloning and Sequencing HBV DNA in Chimeric Mice Coinfected with HBV/A and HBV/G. HBV DNA clones from sera of ChiM92-9 sampled at 26 weeks (Fig. 4A) and ChiM169-8 inoculated with serum passage in it (Fig. 7A) included those of HBV/A and G invariably. They confirmed the results of real-time detection PCR and PCR-RFLP and did not possess any mutations in comparison with the original inoculum of either genotype. No recombinations between HBV/A and G were detected, either. At least 5 clones of each genotype were propagated and sequenced in both sera.

Cotransfection of Huh7 Cells with Plasmids Carrying the Core Gene of Genotype A and the Entire Genome of Genotype G. Huh7 cells were transfected with 2 plasmids that were pcDNA_core clones that expressed the core protein of genotype A2, under the control of cytomegalovirus promoter, and the pUC19/G clone incorporated with 1.24-fold the genome of genotype G. Transfection only with genotype G induced its replication in a weak level (Fig. 8). When Huh7 cells were cotransfected with the genotype G clone and the genotype A core clone, however, the replication was enhanced in a dose-dependent manner.

Liver Pathology of ChiM Mice Infected with HBV/A and/or HBV/G. Figure 9 shows the histology of liver in representative ChiM mice either simultaneously coinfected with genotypes A and G (viremia of only ge-

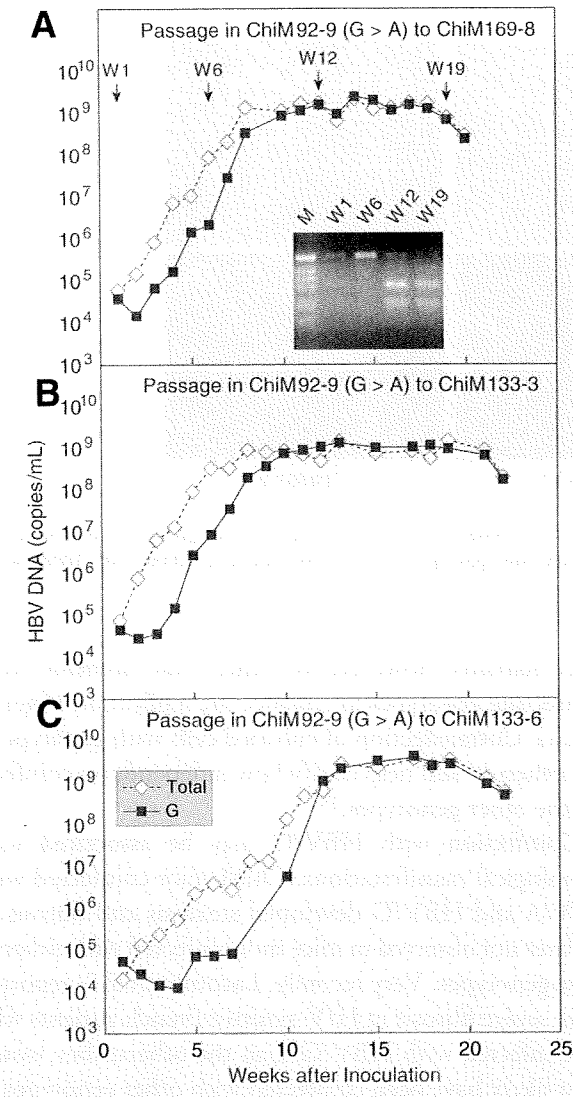

Fig. 7. (A) ChiM169-8, (B) ChiM133-3, and (C) ChiM133-6 mice inoculated with a serum passage from a mouse coinfected with genotypes A and G (ChiM92-9 in Fig. 4A).

notype A in ChiM93-12) or superinfected with genotypes G-on-A (ChiM92-9) and monoinfected with genotype G (ChiM92-3) during 32-39 weeks. HBV infection was demonstrated by double staining for HBcAg and human albumin (Supplementary Fig. 2). The mouse coinfected with genotypes A and G revealed steatosis of hepatocytes with hematoxylin-eosin stain and fibrosis of stage 2 (F2) with Masson's trichrome stain. In contrast, the mice monoinfected with genotype A (ChiM93-12) or G (ChiM92-3) had neither steatosis nor fibrosis. Table 1 summarizes the liver pathology of all autopsied mice. Steatosis in 30%-80% of repopulated human hepatocytes and stage F1-F2 fibrosis were observed in the majority of mice superinfected or coinfected with genotypes G and A or C.

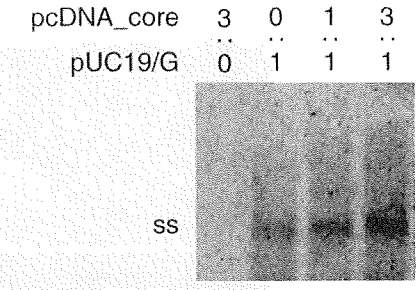

Fig. 8. *Trans*-complementation of the core gene of genotype A for enhanced replication of genotype G. Huh7 cells were cotransfected with plasmids constructed with 1.24-fold the genome of genotype G (pUC19/G) and plasmids expressing the core gene of genotype A (pcDNA_core) in an increasing ratio. Gel strips were Southern-blotted by the complete HBV probe of genotype G. The far left lane represents negative control with pcDNA_core alone. The migration position of single-stranded (ss) HBV DNA is indicated on the left.

Discussion

Using ChiM mice infected with pedigreed HBV DNA in the standardized copy number, we have determined early viral dynamics of HBV/G in detail. Due to constraints on securing ChiM mice with a satisfactory rate of replacement for human hepatocytes (>60%), only 2 or 3 of them were used for each experiment. Concordance of viral dynamics among them, however, would give credence to the reproducibility of obtained results.

HBV/G infected ChiM mice by itself in corroboration with its monoinfection in human beings.[21] The replication was very slow, however, and did not elevate serum HBV DNA to levels detectable by the method used (>10^3 copies/ml). Coinfection with HBV/A enhanced the replication of HBV/G remarkably. HBV/G replicated vividly when coinfected with HBV/C, as well. However, the time required for a 10-fold increase (log time) is 2-fold longer in mice initially infected with HBV/C versus HBV/A (3.3 versus 1.6 weeks). Combined, these results would indicate that HBV/G can thrive at the expense of other genotypes, and coinfection with HBV/A is much more advantageous for its enhanced replication than the other genotypes, including HBV/C. In support of this view, coinfection with HBV/A is frequent in individuals infected with HBV/G.[16,34] Such a heavy dependence of HBV/G on HBV/A does not require recombination between them, because no recombination events occurred in ChiM mice coinfected with them.

The initial replication of HBV/G was much slower than that of HBV/A, even in simultaneous coinfection. This was typically observed in three ChiM mice inoculated with a mouse passage of G-on-A superinfection containing HBV/G in the concentration a few logs higher than that of HBV/A. Despite such an enormous difference in introduced virions, the replication of HBV/A far

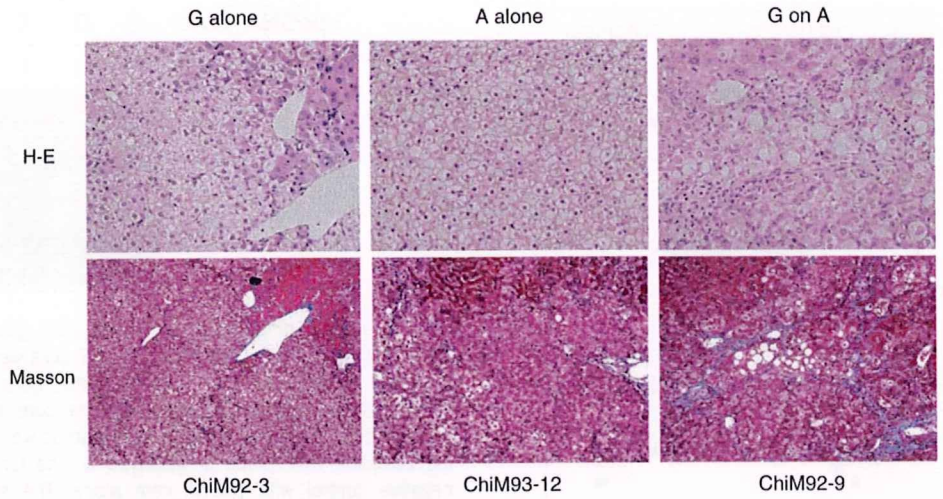

Fig. 9. Liver histology in a ChiM92-3 mouse monoinfected with genotype G, a ChiM93-12 mouse coinfected with genotypes A and G (but persistently infected with genotype A alone), and a ChiM92-9 mouse superinfected with genotype G-on-A. Liver sections stained with hematoxylin-eosin or Masson's trichrome stain are shown.

exceeded that of HBV/G in the initial several weeks. Thereafter, HBV/G caught up with HBV/A, then took it over almost completely. Such a replacement was observed when HBV/G was superinfected on HBV/A, or vice versa.

The mechanism by which genotype G depends on genotype A for replication was pursued in cotransfection experiments in Huh7 cells. Cotransfection with the pcDNA_core clone carrying the core gene of genotype A2 increased the replication of the pUC19/G clone constructed with 1.24-fold the genome of genotype G in a dose-dependent manner (Fig. 8). Hence, *trans*-complementation with the core protein of genotype A would be required for genotype G to replicate actively. The possibility remains, however, for other viral elements from coinfecting genotypes to enhance the replication of genotype G. Cotransfection of cultured cells with genotype G and others would help clarify how it depends on coinfecting the other genotypes.[35]

Table 1. Steatosis and Fibrosis in Human Hepatocytes in the Liver of Chimeric Mice Monoinfected or Coinfected with HBV/G

		Features	
Inoculation	Mouse No.	Steatosis (%)*	Fibrosis Stage
G alone	ChiM92-3	<5	F0
	ChiM 184-4	<5	F0
A alone	ChiM 93-12†	<5	F0
A-on-G	ChiM 93-4	50	F1
	ChiM 172-3	40	F1
G-on-A	ChiM 92-9	40	F2
	ChiM 124-11	50	F1
G-on-C	ChiM 91-21	80	F2
	ChiM 95-11	NA	NA
A plus G	ChiM 93-10	30	F0
Passage	ChiM 169-8	50	F1
A plus G	ChiM 133-3	<5	F2
	ChiM 133-6	30	F2

Abbreviation: NA, not available.

*Percentage of human hepatocytes with steatosis. †Simultaneously inoculated with A plus G but became infected with genotype A only (Fig. 6B).

Coinfection with HBV/G may be associated with pathological manifestations. ChiM mice coinfected with HBV/A and HBV/G developed steatosis and fibrosis in the liver not observed in mice monoinfected with either of these genotypes. Very recently, Lacombe et al.[22] reported more severe fibrosis in HIV-positive French patients who were infected with HBV/G than the others; they would most likely have been coinfected with other genotypes in undetectable levels. On the basis of clinical and experimental pieces of evidence, it does seem that HBV/G has a strong disease-inducing capacity, which would be operable only when it is coinfected with other genotypes. High levels of HBcrAg in mice with HBV/G (Figs. 3-6) under immunocompromised states would implicate accumulation of the product of the core gene in the fibrosis of patients coinfected with it and HIV. Patients with HIV are infected with HBV at a frequency of 6%-9%, and liver-related deaths happen more often in coinfected patients.[36,37] Fibrosis proceeds faster in patients coinfected with HIV and HBV, as in those with HCV.[38,39] Therapeutic intervention to prevent fibrosis would be required in patients coinfected with HIV and HBV, particularly in HBV/G patients.

In conclusion, the early viral dynamics of HBV/G have been characterized in ChiM mice monoinfected with HBV/G or coinfected with other genotypes. The replication of HBV/G is very slow and depends heavily on coinfection with other genotypes. HBV/G rapidly takes over

the other genotypes, though they are indispensable. Infection with HBV/G may induce steatosis and fibrosis in the liver—but again, only in the case of coinfection with other genotypes. However, it is still unclear whether or not such an increased pathogenicity of HBV/G is expressed exclusively in animals and patients with genetic or acquired immune deficiency.

References

1. Lee WM. Hepatitis B virus infection. N Engl J Med 1997;337:1733-1745.
2. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol 2002;83:2059-2073.
3. Norder H, Hammas B, Lofdahl S, Courouce AM, Magnius LO. Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. J Gen Virol 1992;73:1201-1208.
4. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, et al. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J Gen Virol 1988;69: 2575-2583.
5. Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, et al. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol 2000;81:67-74.
6. Naumann H, Schaefer S, Yoshida CF, Gaspar AM, Repp R, Gerlich WH. Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J Gen Virol 1993;74:1627-1632.
7. Miyakawa Y, Mizokami M. Classifying hepatitis B virus genotypes. Intervirology 2003;46:329-338.
8. Schaefer S. Hepatitis B virus: significance of genotypes. J Viral Hepat 2005;12:111-124.
9. Chu CJ, Lok AS. Clinical significance of hepatitis B virus genotypes. HEPATOLOGY 2002;35:1274-1276.
10. Liu CJ, Kao JH, Chen DS. Therapeutic implications of hepatitis B virus genotypes. Liver Int 2005;25:1097-1107.
11. Sugauchi F, Kumada H, Sakugawa H, Komatsu M, Niitsuma H, Watanabe H, et al. Two subtypes of genotype B (Ba and Bj) of hepatitis B virus in Japan. Clin Infect Dis 2004;38:1222-1228.
12. Tanaka Y, Orito E, Yuen MF, Mukaide M, Sugauchi F, Ito K, et al. Two subtypes (subgenotypes) of hepatitis B virus genotype C: A novel subtyping assay based on restriction fragment length polymorphism. Hepatol Res 2005;33:216-224.
13. Sugauchi F, Kumada H, Acharya SA, Shrestha SM, Gamutan MT, Khan M, et al. Epidemiological and sequence differences between two subtypes (Ae and Aa) of hepatitis B virus genotype A. J Gen Virol 2004;85:811-820.
14. Akuta N, Suzuki F, Kobayashi M, Tsubota A, Suzuki Y, Hosaka T, et al. The influence of hepatitis B virus genotype on the development of lamivudine resistance during long-term treatment. J Hepatol 2003;38:315-321.
15. Tanaka Y, Hasegawa I, Kato T, Orito E, Hirashima N, Acharya SK, et al. A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. HEPATOLOGY 2004;40:747-755.
16. Kato H, Orito E, Gish RG, Sugauchi F, Suzuki S, Ueda R, et al. Characteristics of hepatitis B virus isolates of genotype G and their phylogenetic differences from the other six genotypes (A through F). J Virol 2002;76:6131-6137.
17. Kato H, Orito E, Gish RG, Bzowej N, Newsom M, Sugauchi F, et al. Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. HEPATOLOGY 2002;35:922-929.
18. Perez-Olmeda M, Nunez M, Garcia-Samaniego J, Rios P, Gonzalez-Lahoz J, Soriano V. Distribution of hepatitis B virus genotypes in HIV-infected patients with chronic hepatitis B: therapeutic implications. AIDS Res Hum Retroviruses 2003;19:657-659.
19. Suwannakarn K, Tangkijvanich P, Theamboonlers A, Abe K, Poovorawan Y. A novel recombinant of hepatitis B virus genotypes G and C isolated from a Thai patient with hepatocellular carcinoma. J Gen Virol 2005;86:3027-3030.
20. Sanchez LV, Tanaka Y, Maldonado M, Mizokami M, Panduro A. Difference of hepatitis B virus genotype distribution in two groups of Mexican patients with different risk factors. High prevalence of genotype H and G. Intervirology 2007;50:9-15.
21. Chudy M, Schmidt M, Czudai V, Scheiblauer H, Nick S, Mosebach M, et al. Hepatitis B virus genotype G monoinfection and its transmission by blood components. HEPATOLOGY 2006;44:99-107.
22. Lacombe K, Massari V, Girard PM, Serfaty L, Gozlan J, Pialoux G, et al. Major role of hepatitis B genotypes in liver fibrosis during coinfection with HIV. AIDS 2006;20:419-427.
23. Heckel JL, Sandgren EP, Degen JL, Palmiter RD, Brinster RL. Neonatal bleeding in transgenic mice expressing urokinase-type plasminogen activator. Cell 1990;62:447-456.
24. Rhim JA, Sandgren EP, Degen JL, Palmiter RD, Brinster RL. Replacement of diseased mouse liver by hepatic cell transplantation. Science 1994; 263:1149-1152.
25. Tateno C, Yoshizane Y, Saito N, Kataoka M, Utoh R, Yamasaki C, et al. Near completely humanized liver in mice shows human-type metabolic responses to drugs. Am J Pathol 2004;165:901-912.
26. Mercer DF, Schiller DE, Elliott JF, Douglas DN, Hao C, Rinfret A, et al. Hepatitis C virus replication in mice with chimeric human livers. Nat Med 2001;7:927-933.
27. Tsuge M, Hiraga N, Takaishi H, Noguchi C, Oga H, Imamura M, et al. Infection of human hepatocyte chimeric mouse with genetically engineered hepatitis B virus. HEPATOLOGY 2005;42:1046-1054.
28. Sugiyama M, Tanaka Y, Kato T, Orito E, Ito K, Acharya SK, et al. Influence of hepatitis B virus genotypes on the intra- and extracellular expression of viral DNA and antigens. HEPATOLOGY 2006;44:915-924.
29. Kimura T, Ohno N, Terada N, Rokuhara A, Matsumoto A, Yagi S, et al. Hepatitis B virus DNA-negative dane particles lack core protein but contain a 22-kDa precore protein without C-terminal arginine-rich domain. J Biol Chem 2005;280:21713-21719.
30. Shinkai N, Tanaka Y, Orito E, Ito K, Ohno T, Hirashima N, et al. Measurement of hepatitis B virus core-related antigen as predicting factor for relapse after cessation of lamivudine therapy for chronic hepatitis B virus infection. Hepatol Res 2006;36:272-276.
31. Abe A, Inoue K, Tanaka T, Kato J, Kajiyama N, Kawaguchi R, et al. Quantitation of hepatitis B virus genomic DNA by real-time detection PCR. J Clin Microbiol 1999;37:2899-2903.
32. Mason AL, Xu L, Guo L, Kuhns M, Perrillo RP. Molecular basis for persistent hepatitis B virus infection in the liver after clearance of serum hepatitis B surface antigen. HEPATOLOGY 1998;27:1736-1742.
33. Kato H, Orito E, Sugauchi F, Ueda R, Gish RG, Usuda S, et al. Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. J Virol Methods 2001;98:153-159.
34. Osiowy C, Giles E. Evaluation of the INNO-LiPA HBV genotyping assay for determination of hepatitis B virus genotype. J Clin Microbiol 2003;41: 5473-5477.
35. Kremsdorf D, Garreau F, Capel F, Petit MA, Brechot C. In vivo selection of a hepatitis B virus mutant with abnormal viral protein expression. J Gen Virol 1996;77:929-939.
36. Konopnicki D, Mocroft A, de Wit S, Antunes F, Ledergerber B, Katlama C, et al. Hepatitis B and HIV: prevalence, AIDS progression, response to highly active antiretroviral therapy and increased mortality in the EuroSIDA cohort. AIDS 2005;19:593-601.
37. Thio CL, Seaberg EC, Skolasky R Jr, Phair J, Visscher B, Munoz A, et al. HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). Lancet 2002;360:1921-1926.
38. Benhamou Y, Bochet M, Di Martino V, Charlotte F, Azria F, Coutellier A, et al. Liver fibrosis progression in human immunodeficiency virus and hepatitis C virus coinfected patients. The Multivirc Group. HEPATOLOGY 1999;30:1054-1058.
39. Colin JF, Cazals-Hatem D, Loriot MA, Martinot-Peignoux M, Pham BN, Auperin A, et al. Influence of human immunodeficiency virus infection on chronic hepatitis B in homosexual men. HEPATOLOGY 1999;29:1306-1310.

DOI: 10.1534/genetics.106.069484

Transcriptional Interferences in *cis* Natural Antisense Transcripts of Humans and Mice

Naoki Osato, Yoshiyuki Suzuki, Kazuho Ikeo and Takashi Gojobori[1]

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan

Manuscript received December 9, 2006
Accepted for publication March 21, 2007

ABSTRACT

For a significant fraction of mRNAs, their expression is regulated by other RNAs, including *cis* natural antisense transcripts (*cis*-NATs) that are complementary mRNAs transcribed from opposite strands of DNA at the same genomic locus. The regulatory mechanism of mRNA expression by *cis*-NATs is unknown, although a few possible explanations have been proposed. To understand this regulatory mechanism, we conducted a large-scale analysis of the currently available data and examined how the overlapping arrangements of *cis*-NATs affect their expression level. Here, we show that for both human and mouse the expression level of *cis*-NATs decreases as the length of the overlapping region increases. In particular, the proportions of the highly expressed *cis*-NATs in all *cis*-NATs examined were ~36 and 47% for human and mouse, respectively, when the overlapping region was <200 bp. However, both proportions decreased to virtually zero when the overlapping regions were >2000 bp in length. Moreover, the distribution of the expression level of *cis*-NATs changes according to different types of the overlapping pattern of *cis*-NATs in the genome. These results are consistent with the transcriptional collision model for the regulatory mechanism of gene expression by *cis*-NATs.

BIOLOGICAL processes such as development, metabolism, and response to external stimuli are conducted by the cooperative activities of many genes. To understand a biological process, it is essential to understand the regulatory network of genes composing the biological process. After genome sequences had been determined, attempts to reveal regulatory networks of genes were started (Wyrick and Young 2002; Encode Project Consortium 2004; Carninci *et al.* 2005; Levine and Davidson 2005). Regulation of gene expression can be conducted mainly by proteins such as transcription factors. However, it has been found that ~20–30% of mammalian transcripts are targets of microRNAs, which bind to complementary mRNAs and inhibit their activation (Krek *et al.* 2005; Lewis *et al.* 2005; Stark *et al.* 2005; Carthew 2006). This suggests that the regulation of gene expression by RNAs is more ubiquitous and important than we thought (Mattick 2001, 2004).

Cis natural antisense transcripts (*cis*-NATs) are composed of a pair of mRNAs that are transcribed from the opposite strands of DNA at the same genomic locus. The antisense mRNA regulates the expression level of the sense mRNA in a pair. As a result, *cis*-NATs affect the developmental processes such as neural, eye, and tooth formation (Potter and Branford 1998; Korneev *et al.* 1999; Alfano *et al.* 2005; Coudert *et al.* 2005; Korneev and O'Shea 2005), and various molecular functions such as X-inactivation, genomic imprinting, DNA methylation, RNA editing, and alternative splicing (Munroe and Lazar 1991; Kumar and Carmichael 1997; Moore *et al.* 1997; Lee *et al.* 1999; Tufarelli *et al.* 2003).

Although the number of experimentally verified *cis*-NATs was ~40, >2000 *cis*-NATs were predicted in the analyses of the genomic and cDNA sequences in humans and mice (Lehner *et al.* 2002; Shendure and Church 2002; Kiyosawa *et al.* 2003; Yelin *et al.* 2003; Katayama *et al.* 2005). Recently, an analysis of a human oligo microarray showed that as much as 60% of surveyed loci on human chromosome 10 were predicted to encode *cis*-NATs (Cheng *et al.* 2005). Other chromosomes were also expected to encode as many *cis*-NATs. In addition, the presence of *cis*-NATs has been predicted for other eukaryotes as well as prokaryotes (Wagner and Simons 1994; Vanhee-Brossollet and Vaquero 1998; Makalowska *et al.* 2005). These observations implied that the regulation of gene expression by *cis*-NATs would occur more frequently than previously considered. Regulation of gene expression by RNAs may be evolutionarily advantageous, because it regulates gene expression quickly and saves energy and time in synthesizing proteins. Chen *et al.* (2005b) showed that *cis*-NATs were encoded in genes with shorter intron sequences than other mRNAs.

[1]*Corresponding author:* Center for Information Biology and DNA Data Bank of Japan, National Institutes of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan. E-mail: tgojobor@genes.nig.ac.jp

Although a large number of *cis*-NATs have been predicted from various species, the regulatory mechanisms of gene expression by *cis*-NATs remain unclear. To understand the regulatory mechanisms, it will be crucial to know the features of *cis*-NATs that are important in the regulation of gene expression. Thus far, three models have been proposed for the regulation of gene expression by *cis*-NATs (Lavorgna *et al.* 2004).

The first model asserts that *cis*-NATs form a double strand through their complementary sequences, which leads to the inhibition of the function of mRNAs, including protein synthesis. In this model, it is expected that *cis*-NATs are overlapped in at least 6- to 8-bp regions to form stable double strands of RNA (Lai 2002; Lewis *et al.* 2005).

The second model involves epigenetic regulations such as the methylation of promoters and the conversion of the chromosome structure (Wutz *et al.* 1997; Reik and Walter 2001; Tufarelli *et al.* 2003). Through unknown mechanisms, the antisense mRNAs methylate promoters of sense mRNAs and inhibit the transcription of sense mRNAs. In addition, the antisense mRNAs convert the chromosomal structures in which *cis*-NATs are located and regulate the expression of the sense mRNAs. For the model of epigenetic regulations, the features of *cis*-NATs that are essential for the regulations are unclear.

The third model is transcriptional collisions. In the transcription of *cis*-NATs, RNA polymerases bind to the promoters of genes encoding sense and antisense mRNAs and synthesize mRNAs, moving toward the 3′-end of the genes. RNA polymerases clash in the overlapping region and inhibit their transcription (Figure 4). This model was implied from the analyses of the expression levels of *cis*-NATs in yeast (Peterson and Myers 1993; Puig *et al.* 1999; Prescott and Proudfoot 2002). In the analyses, the expression level of *cis*-NATs decreased as the length of the overlapping region of the *cis*-NATs increased. Moreover, the expression level of adjacent transcripts in tandem on the same strand of the yeast genome decreased when the terminator of the upstream transcript was removed. This suggested that RNA polymerases did not stop at the terminator of the upstream transcript and affected the transcription of the downstream transcript. Recently, the collision of *Escherichia coli* RNA polymerases was observed by atomic force microscopy (Crampton *et al.* 2006). This observation showed that RNA polymerases do not pass each other or displace one another, but instead stall against each other.

Here, we report the effects of length and pattern of the overlapping regions of *cis*-NATs for humans and mice on their expression level. Moreover, human and mouse adjacent transcripts, in a particular position, affect their expression level. These results are consistent with the transcriptional collision model, implying that the regulation of the expression of *cis*-NATs by transcriptional collisions is common among species.

MATERIALS AND METHODS

Analysis of the expression level of human *cis*-NATs: To predict *cis*-NATs from human cDNA sequences, we collected a total of 46,675 human cDNA sequences, which consisted of 39,530 human Ensembl cDNA sequences (February 2006) (Hubbard *et al.* 2005), 6501 human Ensembl non-protein-coding sequences (February 2006), and 644 non-protein-coding sequences in RNAdb (Pang *et al.* 2005) that were mapped to the human genome by BLAT software (Kent 2002). We predicted 8964 *cis*-NATs, which had an at least 1-bp-long overlapping region, on the basis of their genomic location using in-house Perl scripts. Redundant *cis*-NATs were merged into the same group when *cis*-NATs overlapped in an at least 1-bp-long region in the genome. The number of the groups of *cis*-NATs was 2496. To examine the expression levels of human *cis*-NATs, we employed 15.5 million *Nla*III human serial analysis of gene expression (SAGE) tags (November 2005) of all tissues in the NCBI SAGEmap database (Lash *et al.* 2000). Human SAGE tags were searched against a total of 46,675 human cDNA sequences according to the protocol for SAGE (Velculescu *et al.* 1995) using in-house Perl scripts. When a SAGE tag matched more than one transcript, these transcripts were removed from further analyses. As a result, among the 46,675 human cDNA sequences, 28,009 had unique SAGE tag assignments, and among the 2496 groups of human *cis*-NATs, 728 groups had unique SAGE tag assignments to all transcripts in each group. Some SAGE tags are supposed to be assigned to overlapping regions of *cis*-NATs. Although cDNA microarrays cannot distinguish the expression of mRNAs encoded in plus and minus strands of the genome in the same locus, SAGE tags are strand specific. Therefore, even though SAGE tags are produced from the overlapping regions of *cis*-NATs, the expression levels of sense and antisense mRNAs of *cis*-NATs can be measured separately.

To examine the expression levels of the human *cis*-NATs, we compared the expression level of human *cis*-NATs with that of other human transcripts (*i.e.*, human transcripts excluding *cis*-NATs, pseudogenes, and non-protein-coding mRNAs). We calculated the ratio of the expression level of a human *cis*-NAT to that of the other human transcripts. However, the expression levels of transcripts are known to be affected by the overall length of the transcripts (Castillo-Davis *et al.* 2002). Thus, we compensated the expression level of *cis*-NATs according to their overall length. The expression level of a human *cis*-NAT was compared with that of the other human transcripts with almost the same overall length of the *cis*-NAT. We removed human *cis*-NATs, pseudogenes, and non-protein-coding mRNAs from human cDNA sequences and selected 2000 human transcripts of which overall length in the genome was close to the overall length of a human *cis*-NAT in the genome. We calculated the median of the expression levels of the selected 2000 human transcripts (supplemental Figure 1A at http://www.genetics.org/supplemental/). The median of the expression levels of the selected human transcripts was defined as a ratio of 1.0, and then the ratio of the expression level of a *cis*-NAT was calculated as follows:

$$R_{cis\text{-NAT}} = T_{cis\text{-NAT}} / T_{(\text{all}-cis\text{-NAT}-\text{pseudo}-\text{noncoding})},$$

where $R_{cis\text{-NAT}}$ is ratio of the expression level of a human *cis*-NAT to that of other human protein-coding transcripts, $T_{cis\text{-NAT}}$ is the expression level of a human *cis*-NAT, and $T_{(\text{all}-cis\text{-NAT}-\text{pseudo}-\text{noncoding})}$ is the median of the expression level of other human transcripts (*i.e.*, human transcripts excluding *cis*-NATs, pseudogenes, and non-protein-coding mRNAs) with almost the same overall length of the *cis*-NAT.

Sense mRNAs of *cis*-NATs are located not only in the plus or the minus strand of the genome, but also in both strands

of the genome. Some of the sense mRNAs encode proteins and others encode non-protein-coding mRNAs that may have some biological function such as the regulation of mRNA translation and stability (MATTICK and MAKUNIN 2006). The antisense mRNA in each group of *cis*-NATs is known to decrease the expression level of the sense mRNA in the group and to inhibit the activation of the sense mRNA (WAGNER and SIMONS 1994; KUMAR and CARMICHAEL 1998; VANHEE-BROSSOLLET and VAQUERO 1998). Therefore, we recognized the *cis*-NAT with the lowest expression level in each group as the sense mRNA in the group. In this study, we used the expression level of the sense mRNA in each group to examine the expression levels of *cis*-NATs according to the overlapping arrangements in the genome. However, there is an assumption that under some regulatory mechanisms of *cis*-NATs, such as RNA masking and a double-stranded RNA-dependent mechanism, the expression level of a sense mRNA may not be the lowest in the group of *cis*-NATs and an inverse correlation of the expression levels may not be found between a sense and an antisense mRNA (CHEN *et al.* 2005a; KATAYAMA *et al.* 2005; LAPIDOT and PILPEL 2006). Therefore, we also examined the expression level of *cis*-NATs randomly selected (*i.e.*, selected in a non-expression-level-dependent manner) (see supplemental material at http://www.genetics.org/supplemental/).

We examined the expression levels of *cis*-NATs as the overlapping regions increased in length. When *cis*-NATs included more than two transcripts (*i.e.*, sense and antisense transcripts on both strands of the genome), the length of the overlapping region was defined as the distance between the farthest upstream and the farthest downstream genomic locations of the overlapping regions of *cis*-NATs in a group.

Analysis of the expression level of mouse *cis*-NATs: To predict the mouse *cis*-NATs, we collected a total of 35,486 mouse cDNA sequences, which consisted of 33,252 mouse Ensembl cDNA sequences (April 2006), 1752 mouse Ensembl non-protein-coding sequences (April 2006), and 482 non-protein-coding sequences in RNAdb that were mapped to the mouse genome using BLAT. We predicted 5491 *cis*-NATs from the mouse cDNA sequences and redundant *cis*-NATs were merged into the same group. There were 1868 groups of mouse *cis*-NATs. We examined the expression levels of mouse *cis*-NATs using 3.6 million *Nla*III mouse SAGE tags of all tissues in the NCBI SAGEmap database (November 2005) to compare them with the expression levels of human *cis*-NATs. Among the 35,486 mouse cDNA sequences, 21,982 had unique SAGE tag assignments, and among the 1868 groups of mouse *cis*-NATs, 704 groups had unique SAGE tag assignments to all transcripts, including alternative forms in each group. We examined the expression levels of mouse *cis*-NATs in the same way we examined those of humans.

Comparison of *cis*-NATs of humans and mice at the nucleotide level: To find *cis*-NATs conserved between humans and mice, we compared 46,675 human cDNA sequences with 35,486 mouse cDNA sequences and vice versa using the all-against-all FASTA (PEARSON and LIPMAN 1988) procedure. An expected (E)-value cutoff of 1.0×10^{-20} was used. We selected the human and mouse cDNA sequences that matched reciprocally with an E-value less than the square root of the lowest E-value as candidates of orthologs. When human *cis*-NATs in both strands of the genome are orthologous to mouse *cis*-NATs in both strands of the genome, we recognized the *cis*-NATs as conserved between human and mouse.

RESULTS

Length of the overlapping region of human *cis*-NATs affects their expression level: To predict *cis*-NATs from human cDNA sequences, we searched 46,675 human cDNA sequences. A total of 8964 *cis*-NATs were predicted and were clustered into 2496 groups, each of which consisted of sense and antisense transcripts as well as their alternative forms. To examine the expression levels of human cDNA sequences, ~15.5 million *Nla*III SAGE tags were collected from all the human tissues available in the NCBI SAGEmap database (November 2005) (LASH *et al.* 2000) and were compared to 46,675 human cDNA sequences. Among the 8964 (2496 groups) *cis*-NATs, 2038 (728 groups) had unique SAGE tag assignments to all transcripts in each group.

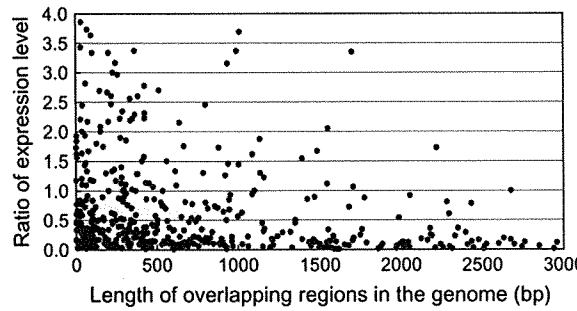

FIGURE 1.—Distribution of the expression level of human *cis*-NATs as the overlapping region in the genome increased in length. The *x*-axis shows the length of the overlapping exon and intron regions of human *cis*-NATs in the genome. The *y*-axis shows the ratio of the expression level of a human *cis*-NAT to the median of the expression levels of human transcripts of which overall lengths in the genome are close to the overall length of the human *cis*-NAT in the genome. The median of the expression levels of human transcripts is defined as a ratio of 1.0. Because the expression levels of human transcripts are affected by their overall length, we compensated for the expression level of a *cis*-NAT according to its overall length (see MATERIALS AND METHODS).

To examine whether the length of overlapping regions in *cis*-NATs affects their expression level, we investigated the relationship between the expression levels of *cis*-NATs and the length of the overlapping region in the genome. It should be noted that CASTILLO-DAVIS *et al.* (2002) found that highly expressed transcripts tended to have short introns, implying that the short *cis*-NATs may be expressed at a higher level than the long *cis*-NATs. To eliminate the effect of the overall length of a transcript on its expression level, we selected nonoverlapping transcripts whose overall lengths in the genome were almost the same as that of a *cis*-NAT in the genome and then compared the expression level of the *cis*-NATs with that of the selected transcripts (see MATERIALS AND METHODS).

In Figure 1, the *y*-axis represents the ratio of the expression level of the human *cis*-NATs to that of the selected human transcripts. The proportion of highly expressed *cis*-NATs (ratio >1.0) in all *cis*-NATs examined was 36% when the overlapping region was between 1 and 200 bp. However, the proportion decreased to virtually zero when the overlapping regions were >2000

bp long (chi-square $P < 10^{-15}$). This result suggests that the expression level of *cis*-NATs decreases as the length of the overlapping region increases. In addition, the expression levels of *cis*-NATs may be influenced by other factors such as alternative transcripts of *cis*-NATs, the difference of the expression levels among SAGE tag libraries, GC content bias of SAGE tags (Margulies *et al.* 2001), the experimental methods for the analysis of gene expression, the ways of selecting sense mRNAs, the criteria for the length of the overlapping regions, and a set of human cDNA sequences used in this analysis. However, these factors did not change the overall distribution of the expression levels significantly (see supplemental material at http://www.genetics.org/supplemental/).

Overlapping pattern of human *cis*-NATs affects their expression level: *Cis*-NATs are classified into three types on the basis of overlapping patterns in the genome: head to head, tail to tail, and full overlap (Figure 2). "Full overlap" describes *cis*-NATs where the sense mRNA entirely overlaps within the antisense mRNA. The numbers of head-to-head, tail-to-tail, and full-overlap types of human *cis*-NATs were 254, 476, and 1766 groups of which 126, 230, and 356 groups had unique SAGE tag assignments to all transcripts in each group, respectively. To examine whether the overlapping pattern of *cis*-NATs affects their expression level, we analyzed the expression level of *cis*-NATs of humans according to the overlapping patterns in the genome. Figure 3, A–C, shows the expression levels of human *cis*-NATs in the head-to-head, tail-to-tail, and full-overlap manners, respectively. The highly expressed *cis*-NATs decreased in quantity as the overlapping region increased in length for all types of *cis*-NATs. However, highly expressed *cis*-NATs in head-to-head and full-overlap manners decreased in quantity more than those in a tail-to-tail manner did. When the length of the overlapping region was <600 bp, 26.7% of *cis*-NATs in a head-to-head manner showed high expression (ratio >1.0) and 43.4% of *cis*-NATs in a tail-to-tail manner showed high expression. The proportion of highly expressed *cis*-NATs in a head-to-head manner (26.7%) was 1.6 times smaller than that in a tail-to-tail manner (43.4%) (Mann–Whitney *U*-test: $P < 10^{-2}$). Similarly, when the length of the overlapping region was <600 bp, 24.5% of *cis*-NATs in a full-overlap manner showed high expression. The proportion of highly expressed *cis*-NATs in a full-overlap manner (24.5%) was 1.7 times smaller than that in a tail-to-tail manner (43.4%) (Mann–Whitney *U*-test: $P < 10^{-3}$). Among the 356 *cis*-NATs in a full-overlap manner, 314 were *cis*-NATs where a sense transcript overlapped only in the intron regions of the antisense transcript in the genome. The mRNAs that overlapped in the intron regions showed the same feature of expression.

As many as 1450 human transcripts were found to be located within a distance of <1 kbp in the genome (Adachi and Lieber 2002; Koyanagi *et al.* 2005). To examine whether the expression levels of nearby transcripts decreased, we investigated the expression levels of human nearby transcripts. Figure 3D shows the expression levels of nearby transcripts where the 5′-end of a transcript is near the 5′-end of another transcript in the genome. Here, we call them "nearby transcripts in a head-to-head manner" (Figure 2). When the distance between the 5′-ends of transcripts was < ~50 bp, highly expressed transcripts (a ratio >1.0) were not observed (chi-square $P < 10^{-7}$).

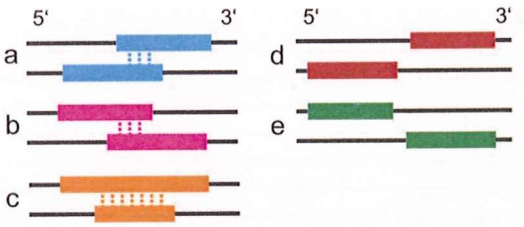

Figure 2.—Classification of *cis*-NATs and nearby transcripts on the basis of their relative positions in the genome. *Cis*-NATs are classified on the basis of their relative positions in the genome: (a) *cis*-NATs in a head-to-head manner (5′-end to 5′-end), (b) those in a tail-to-tail manner (3′-end to 3′-end), and (c) those in a full-overlap manner. Full overlap describes the *cis*-NATs where a transcript on a strand of the genome is overlapped by the entire length of the other transcript on the opposite strand of the genome. (d) Nearby transcripts in a head-to-head manner where the 5′-end of a transcript is near the 5′-end of another transcript in the genome. (e) Nearby transcripts in a tail-to-tail manner where the 3′-end of a transcript is near the 3′-end of another transcript in the genome.

Figure 3E shows the expression levels of nearby transcripts where the 3′-end of a transcript is near the 3′-end of another transcript in the genome. Here, we call them "nearby transcripts in a tail-to-tail manner." Contrary to nearby transcripts in a head-to-head manner, the expression levels of nearby transcripts in a tail-to-tail manner did not change, regardless of the distance of the nearby transcripts in the genome (chi-square $P = 0.7$).

There is a possibility that the length of some human transcripts registered in a database such as the Ensembl database may be shorter than natural transcripts (Makalowska *et al.* 2005). Nearby transcripts found in the Ensembl database may, in fact, overlap in the genome, such that the expression levels of such artificial nearby transcripts seemed to decrease. However, almost all nearby transcripts in a head-to-head manner were expressed at a low level when the distance of the transcripts was < ~50 bp. Therefore, the decrease in the expression level of nearby transcripts will be a natural phenomenon.

Overlapping arrangements of mouse *cis*-NATs affect their expression levels: *Cis*-NATs have been predicted for various species (Wagner and Simons 1994; Vanhee-Brossollet and Vaquero 1998; Wagner and Flardh 2002; Makalowska *et al.* 2005). If the regulatory

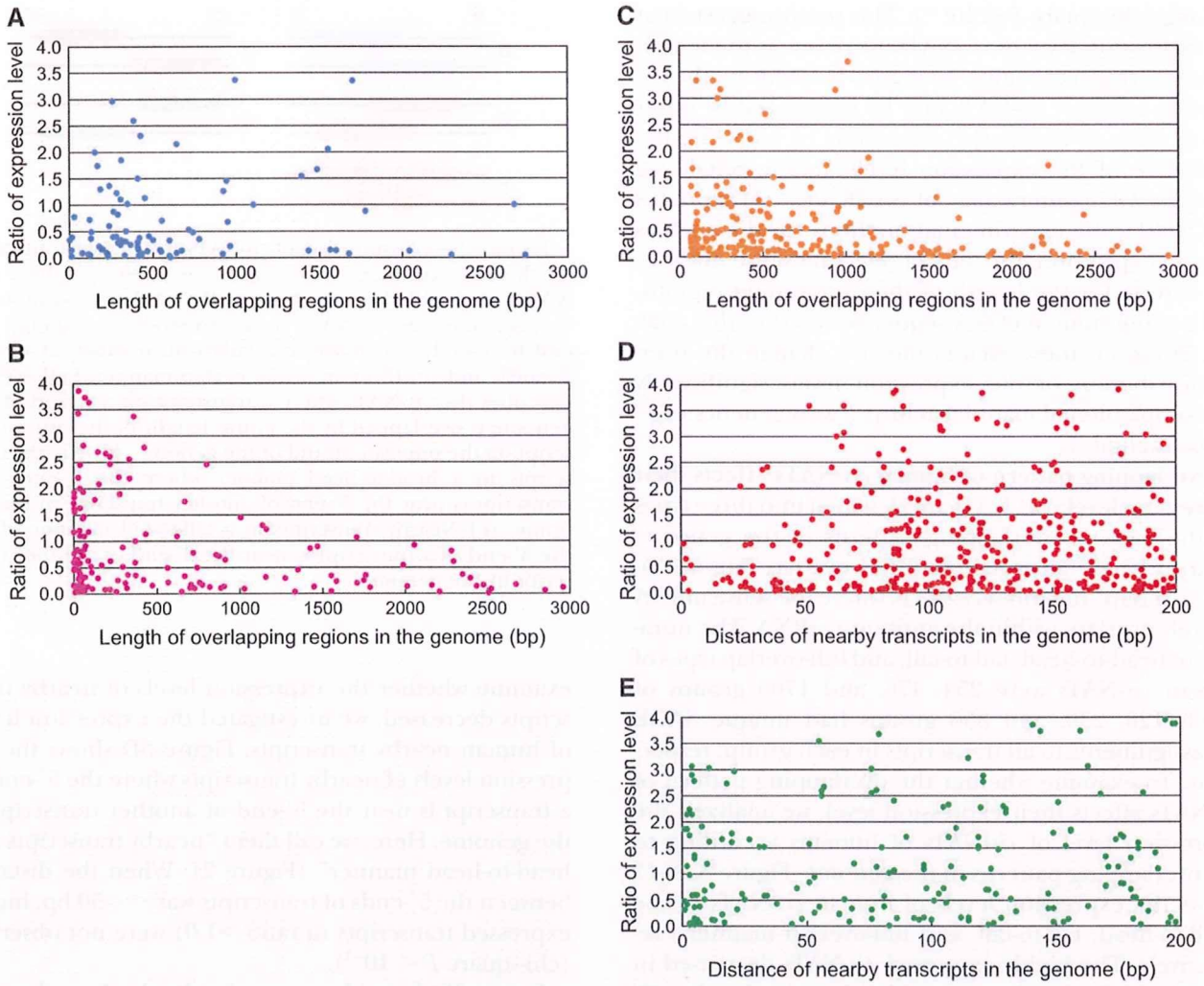

FIGURE 3.—Distribution of the expression level of human *cis*-NATs according to overlapping patterns in the genome. The *x*-axis shows the length of the overlapping exon and intron regions of human *cis*-NATs in the genome. The *y*-axis shows the ratio of the expression level of a *cis*-NAT to the median of the expression levels of human transcripts whose overall lengths in the genome are close to those of the human *cis*-NAT in the genome. The median of the expression levels of human transcripts is defined as a ratio of 1.0. Because the expression levels of human transcripts are affected by their overall length, we compensated the expression level of a *cis*-NAT according to its overall length (see MATERIALS AND METHODS). (A) *cis*-NATs in a head-to-head manner, (B) those in a tail-to-tail manner, (C) those in a full-overlap manner, (D) nearby transcripts in a head-to-head manner, and (E) nearby transcripts in a tail-to-tail manner.

mechanisms of *cis*-NATs in gene expression are conserved among species, *cis*-NATs of another species are expected to show similar effects on their expression levels. To evaluate whether the relationship between the overlapping arrangements and the expression levels of *cis*-NATs is conserved among species, we examined the expression levels of mouse *cis*-NATs. Almost the same number of *cis*-NATs (1771 groups) as that of humans was found in mouse cDNA sequences (KIYOSAWA *et al.* 2003; YELIN *et al.* 2003). Mouse *cis*-NATs were compared to 3.6 million *Nla*III mouse SAGE tags in the NCBI SAGEmap database. Although the number of mouse SAGE tags and the number of mouse *cis*-NATs (705 groups) assigned to unique SAGE tags was smaller than for humans, the distribution of the expression levels of mouse *cis*-NATs showed the same features as that of humans (supplemental Figure 2A at http://www.genetics.org/supplemental/): highly expressed (ratio >1.0) *cis*-NATs decreased in quantity when the overlapping regions in the genome increased in length. The proportion of highly expressed *cis*-NATs in all *cis*-NATs examined was 47% when the overlapping region was between 1 and 200 bp, and the proportion decreased virtually to zero when the overlapping regions were >2000 bp long (chi-square $P < 10^{-15}$).

For overlapping patterns in the genome, the distribution of the expression levels in mice changed in the same way as in humans (Mann–Whitney *U*-test: $P = 0.52$ between human and mouse *cis*-NATs in a head-to-head manner, $P = 0.92$ between those in a tail-to-tail manner, and $P = 0.17$ between those in a full-overlap manner) (supplemental Figure 2, B–D, at http://www.genetics.org/

supplemental/). When the length of the overlapping region was <600 bp, 27.6% of *cis*-NATs in a head-to-head manner showed high expression (ratio >1.0) and 44.1% of *cis*-NATs in a tail-to-tail manner showed high expression. The proportion of highly expressed *cis*-NATs in a head-to-head manner (27.6%) was 1.6 times smaller than that of those in a tail-to-tail manner (44.1%) (Mann–Whitney *U*-test: $P < 10^{-2}$). Similarly, 25.9% of *cis*-NATs in a full-overlap manner showed high expression (ratio >1.0) and the proportion of highly expressed *cis*-NATs in a full-overlap manner (25.9%) was 1.7 times smaller than that in a tail-to-tail manner (44.1%) (Mann–Whitney *U*-test: $P < 0.05$). With nearby transcripts, when the distance between nearby transcripts in a head-to-head manner was $< \sim 50$ bp, highly expressed transcripts were not observed as found in humans (chi-square $P < 10^{-5}$) (supplemental Figure 2E at http://www.genetics.org/supplemental/). The expression levels of nearby transcripts in a tail-to-tail manner did not change, regardless of the distance of the nearby transcripts in the genome (chi-square $P = 0.9$) (supplemental Figure 2F at http://www.genetics.org/supplemental/). These results suggest that the expression levels of mouse *cis*-NATs are affected by the overlapping arrangements in the genome in the same way as those of humans. This implies that the regulatory mechanisms of *cis*-NATs in gene expression are conserved between humans and mice.

However, there was a possibility that *cis*-NATs showed a similar distribution of the expression level between humans and mice because most human and mouse *cis*-NATs were orthologous (Liao and Zhang 2006). To address this possibility, we compared the distribution of the expression levels of *cis*-NATs that are not conserved between human and mouse. First, we compared human and mouse cDNA sequences by using FASTA (Pearson and Lipman 1988) and found that only 329 groups (11.9%) of human *cis*-NATs were conserved in mouse *cis*-NATs, although 34,670 (76.0%) of human cDNA sequences were conserved in mice. Human and mouse *cis*-NATs were supposed to be highly divergent in terms of cDNA sequences (Veeramachaneni *et al.* 2004; Makalowska *et al.* 2005). We removed the 329 groups of *cis*-NATs from 2765 groups of human and 1704 groups of mouse *cis*-NATs, which left 710 groups of human and 605 groups of mouse *cis*-NATs with SAGE tags assigned to all transcripts in each group. We examined the expression level of the human and mouse *cis*-NATs according to the length and the pattern of the overlapping region in the genome. They showed almost the same distribution of the expression level as those including *cis*-NATs conserved between humans and mice (Mann–Whitney *U*-test: $P = 0.22$ and $P = 0.88$ for humans and mice, respectively). These results suggested that the similarity of the distribution of the expression level of human and mouse *cis*-NATs was not due to the conservation of the cDNA sequences of the *cis*-NATs.

DISCUSSION

We found that the expression level of *cis*-NATs changed according to the overlapping arrangements of *cis*-NATs in the human and mouse genomes. The expression level of *cis*-NATs decreased when the overlapping regions increased in length. Moreover, the overlapping pattern of *cis*-NATs affects their expression level. Nearby transcripts in a particular position decreased their expression levels.

Here, we examined the expression level of *cis*-NATs using SAGE tags and oligonucleotide arrays in public databases. We obtained the same distribution of the expression level of *cis*-NATs at least in the same tissue or cell such as fetal brain, embryonic stem cell, and liver (supplemental material and supplemental Figures 3 and 4 at http://www.genetics.org/supplemental/). However, all SAGE libraries and the expression data of oligonucleotide arrays were produced from cells and tissues of humans and mice, not from a single cell. In addition, some *cis*-NATs may be expressed only in some developmental stages or at a specific time. Therefore, some *cis*-NATs may not be expressed concurrently in the same single cell.

Thus far, three models have been proposed for the regulation of gene expression by *cis*-NATs. Forming double strands of *cis*-NATs requires a minimum 6- to 8-bp overlapping region of *cis*-NATs. We found that the expression level of human and mouse *cis*-NATs decreased consecutively as the length of the overlapping region increased. However, currently there is no report that this result is brought about by this model. In addition, this model does not intend to explain the change of the expression levels of nearby transcripts. Epigenetic regulations are also considered to be involved in the regulation of *cis*-NATs in gene expression. However, currently there is no report that by this model the expression level of human and mouse *cis*-NATs decreases consecutively as the length of the overlapping region increases.

Our results are consistent with the transcriptional collision model that has been proposed following the analyses of adjacent transcripts and *cis*-NATs in yeast (Puig *et al.* 1999; Prescott and Proudfoot 2002) and the observation of the collision of RNA polymerases by atomic force microscopy (Crampton *et al.* 2006). RNA polymerases bind to the upstream regions of genes and synthesize mRNAs, moving toward the 3′-ends of the genes. When opposite genomic strands in the same locus encode complementary mRNAs like *cis*-NATs, an RNA polymerase bound on a strand of the genome collides with the RNA polymerase bound on the opposite strand during the transcription of both strands of mRNAs (Figure 4). This leads to the inhibition of transcription. From this model, the frequency of the collisions of RNA polymerases is expected to increase when the overlapping regions increase in length. Moreover, overlapping patterns in the genome would affect the

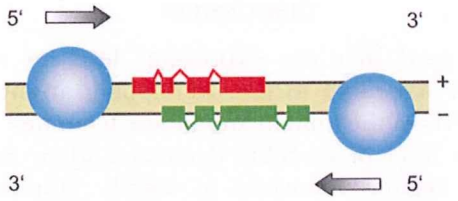

FIGURE 4.—Transcriptional collision model When *cis*-NATs are transcribed by RNA polymerases, RNA polymerases bind to the upstream region of a gene encoding a sense mRNA and synthesize the complementary mRNA, moving to the 3′-end of the gene. Similarly, RNA polymerases that are bound to the upstream region of a gene encoding an antisense mRNA move to the 3′-end of the gene. Then, RNA polymerases collide with each other in the overlapping region of the genes, thereby inhibiting the transcription.

frequency of the collisions of RNA polymerases. In the case of *cis*-NATs in a head-to-head manner, the 5′-ends of mRNAs are the start position for the transcription of mRNAs. Overlapping at the 5′-end would inhibit the initiation of transcription and decrease the expression level of the *cis*-NATs (Figure 3A). Contrary to *cis*-NATs in a head-to-head manner, overlapping at the 3′-end would not decrease the expression level significantly when the overlapping region is short (Figure 3B). In the case of *cis*-NATs in a full-overlap manner, both the 5′- and 3′-ends of a transcript are overlapped. This would decrease the expression level, even when the length of overlapping regions is short (Figure 3C). For nearby transcripts, highly expressed nearby transcripts in a head-to-head manner decreased in quantity when the distance between the nearby transcripts was <50 bp (Figure 3D). However, the level of highly expressed nearby transcripts in a tail-to-tail manner did not decrease (Figure 3E). These results would occur if the start or end positions of transcripts were close to each other. In the initiation of transcription, RNA polymerases bind to the start position of transcription of mRNAs and cover the region between 55 bp upstream and 20 bp downstream (−55 to 20) of the start position (KORZHEVA *et al.* 2000; LEE and YOUNG 2000; MURAKAMI *et al.* 2002). Therefore, these findings suggest that nearby transcripts in a head-to-head manner inhibited the binding of RNA polymerases to the upstream regions of the transcripts, when the distance between the nearby transcripts in a head-to-head manner was <50 bp. In addition, the model of transcriptional collisions for *cis*-NATs may explain an observation that experiments of Northern hybridization showed smear bands of mRNAs at the genomic regions where *cis*-NATs were located (KIYOSAWA *et al.* 2005). This implies that various lengths of single-stranded mRNAs may be produced by the inhibition of the transcription and unusual movements of RNA polymerases.

Among the 2462 groups of human *cis*-NATs, 874 groups did not include alternative forms of sense and antisense mRNAs, and among the 874 groups, 542 (62%) groups consisted of *cis*-NATs where a sense mRNA overlapped only the intron regions of the gene encoding the antisense mRNA in the genome. As shown in Figures 1 and 3, the expression level of *cis*-NATs overlapping the intron regions also decreased as the length of the overlapping region increased. As for the regulatory mechanisms of gene expression by *cis*-NATs overlapping the intron regions, a double-stranded RNA-dependent mechanism would be difficult to use in explaining the decrease of the expression level of the *cis*-NATs, because they cannot form double strands of mRNAs after transcription and pre-mRNA splicing of mRNAs. Although double strands of mRNAs may be formed before pre-mRNA splicing after transcription, it is unclear whether it occurs. In the meantime, transcriptional collisions reasonably explain that *cis*-NATs overlapping the intron regions affected the expression of the *cis*-NATs.

Our observations are consistent with the transcriptional collision model. However, this does not mean that they exclude other regulatory mechanisms from the regulation of *cis*-NATs in gene expression. In addition to the regulation by general transcription factors in gene expression, *cis*-NATs employ several regulatory mechanisms, including transcriptional collisions (LAVORGNA *et al.* 2004). Our findings will be useful for the examination and understanding of the regulatory mechanisms of *cis*-NATs in gene expression and furthermore will help in elucidating the regulatory network of genes and their evolution.

We are grateful to members of the Laboratory for DNA Data Analysis at the National Institute of Genetics for discussion and comments on the manuscript. This work was supported by a research grant from the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency, and by a grant of the Genome Network Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

LITERATURE CITED

ADACHI, N., and M. R. LIEBER, 2002 Bidirectional gene organization: a common architectural feature of the human genome. Cell **109:** 807–809.

ALFANO, G., C. VITIELLO, C. CACCIOPPOLI, T. CARAMICO, A. CAROLA *et al.*, 2005 Natural antisense transcripts associated with genes involved in eye development. Hum. Mol. Genet. **14:** 913–923.

CARNINCI, P., T. KASUKAWA, S. KATAYAMA, J. GOUGH, M. C. FRITH *et al.*, 2005 The transcriptional landscape of the mammalian genome. Science **309:** 1559–1563.

CARTHEW, R. W., 2006 Gene regulation by microRNAs. Curr. Opin. Genet. Dev. **16:** 203–208.

CASTILLO-DAVIS, C. I., S. L. MEKHEDOV, D. L. HARTL, E. V. KOONIN and F. A. KONDRASHOV, 2002 Selection for short introns in highly expressed genes. Nat. Genet. **31:** 415–418.

CHEN, J., M. SUN, L. D. HURST, G. G. CARMICHAEL and J. D. ROWLEY, 2005a Genome-wide analysis of coordinate expression and evolution of human *cis*-encoded sense-antisense transcripts. Trends Genet. **21:** 326–329.

CHEN, J., M. SUN, L. D. HURST, G. G. CARMICHAEL and J. D. ROWLEY, 2005b Human antisense genes have unusually short introns: evidence for selection for rapid transcription. Trends Genet. **21:** 203–207.

CHENG, J., P. KAPRANOV, J. DRENKOW, S. DIKE, S. BRUBAKER *et al.*, 2005 Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science **308:** 1149–1154.

COUDERT, A. E., L. PIBOUIN, B. VI-FANE, B. L. THOMAS, M. MACDOUGALL *et al.*, 2005 Expression and regulation of the Msx1 natural antisense transcript during development. Nucleic Acids Res. **33:** 5208–5218.

CRAMPTON, N., W. A. BONASS, J. KIRKHAM, C. RIVETTI and N. H. THOMSON, 2006 Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res. **34:** 5416–5425.

ENCODE PROJECT CONSORTIUM, 2004 The ENCODE (ENCyclopedia Of DNA Elements) Project. Science **306:** 636–640.

HUBBARD, T., D. ANDREWS, M. CACCAMO, G. CAMERON, Y. CHEN *et al.*, 2005 Ensembl 2005. Nucleic Acids Res. **33:** D447–D453.

KATAYAMA, S., Y. TOMARU, T. KASUKAWA, K. WAKI, M. NAKANISHI *et al.*, 2005 Antisense transcription in the mammalian transcriptome. Science **309:** 1564–1566.

KENT, W. J., 2002 BLAT: the BLAST-like alignment tool. Genome Res. **12:** 656–664.

KIYOSAWA, H., I. YAMANAKA, N. OSATO, S. KONDO and Y. HAYASHIZAKI, 2003 Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res. **13:** 1324–1334.

KIYOSAWA, H., N. MISE, S. IWASE, Y. HAYASHIZAKI and K. ABE, 2005 Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res. **15:** 463–474.

KORNEEV, S., and M. O'SHEA, 2005 Natural antisense RNAs in the nervous system. Rev. Neurosci. **16:** 213–222.

KORNEEV, S. A., J. H. PARK and M. O'SHEA, 1999 Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. J. Neurosci. **19:** 7711–7720.

KORZHEVA, N., A. MUSTAEV, M. KOZLOV, A. MALHOTRA, V. NIKIFOROV *et al.*, 2000 A structural model of transcription elongation. Science **289:** 619–625.

KOYANAGI, K. O., M. HAGIWARA, T. ITOH, T. GOJOBORI and T. IMANISHI, 2005 Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene **353:** 169–176.

KREK, A., D. GRUN, M. N. POY, R. WOLF, L. ROSENBERG *et al.*, 2005 Combinatorial microRNA target predictions. Nat. Genet. **37:** 495–500.

KUMAR, M., and G. G. CARMICHAEL, 1997 Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc. Natl. Acad. Sci. USA **94:** 3542–3547.

KUMAR, M., and G. G. CARMICHAEL, 1998 Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. Microbiol. Mol. Biol. Rev. **62:** 1415–1434.

LAI, E. C., 2002 Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. **30:** 363–364.

LAPIDOT, M., and Y. PILPEL, 2006 Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. **7:** 1216–1222.

LASH, A. E., C. M. TOLSTOSHEV, L. WAGNER, G. D. SCHULER, R. L. STRAUSBERG *et al.*, 2000 SAGEmap: a public gene expression resource. Genome Res. **10:** 1051–1060.

LAVORGNA, G., D. DAHARY, B. LEHNER, R. SOREK, C. M. SANDERSON *et al.*, 2004 In search of antisense. Trends Biochem. Sci. **29:** 88–94.

LEE, J. T., L. S. DAVIDOW and D. WARSHAWSKY, 1999 Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. **21:** 400–404.

LEE, T. I., and R. A. YOUNG, 2000 Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. **34:** 77–137.

LEHNER, B., G. WILLIAMS, R. D. CAMPBELL and C. M. SANDERSON, 2002 Antisense transcripts in the human genome. Trends Genet. **18:** 63–65.

LEVINE, M., and E. H. DAVIDSON, 2005 Gene regulatory networks for development. Proc. Natl. Acad. Sci. USA **102:** 4936–4942.

LEWIS, B. P., C. B. BURGE and D. P. BARTEL, 2005 Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell **120:** 15–20.

LIAO, B. Y., and J. ZHANG, 2006 Evolutionary conservation of expression profiles between human and mouse orthologous genes. Mol. Biol. Evol. **23:** 530–540.

MAKALOWSKA, I., C. F. LIN and W. MAKALOWSKI, 2005 Overlapping genes in vertebrate genomes. Comput. Biol. Chem. **29:** 1–12.

MARGULIES, E. H., S. L. KARDIA and J. W. INNIS, 2001 Identification and prevention of a GC content bias in SAGE libraries. Nucleic Acids Res. **29:** E60.

MATTICK, J. S., 2001 Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. **2:** 986–991.

MATTICK, J. S., 2004 RNA regulation: A new genetics? Nat. Rev. Genet. **5:** 316–323.

MATTICK, J. S., and I. V. MAKUNIN, 2006 Non-coding RNA. Hum. Mol. Genet. **15** (Spec. no. 1): R17–R29.

MOORE, T., M. CONSTANCIA, M. ZUBAIR, B. BAILLEUL, R. FEIL *et al.*, 1997 Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse Igf2. Proc. Natl. Acad. Sci. USA **94:** 12509–12514.

MUNROE, S. H., and M. A. LAZAR, 1991 Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J. Biol. Chem. **266:** 22083–22086.

MURAKAMI, K. S., S. MASUDA, E. A. CAMPBELL, O. MUZZIN and S. A. DARST, 2002 Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science **296:** 1285–1290.

PANG, K. C., S. STEPHEN, P. G. ENGSTROM, K. TAJUL-ARIFIN, W. CHEN *et al.*, 2005 RNAdb: a comprehensive mammalian noncoding RNA database. Nucleic Acids Res. **33:** D125–D130.

PEARSON, W. R., and D. J. LIPMAN, 1988 Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:** 2444–2448.

PETERSON, J. A., and A. M. MYERS, 1993 Functional analysis of mRNA 3′ end formation signals in the convergent and overlapping transcription units of the *S. cerevisiae* genes RHO1 and MRP2. Nucleic Acids Res. **21:** 5500–5508.

POTTER, S. S., and W. W. BRANFORD, 1998 Evolutionary conservation and tissue-specific processing of Hoxa 11 antisense transcripts. Mamm. Genome **9:** 799–806.

PRESCOTT, E. M., and N. J. PROUDFOOT, 2002 Transcriptional collision between convergent genes in budding yeast. Proc. Natl. Acad. Sci. USA **99:** 8796–8801.

PUIG, S., J. E. PEREZ-ORTIN and E. MATALLANA, 1999 Transcriptional and structural study of a region of two convergent overlapping yeast genes. Curr. Microbiol. **39:** 369–373.

REIK, W., and J. WALTER, 2001 Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. **2:** 21–32.

SHENDURE, J., and G. M. CHURCH, 2002 Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol. **3:** RESEARCH0044.

STARK, A., J. BRENNECKE, N. BUSHATI, R. B. RUSSELL and S. M. COHEN, 2005 Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. Cell **123:** 1133–1146.

TUFARELLI, C., J. A. STANLEY, D. GARRICK, J. A. SHARPE, H. AYYUB *et al.*, 2003 Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat. Genet. **34:** 157–165.

VANHEE-BROSSOLLET, C., and C. VAQUERO, 1998 Do natural antisense transcripts make sense in eukaryotes? Gene **211:** 1–9.

VEERAMACHANENI, V., W. MAKALOWSKI, M. GALDZICKI, R. SOOD and I. MAKALOWSKA, 2004 Mammalian overlapping genes: the comparative perspective. Genome Res. **14:** 280–286.

VELCULESCU, V. E., L. ZHANG, B. VOGELSTEIN and K. W. KINZLER, 1995 Serial analysis of gene expression. Science **270:** 484–487.

WAGNER, E. G., and K. FLARDH, 2002 Antisense RNAs everywhere? Trends Genet. **18:** 223–226.

WAGNER, E. G., and R. W. SIMONS, 1994 Antisense RNA control in bacteria, phages, and plasmids. Annu. Rev. Microbiol. **48:** 713–742.

WUTZ, A., O. W. SMRZKA, N. SCHWEIFER, K. SCHELLANDER, E. F. WAGNER *et al.*, 1997 Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature **389:** 745–749.

WYRICK, J. J., and R. A. YOUNG, 2002 Deciphering gene expression regulatory networks. Curr. Opin. Genet. Dev. **12:** 130–136.

YELIN, R., D. DAHARY, R. SOREK, E. Y. LEVANON, O. GOLDSTEIN *et al.*, 2003 Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. **21:** 379–386.

Communicating editor: S. YOKOYAMA

Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 364 (2007) 78–85
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay

Nao Nishida [a,*], Tetsuya Tanabe [b], Miwa Takasu [a], Akira Suyama [c], Katsushi Tokunaga [a]

[a] *Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b] *Bio Business Division, Olympus Corporation, Hachioji, Tokyo 192-8512, Japan*
[c] *Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan*

Received 21 December 2006
Available online 13 February 2007

Abstract

A number of single nucleotide polymorphisms (SNPs) are considered to be candidate susceptibility or resistance genetic factors for multifactorial disease. Genome-wide searches for disease susceptibility regions followed by high-resolution mapping of primary genes require cost-effective and highly reliable technology. To accomplish successful and low-cost typing for candidate SNPs, new technologies must be developed. We previously reported a multiplex SNP typing method, designated the DigiTag assay, that has the potential to analyze nearly any SNP with high accuracy and reproducibility. However, the DigiTag assay requires multiple washing steps in manipulation and uses genotyping probes modified with biotin for each target SNP. Here we describe the next version of the assay, DigiTag2, which works with simple protocols and uses unmodified genotyping probes. We investigated the feasibility of the DigiTag2 assay by genotyping 96 target SNPs spanning a 610-kb region of human chromosome 5. The DigiTag2 assay is suitable for genotyping an intermediate number of SNPs (tens to hundreds of sites) with a high conversion rate (> 90%), high accuracy, and low cost.

Keywords: Multiplex genotyping; SNPs; Mutation; Oligonucleotide ligation assay

As a consequence of the Human Genome Project and single nucleotide polymorphism (SNP)[1] discovery projects, several million SNPs have been uploaded onto public SNP databases. It is estimated that there are 5 million SNPs with a greater than 10% minor allele frequency and 11 million SNPs with a greater than 1% minor allele frequency in the human genome [1]. Among these SNPs, many are candidate susceptibility or resistance genetic factors for multifactorial diseases and have been identified based on linkage analysis in families or association analysis with unrelated patients (cases) and healthy controls [2–6]. Large-scale case–control analyses using a dense set of SNP markers across the human genome have revealed associations between various diseases and SNPs with the highest detection power [7–9].

During recent years, genome-wide association studies using SNP markers have attempted to search for susceptibility and/or resistance genes by using emerging genome-wide SNP typing technologies such as Affymetrix GeneChip arrays and Illumina BeadArray genotyping technology [10–13]. These genome-wide SNP typing technologies would detect candidate regions, including susceptibility or resistance genes. However, to identify primary SNPs or genes, it is necessary to perform association analysis using an intermediate number of SNPs (tens to hundreds of sites) located within the candidate regions. Currently, there are a variety of SNP genotyping methods that are suitable for genotyping large numbers of samples for a modest number of SNPs such as 5′ exonuclease

* Corresponding author. Fax: +81 3 5802 8619.
E-mail address: nishida-75@umin.ac.jp (N. Nishida).

[1] *Abbreviations used:* SNP, single nucleotide polymorphism; MALDI–TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; ED, end digit; D1, first digit; PCR, polymerase chain reaction; dNTP, deoxynucleoside triphosphate; ATP, adenosine triphosphate; DTT, dithiothreitol; NAD, nicotinomide adenosine dinucleotide; EDTA, ethylenediaminetetraacetic acid; Cy3–ED-1, Cy3-labeled ED-1; Cy5–ED-2, Cy5-labeled ED-2; SDS, sodium dodecyl sulfate; DCN, DNA coded number.

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.02.005

fluorescence-based assay (TaqMan) [14], pyrosequencing [15], single-base extension [16], matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI–TOF MS) [17,18], and SNPlex assay [19]. However, many applications need to select relevant SNPs for their assay by in silico assay design, and some candidate SNPs are then excluded from investigation. Moreover, it is difficult or impossible for some assays to perform multiplex SNP genotyping.

To accomplish successful SNP typing for all candidate SNPs at low cost, new technologies must be developed. We previously reported a multiplex SNP typing method, designated the DigiTag assay, that has a high conversion rate (>90%) and reliable accuracy [20]. However, the DigiTag assay requires improvement with regard to simplifying assay protocols and reducing assay cost. In this study, we developed the DigiTag2 assay, which has simplified assay protocols, and performed typing for 96 SNP sites located in a 610-kb region on human chromosome 5 using 48 individual genomic DNA samples.

Materials and methods

DNA samples

Genomic DNA samples from 48 unrelated healthy donors were obtained from the Japan Health Science Foundation (Osaka, Japan). All donors provided written informed consent, and samples were anonymized. For each sample, 1 μg of purified genomic DNA was dissolved in 20 μl of TE buffer (pH 8.0, Wako, Osaka, Japan) for use and was stored at −20 °C.

End digits and first digits

We designed the end digits (EDs) and first digits (D1s) to be 23-mer oligonucleotides and attached the EDs and D1s to 5′ query probes and 3′ query probes, respectively. We prepared two EDs (ED-1 and ED-2) for two alleles at each SNP. All EDs and D1s are used for the priming site in the labeling step, and D1s are also used as probes that are attached to DNA microarray in the detection step. The EDs and D1s have the following properties: (i) uniform melting temperature (58.8 ± 1.0 °C) and length, (ii) specific hybridization only to complementary EDs and D1s, (iii) minimal interaction with other EDs and D1s, and (iv) no formation of secondary structures [21]. These properties ensure uniform polymerase chain reaction (PCR) efficiency, even if all of the EDs and D1s are used in multiplex PCR. Furthermore, precise hybridization on DNA microarray is possible using a set of D1s with high reproducibility. Sequence information for EDs and D1s is listed in Supplementary Table 1.

Multiplex PCR from sample DNA

We designed multiplex PCR primers for each of the 96 SNP sites to have relatively long length (average length 40-mer) and to give PCR products of between 181 and 798 bp (average length 527 bp). Sequence information for the multiplex PCR primers is listed in Supplementary Table 2.

We performed multiplex PCR using a two-step protocol (denature and extension steps) with a 6-min extension step using specifically designed primer pairs. Multiplex PCR was performed with 2.5 μl genomic DNA and 250 fmol of each primer for 96 SNP sites in 10 μl of 2× Qiagen Multiplex PCR Master Mix containing HotStarTaq DNA polymerase, multiplex PCR buffer and deoxynucleoside triphosphate (dNTP) mix (Qiagen Multiplex PCR Kit, Qiagen, Valencia, CA, USA). Cycling was performed using a Bio-Rad PTC-200 Peltier thermal cycler (Bio-Rad, Hercules, CA, USA) as follows: 95 °C for 15 min, followed by 40 cycles of 95 °C for 30 s and 68 °C for 6 min. When necessary, fragment length of the 96 PCR products was confirmed by capillary electrophoresis (Agilent 2100 Bioanalyzer, Agilent, Palo Alto, CA, USA) to evaluate PCR efficiency.

Encoding reaction

We performed multiplex oligonucleotide ligation assay using the multiplex PCR products as targets. For 96-plex oligonucleotide ligation assay, we prepared mismatch-induced 5′ query probes for 91 target SNPs and perfect match 5′ query probes for 5 target SNPs (SNP 7, SNP 9, SNP 18, SNP 49, and SNP 93). The assignment of D1s to the SNPs analyzed in this study and sequence information for the probes are listed in Supplementary Table 3.

Prior to the encoding reaction, 96 unmodified 3′ query probes were simultaneously phosphorylated at the 5′ end in 40 μl of 1× protruding end kinase buffer containing 30 mM adenosine triphosphate (ATP), 40 U polynucleotide kinase, and 4 pmol of 3′ query probes for 96 SNP sites (Kination Kit, Toyobo, Osaka, Japan). The reaction mixture was incubated for 30 min at 37 °C and for 3 min at 95 °C using a Bio-Rad PTC-200 Peltier thermal cycler. The encoding reaction was prepared by mixing 1 μl of multiplex PCR products in 15 μl of *Taq* DNA ligase buffer containing 20 mM Tris–HCl (pH 7.6), 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM dithiothreitol (DTT), 1 mM nicotinomide adenosine dinucleotide (NAD), and 0.1% Triton X-100 (New England Biolabs, Beverly, MA, USA) with 10 fmol of probes (192 5′ query probes and 96 phosphorylated 3′ query probes) and 10 U *Taq* DNA ligase. All components of the encoding reaction were mixed on ice. The encoding reaction initially was held at 95 °C for 5 min, followed by 58 °C for 15 min using a Bio-Rad PTC-200 Peltier thermal cycler. The reaction was stopped by holding the temperature at 10 °C.

Labeling reaction

For the labeling reaction, 6 μl of ligation products was directly mixed in 12 μl of *Ex Taq* buffer containing 20 mM Tris–HCl (pH 8.0), 100 mM KCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, 0.5% Tween 20, 0.5%