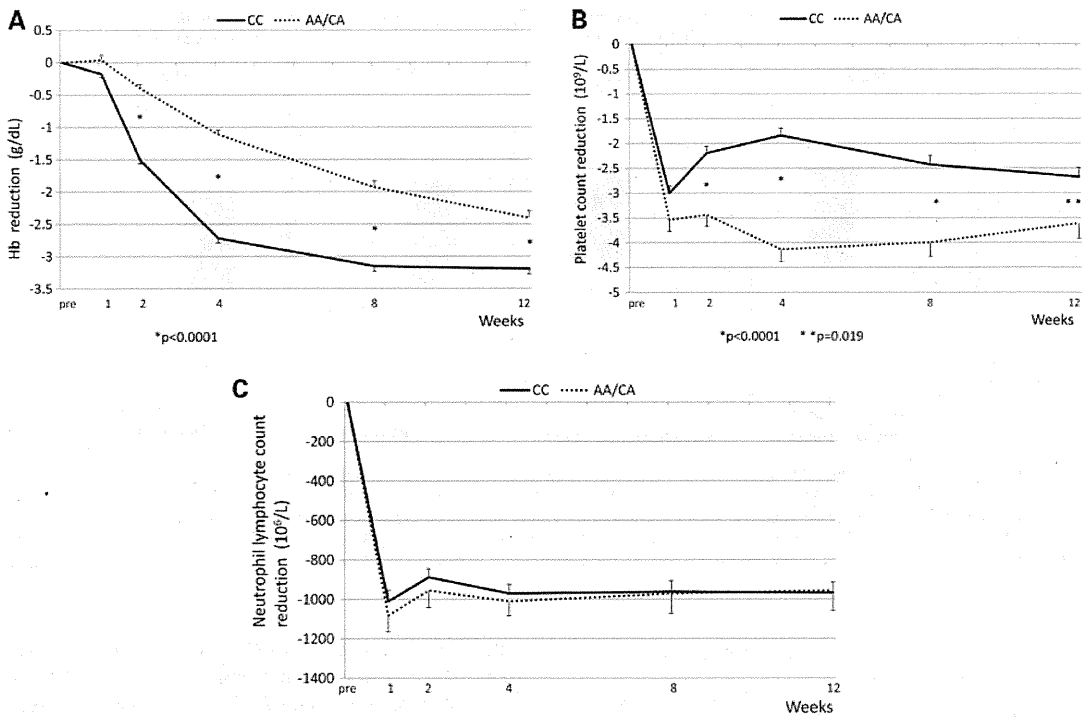

Figure 3. *ITPA* rs1127354 genotypes and the quantitative reduction of blood cells from baseline. Mean reduction of (**A**) Hb levels, (**B**) platelet counts and (**C**) neutrophil leukocyte counts during treatment according to rs1127354 genotype is shown. Solid and dotted lines indicate patients with CC and AA/CA genotypes, respectively. Error bars indicate standard error. CC genotype had more reduction in mean Hb levels during therapy compared with the AA/CA genotype ($^*P < 0.0001$ for weeks 2, 4, 8, 12). CC genotype had less of a reduction in mean platelet counts ($^*P < 0.0001$ for weeks 2, 4, 8, and $^{**}P = 0.019$ for week 12), and showed a reactive increase of platelet counts through weeks 1–4.

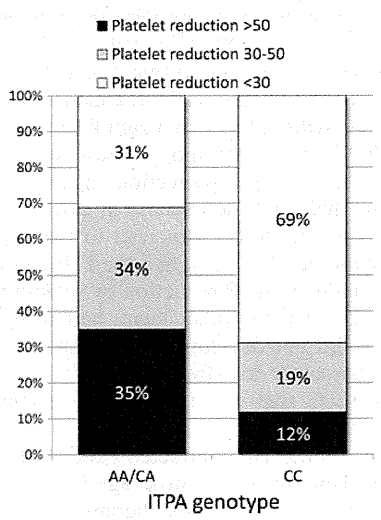

Figure 4. *ITPA* rs1127354 genotypes and reduction of platelet counts at week 4 of PEG-IFN/RBV therapy. The percentage of patients with platelet count reduction of >50 (10^9/l) (black bar), 30–50 (10^9/l) (gray bar) and <30 (10^9/l) (white bar) at week 4 is shown for rs1127354 genotypes. The incidence of platelet count reduction of >50 and <30 was significantly lower in patients with the rs1127354 genotypes CC compared with AA/CA genotypes: 12 versus 35%, $P < 0.0001$, and 69 versus 31%, $P < 0.0001$, respectively.

In this study, two SNPs, rs11697186 and rs6139030, which were within and around *DDRGK1* gene on chromosome 20, were strongly associated with thrombocytopenia as well as with Hb reduction at week 4. In clinical practice, the positive predictive value and negative predictive value by rs11697186 genotypes were 66.5 and 69.4% for thrombocytopenia, as well as 97.2 and 45% for RBV-induced anemia at week 4. As previously reported (22,26), a functional SNP (rs1127354) in the *ITPA* locus, which is in strong LD with rs11697186, was the most significant SNP associated with RBV-induced anemia and, in this study, IFN-induced thrombocytopenia in Japanese genetic populations. Note that severe Hb decline, which is mainly found in ITPA-CC patients, was inversely correlated with platelet reduction. This would contribute to an association between severe anemia and relative reactive increase of platelet count in this population, which attenuated the IFN effect on the platelet count. Our data supported a previous report which described that the current use of RBV, inducing severe anemia, might blunt the thrombocytopenic effect of IFNs as a result of reactive increase of platelet counts (27).

A previous paper showed hematological and bone marrow effects of RBV in rhesus monkeys (28). Hb values decreased significantly during RBV administration due to dose-related erythroid hypoplasia in bone marrow and returned to normal following withdrawal. On the other hand, increase of the platelet count occurred in both low- and high-dose treatment groups during RBV administration, with a fall of the platelet count to normal after drug withdrawal. The effect on platelet count was clearly dose related, with maximum counts rising to twice and three times above baseline levels in the low- and high-dose groups, respectively. This caused a significant increase of

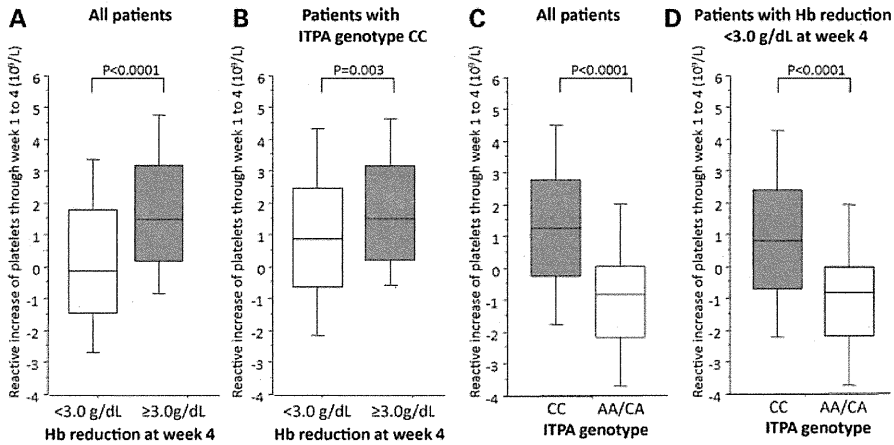

Figure 5. Reactive increase of platelet counts through weeks 1–4. Box plots of reactive increase of platelet count through weeks 1–4 according to the degree of anemia at week 4 are shown for all patients (**A**) and a subgroup of patients with the rs1127354 genotypes CC (**B**). Patients with anemia (Hb reduction ≥3.0 g/dl) at week 4 had a significantly higher degree of reactive increase of platelet count than those without anemia ($P < 0.0001$). Box plots of reactive increase of platelet counts according to the rs1127354 genotype CC are shown for all patients (**C**) and a subgroup of patients without anemia (**D**) (Hb reduction <3.0 g/dl) at week 4. Patients with the rs1127354 genotypes CC had a significantly high degree of reactive increase of platelet counts compared with those with genotypes AA/CA ($P < 0.0001$).

Table 4. Multivariate analysis of factors associated with reactive increase of platelets ≥20 (10^9/l) through weeks 1–4

	OR	95% CI	*P*-value
Baseline platelet counts	1.168	1.101–1.239	<0.0001
ITPA AA/CA	0.379	0.168–0.856	0.0196
Platelet reduction ≥30 (10^9/l) at week 4	0.051	0.021–0.120	<0.0001
Hb reduction ≥3.0 g/dl at week 4	1.602	0.914–2.809	0.0996

the platelet count associated with increased numbers of megakaryocytes. Additionally, the sequence homology of thrombopoietin (TPO) and erythropoietin (EPO) may explain the synergy of the physiologic role of TPO and EPO in platelet production. When EPO is elevated, as in iron deficiency anemia, an amino acid sequence similar to TPO may increase the platelet count (29).

Another possibility is a direct association between *ITPA* SNPs or the related SNPs with a strong LD and IFN-induced thrombocytopenia. *DDRGK1* (DDRGK domain-containing protein 1) is a novel C53/LZAP-interacting protein. C53/LZAP (also named as Cdk5rap3) is a putative tumor suppressor that plays important roles in multiple cell signaling pathways, including DNA damage response and NF-kappaB signaling (30); however, it remains largely unknown how the function of *DDRGK1* variants is regulated. Further studies are required to elucidate the possible association between *DDRGK1* variants and thrombocytopenia.

Multivariate analysis demonstrated that rs1127354 in the *ITPA* gene was independently associated with RBV-induced severe anemia and IFN-induced thrombocytopenia. This finding suggests that rs1127354 would be a useful marker to predict these hematological side effects by PEG-IFN/RBV therapy, indicating that genetic testing of *ITPA* variant might be applied to establish personalized dosages of PEG-IFN/RBV therapy. The rate of SVR tended to be higher in patients with *ITPA*-AA/CA genotype than those with *ITPA*-CC in this population. This might reflect decreased treatment efficacy (higher relapse rate) due to dose reduction of RBV in patients with *ITPA*-CC genotype. Our recent paper also demonstrated that the incidence of early dose reduction was significantly higher in *ITPA*-major (CC) patients as expected and, more importantly, that a significantly higher SVR rate was achieved in *ITPA*-hetero/minor (CA/AA) patients with HCV non-1b or low viral load strains (31) and in a subset of Japanese patients with the favorable TT genotype at rs8099917 of *IL28B* (32). Taken together, our results indicate that the *ITPA* minor variant A is not only a protective allele against PEG-IFN and RBV treatment-associated anemia in Japanese population, but also a significant predictor of SVR in certain HCV strains that show good response to IFN. The possible mechanism of protection against RBV-induced hemolysis is that ITP deficiency or low-activity variants (*ITPA* minor variant A) in turn lead to the accumulation of ITP in red blood cells (33,34), and the ITP confers protection against RBV-induced ATP reduction by substituting for erythrocyte GTP (25). On the other hand, half of the *ITPA*-major (CC) patients did not develop a significant Hb decline. This finding suggests other low-frequency *ITPA* variants or SNPs in other enzymes that are involved in erythrocyte purine nucleoside metabolism.

In Japan, the older HCV-infected patients developing liver fibrosis have been prevalent (mean age 62 years) (9). Thrombocytopenia by PEG-IFN/RBV therapy could lead to poor treatment efficiency among such Japanese patients with LC due to the initial or early dose reduction of PEG-IFN. In fact, ~40% of such population in this study had the initial dose reduction of PEG-IFN, resulting in a low SVR rate. Splenectomy or embolization of the splenic artery might be one of the options to increase the SVR rate, but a sufficient treatment outcome had not been obtained at present (35). Based on the recently accumulated SNP data, if patients had favorable *IL28B* genotype and *ITPA*-CC (lower reduction of platelet counts), a standard dose of PEG-IFN might be available for

the patients with lower platelet counts and the SVR rate might be increased due to sufficient dose of PEG-IFN.

Several STAT-C agents (specifically targeted antiviral therapies for hepatitis C) are being tested for clinical efficacy against hepatitis C (12,13,15,16). Most experts believe that when new drugs are approved to treat hepatitis C, they will be used in combination with PEG-IFN and RBV. Moreover, recent clinical trials, including NS3 protease inhibitors, have shown that PEG-IFN plus RBV would be necessary to achieve optimal treatment responses (12,13). Our present results may provide a valuable pharmacogenetic diagnostic tool for tailoring PEG-IFN and RBV dosing to minimize drug-induced adverse events and for further optimization of clinical anti-HCV chemotherapeutics.

MATERIALS AND METHODS

Patients

From April 2007 to April 2010, samples were obtained from 303 patients with chronic HCV (genotype 1) infection who were treated at 14 multi-center hospitals (liver units with hepatologists) throughout Japan. Each patient was treated with PEG-IFN-α2b (1.5 μg/kg body weight, subcutaneously once a week) or PEG-IFN-α2a (180 μg once a week) plus RBV (600–1000 mg daily according to body weight) for 48 weeks. Treatment duration was extended in some patients up to 72 weeks, according to the physicians' preferences. The dose of PEG-IFN or RBV was reduced according to the recommendations on the package inserts or the clinical conditions of the individual patients. EPO or other growth factors were not given. Written informed consent was obtained from each patient and the study protocol conformed to the ethics guidelines of the Declaration of Helsinki and was approved by the institutional ethics review committees. HBsAg-positive and/or anti-HIV-positive patients were excluded from this study.

In the following stage of replication study, SNP genotyping in an independent set of 391 Japanese HCV patients treated with PEG-IFN plus RBV treatment was completed using the DigiTag2 or TaqMan assay (ABI) following the manufacturer's protocol. The characteristics of patients for each GWAS stage and replication stage are summarized in Table 1.

SNP genotyping and data cleaning

In the GWAS stage, we genotyped 303 Japanese HCV patients with and without the decrease of platelet counts from baseline to week 4 of PEG-IFN/RBV treatment [107 patients with a decrease of >30 (10^9/l) in platelet counts and 196 patients without a decrease of >30 (10^9/l) in platelet counts], using the Affymetrix Genome-Wide Human SNP Array 6.0 according to the manufacturer's instructions. The cut-off value was calculated to maximize the difference, which was also close to the median change. The average overall call rate of patients with and without the decrease of PLT reached 98.69 and 98.72%, respectively. We then applied the following thresholds for SNP QC in data cleaning: SNP call rate $\geq$95% for all samples, MAF $\geq$1% for all samples. A total of 595 052 SNPs on autosomal chromosomes passed the QC filters and were used for association analysis. All cluster plots of SNPs showing $P < 0.0001$ in association analyses by comparing allele frequencies in both groups with and without the decrease of PLT were checked by visual inspection, and SNPs with ambiguous genotype calls were excluded.

In the following stage of the replication study and high-density association mapping, we selected 23 tag SNPs from the 44.7 kb region, including *DDRGK1* gene and *ITPA* gene by analyzing LD and haplotype structure based on the HapMap data of Japanese, using the Haploview software. Of these tag SNPs, rs1127354 within the *ITPA* gene, which was associated with RBV-induced anemia (22), was included; however, rs7270101 was excluded because recent papers studying Japanese patients showed no variants in rs7270101 (26,31,32). The SNP genotyping in an independent set of 391 Japanese HCV patients with and without quantitative change in PLT levels from baseline to week 4 of PEG-IFN/RBV treatment (175 patients with quantitative change in PLT and 216 patients without quantitative change in PLT) was completed using the DigiTag2 assay (36). Twenty-two of the 23 SNPs were successfully analyzed and were used for SNP genotyping and data cleaning. All 22 SNPs in the replication study cleared HWE P-value >0.001.

Based on the above SNPs data obtained from 303 Japanese HCV patients, using the Affymetrix Genome-Wide Human SNP Array 6.0, we also performed GWAS between 94 patients with a quantitative change of >3 g of reduction in Hb and 209 patients without quantitative change in Hb levels from baseline to week 4 of PEG-IFN/RBV treatment. SNP genotyping in an independent set of 391 Japanese HCV patients with and without quantitative change in Hb levels from baseline to week 4 of PEG-IFN/RBV treatment (137 patients with quantitative change in Hb and 254 patients without quantitative change in Hb) was also completed using the DigiTag2 assay (36). Twenty-two of the 23 SNPs were successfully analyzed and were used for SNP genotyping and data cleaning.

An application of the Cochrane–Armitage test on all the SNPs showed the genetic inflation factor $\lambda = 1.000$ for thrombocytopenia and $\lambda = 1.006$ for anemia in the GWAS stage (Supplementary Material, Figs S1 and S2). In addition, principal component analysis was performed in 303 samples for the GWAS stage together with the HapMap samples (CEU, YRI, CHB and JPT) (Supplementary Material, Fig. S3). These results implied that the effect of population stratification was negligible, except one sample, which was excluded from further analysis.

Laboratory and histological tests

Blood samples were obtained at baseline, 1, 2, 4, 8 and 12 weeks after the start of therapy and for hematologic tests after the start of therapy and for hematologic tests, blood chemistry and HCV-RNA. Genetic polymorphism in the *IL28B* gene (rs8099917) was determined using the ABI TaqMan assay (Applied Biosystems, Carlsbad, CA, USA). Fibrosis was evaluated on a scale of 0–4 according to the METAVIR scoring system. The SVR was defined as an undetectable HCV-RNA level by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems, CA, USA) or by Cobas Ampliprep/Cobas TaqMan assay (CAP/CTM) with a lower detection limit of

15 IU/ml (Roche Diagnostic Systems) 24 weeks after the completion of therapy.

Statistical analysis

The observed association between an SNP and the decrease of platelets/quantitative change in Hb levels with response to PEG-IFN plus RBV treatment was assessed by χ^2 test with a two-by-two contingency table in three genetic models: allele frequency model, dominant-effect model and recessive-effect model. SNPs on chromosome X were removed because gender was not matched between groups with and without the decrease of PLT and quantitative change in Hb levels. A total of 595 052 SNPs passed the quality control filters in the GWAS stage; therefore, significance levels after Bonferroni correction for multiple testing were $P = 8.40 \times 10^{-8}$ (0.05/595052) in the GWAS stage and $P = 2.27 \times 10^{-3}$ (0.05/22) in the replication stage.

The association between an SNP of the *ITPA* gene (rs1127354) and the incidence of platelet reduction at week 4 was analyzed by Fisher's exact test. The association between *ITPA* polymorphisms and the degree of reduction in platelet counts and Hb levels at each time point during therapy were analyzed by Mann–Whitney *U* test. Multivariable regression analysis was used to analyze the factors associated with *ITPA*, the rs1127354 genotype, factors associated with platelet count reductions and factors associated with the reactive increase in platelet counts. IBM-SPSS software v.15.0 (SPSS, Inc., Chicago, IL, USA) was used for these analyses.

Possible heterogeneity in allele frequencies at rs1127354 was assessed by Tarone's test. The association between the SNP and thrombocytopenia/anemia were analyzed by the Cochran–Mantel–Haenszel test. Both analyses were performed using the R (version 2.9.0) software (Supplementary Material, Table S3).

AUTHORS' CONTRIBUTIONS

Drafting of the paper, statistical analysis and approval of the final draft submitted: M.M.; drafting of the paper, statistical analysis, collecting samples and clinical data and approval of the final draft submitted: Y.T. and M.K.; statistical analysis and approval of the final draft submitted: N.N., M.S. and K.T.; collecting samples and clinical data and approval of the final draft submitted: K.M., N.S., N.E., H.Y., S.N., K.H., S.H., Y.I., E.T., S.M., M.H., Y.H., F.S., S.K. and N.I.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

This study is based on 14 multi-center hospitals throughout Japan: Hokkaido area (Hokkaido University Hospital), Kanto area (Saitama University Hospital; Konodai Hospital; Musashino Red Cross Hospital; Tokyo Medical and Dental University Hospital; Yamanashi University Hospital), Koshin area (Shinshu University Hospital; Kanazawa University Hospital), Tokai area (Nagoya City University Hospital), Kinki area (Kyoto Prefectural University of Medicine Hospital; Hyogo College of Medicine Hospital), Chugoku/Shikoku area (Ehime University Hospital; Kawasaki Medical College Hospital) and Kyushu area (National Nagasaki Medical Center). We thank Ms Yasuka Uehara-Shibata, Yuko Ogasawara-Hirano, Yoshimi Ishibashi, Natsumi Baba and Megumi Yamaoka-Sageshima (Tokyo University) for technical assistance. We also thank Dr Masaaki Korenaga (Kawasaki), Dr Akihiro Matsumoto (Shinshu), Dr Kayoko Naiki (Saitama), Dr Takeshi Nishimura (Kyoto), Dr Hirayuki Enomoto (Hyogo), Dr Minako Nakagawa (Tokyo Medical and Dental University) and Ochanomizu Liver Conference Study Group for collecting samples, and Dr Mamoru Watanabe (Tokyo Medical and Dental University) and Dr Moriichi Onji (Ehime University) for their advice throughout the study.

Conflict of Interest statement. Y.T., E.T. and S.K. are currently conducting research sponsored by Merck Sharp & Dohme, Corp. and Chugai Pharmaceutical Co. Ltd. The other co-authors have no conflict of interest.

FUNDING

This study was supported by a grant-in-aid from the Ministry of Health, Labour, and Welfare of Japan (H22-kannen-005), and the Ministry of Education, Culture, Sports, Science, and Technology.

REFERENCES

1. Global Burden of Hepatitis C Working Group (2004) Global burden of disease (GBD) for hepatitis C. *J. Clin. Phamacol.*, **44**, 20–29.
2. Shiratori, Y., Shiina, S., Imamura, M., Kato, N., Kanai, F., Okudaira, T., Teratani, T., Tohgo, G., Toda, N., Ohashi, M. *et al.* (1995) Characteristic difference of hepatocellular carcinoma between hepatitis B- and C-viral infection in Japan. *Hepatology*, **22**, 1027–1033.
3. Yoshida, H., Tateishi, R., Arakawa, Y., Sata, M., Fujiyama, S., Nishiguchi, S., Ishibashi, H., Yamada, G., Yokosuka, O., Shiratori, Y. *et al.* (2004) Benefit of interferon therapy in hepatocellular carcinoma prevention for individual patients with chronic hepatitis C. *Gut*, **53**, 425–430.
4. George, S.L., Bacon, B.R., Brunt, E.M., Mihindukulasuriya, K.L., Hoffmann, J. and Di Bisceglie, A.M. (2009) Clinical, virologic, histologic, and biochemical outcomes after successful HCV therapy: a 5-year follow-up of 150 patients. *Hepatology*, **49**, 729–738.
5. Fried, M.W., Shiffman, M.L., Reddy, K.R., Smith, C., Marinos, G., Goncales, F.L. Jr, Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D. *et al.* (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.*, **347**, 975–982.
6. Manns, M.P., McHutchison, J.G., Gordon, S.C., Rustgi, V.K., Shiffman, M., Reindollar, R., Goodman, Z.D., Koury, K., Ling, M. and Albrecht, J.K. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*, **358**, 958–965.
7. Hadziyannis, S.J., Sette, H. Jr, Morgan, T.R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H. Jr, Bernstein, D., Rizzetto, M. *et al.* (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.*, **140**, 346–355.
8. Hiramatsu, N., Oze, T., Tsuda, N., Kurashige, N., Koga, K., Toyama, T., Yasumaru, M., Kanto, T., Takehara, T., Kasahara, A. *et al.* (2006) Should aged patients with chronic hepatitis C be treated with interferon and ribavirin combination therapy? *Hepatol. Res.*, **35**, 185–189.
9. Iwasaki, Y., Ikeda, H., Araki, Y., Osawa, T., Kita, K., Ando, M., Shimoe, T., Takaguchi, K., Hashimoto, N., Kobatake, T. *et al.* (2006) Limitation of

combination therapy of interferon and ribavirin for older patients with chronic hepatitis C. *Hepatology*, **43**, 54–63.
10. Sezaki, H., Suzuki, F., Akuta, N., Yatsuji, H., Hosaka, T., Kobayashi, M., Suzuki, Y., Arase, Y., Ikeda, K., Miyakawa, Y. *et al.* (2009) An open pilot study exploring the efficacy of fluvastatin, pegylated interferon and ribavirin in patients with hepatitis C virus genotype 1b in high viral loads. *Intervirology*, **52**, 43–48.
11. Bruno, R., Sacchi, P., Maiocchi, L., Patruno, S. and Filice, G. (2006) Hepatotoxicity and antiretroviral therapy with protease inhibitors: a review. *Dig. Liver Dis.*, **38**, 363–373.
12. Hezode, C., Forestier, N., Dusheiko, G., Ferenci, P., Pol, S., Goeser, T., Bronowicki, J.P., Bourliere, M., Gharakhanian, S., Bengtsson, L. *et al.* (2009) Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N. Engl. J. Med.*, **360**, 1839–1850.
13. McHutchison, J.G., Everson, G.T., Gordon, S.C., Jacobson, I.M., Sulkowski, M., Kauffman, R., McNair, L., Alam, J. and Muir, A.J. (2009) Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N. Engl. J. Med.*, **360**, 1827–1838.
14. Suzuki, F., Akuta, N., Suzuki, Y., Sezaki, H., Yatsuji, H., Kawamura, Y., Hosaka, T., Kobayashi, M., Arase, Y., Ikeda, K. *et al.* (2009) Rapid loss of hepatitis C virus genotype 1b from serum in patients receiving a triple treatment with telaprevir (MP-424), pegylated interferon and ribavirin for 12 weeks. *Hepatol. Res.*, **39**, 1056–1063.
15. Sakamoto, N. and Watanabe, M. (2009) New therapeutic approaches to hepatitis C virus. *J. Gastroenterol.*, **44**, 643–649.
16. Afdhal, N.H., McHutchison, J.G., Zeuzem, S., Mangia, A., Pawlotsky, J.M., Murray, J.S., Shianna, K.V., Tanaka, Y., Thomas, D.L., Booth, D.R. *et al.* (2010) Hepatitis C pharmacogenetics: state of the art in 2010. *Hepatology*, **53**, 336–345.
17. Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S. *et al.* (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.*, **41**, 1105–1109.
18. Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J. *et al.* (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*, **461**, 399–401.
19. Suppiah, V., Moldovan, M., Ahlenstiel, G., Berg, T., Weltman, M., Abate, M.L., Bassendine, M., Spengler, U., Dore, G.J., Powell, E. *et al.* (2009) IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat. Genet.*, **41**, 1100–1104.
20. Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S.I., Alexander, G. *et al.* (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature*, **461**, 798–801.
21. Tanaka, Y., Nishida, N., Sugiyama, M., Tokunaga, K. and Mizokami, M. (2010) lambda-Interferons and the single nucleotide polymorphisms: a milestone to tailor-made therapy for chronic hepatitis C. *Hepatol. Res.*, **40**, 449–460.
22. Fellay, J., Thompson, A.J., Ge, D., Gumbs, C.E., Urban, T.J., Shianna, K.V., Little, L.D., Qiu, P., Bertelsen, A.H., Watson, M. *et al.* (2010) ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. *Nature*, **464**, 405–408.
23. Afdhal, N., McHutchison, J., Brown, R., Jacobson, I., Manns, M., Poordad, F., Weksler, B. and Esteban, R. (2008) Thrombocytopenia associated with chronic liver disease. *J. Hepatol.*, **48**, 1000–1007.
24. Wazny, L.D. and Ariano, R.E. (2000) Evaluation and management of drug-induced thrombocytopenia in the acutely ill patient. *Pharmacotherapy*, **20**, 292–307.
25. Hitomi, Y., Cirulli, E.T., Fellay, J., McHutchison, J.G., Thompson, A.J., Gumbs, C.E., Shianna, K.V., Urban, T.J. and Goldstein, D.B. (2011) Inosine triphosphate protects against ribavirin-induced adenosine triphosphate loss by adenylosuccinate synthase function. *Gastroenterology*, **140**, 1314–1321.
26. Ochi, H., Maekawa, T., Abe, H., Hayashida, Y., Nakano, R., Kubo, M., Tsunoda, T., Hayes, C.N., Kumada, H., Nakamura, Y. *et al.* (2010) ITPA polymorphism affects ribavirin-induced anemia and outcomes of therapy—a genome-wide study of Japanese HCV virus patients. *Gastroenterology*, **139**, 1190–1197.
27. Ong, J.P. and Younossi, Z.M. (2004) Managing the hematologic side effects of antiviral therapy for chronic hepatitis C: anemia, neutropenia, and thrombocytopenia. *Cleve. Clin. J. Med.*, **71** (Suppl. 3), S17–S21.
28. Canonico, P.G., Kastello, M.D., Cosgriff, T.M., Donovan, J.C., Ross, P.E., Spears, C.T. and Stephen, E.L. (1984) Hematological and bone marrow effects of ribavirin in rhesus monkeys. *Toxicol. Appl. Pharmacol.*, **74**, 163–172.
29. Akan, H., Guven, N., Aydogdu, I., Arat, M., Beksac, M. and Dalva, K. (2000) Thrombopoietic cytokines in patients with iron deficiency anemia with or without thrombocytosis. *Acta Haematol.*, **103**, 152–156.
30. Wu, J., Lei, G., Mei, M., Tang, Y. and Li, H. (2010) A novel C53/LZAP-interacting protein regulates stability of C53/LZAP and DDRGK domain-containing protein 1 (DDRGK1) and modulates NF-kappaB signaling. *J. Biol. Chem.*, **285**, 15126–15136.
31. Sakamoto, N., Tanaka, Y., Nakagawa, M., Yatsuhashi, H., Nishiguchi, S., Enomoto, N., Azuma, S., Nishimura-Sakurai, Y., Kakinuma, S., Nishida, N. *et al.* (2010) ITPA gene variant protects against anemia induced by pegylated interferon-alpha and ribavirin therapy for Japanese patients with chronic hepatitis C. *Hepatol. Res.*, **40**, 1063–1071.
32. Kurosaki, M., Tanaka, Y., Tanaka, K., Suzuki, Y., Hoshioka, Y., Tamaki, N., Kato, T. and Yasui, Y. (2011) Analysis of the correlations between genetic polymorphisms of the ITPA gene and hemolytic anemia or outcome after treatment with pegylated-interferon and ribavirin in genotype 1b chronic hepatitis C. *Antivir. Ther.*, in press.
33. Shipkova, M., Lorenz, K., Oellerich, M., Wieland, E. and von Ahsen, N. (2006) Measurement of erythrocyte inosine triphosphate pyrophosphohydrolase (ITPA) activity by HPLC and correlation of ITPA genotype-phenotype in a Caucasian population. *Clin. Chem.*, **52**, 240–247.
34. Fraser, J.H., Meyers, H., Henderson, J.F., Brox, L.W. and McCoy, E.E. (1975) Individual variation in inosine triphosphate accumulation in human erythrocytes. *Clin. Biochem.*, **8**, 353–364.
35. Kumada, H., Okanoue, T., Onji, M., Moriwaki, H., Izumi, N., Tanaka, E., Chayama, K., Sakisaka, S., Takehara, T., Oketani, M. *et al.* (2010) Guidelines for the treatment of chronic hepatitis and cirrhosis due to hepatitis C virus infection for the fiscal year 2008 in Japan. *Hepatol. Res.*, **40**, 8–13.
36. Nishida, N., Tanabe, T., Takasu, M., Suyama, A. and Tokunaga, K. (2007) Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. *Anal. Biochem.*, **364**, 78–85.

Genomics 98 (2011) 280–287

Contents lists available at ScienceDirect

Genomics

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

Genome-wide profiling of DNA methylation in human cancer cells

Katsumi Ogoshi [a,1], Shin-ichi Hashimoto [a,b,2], Yoichiro Nakatani [b,2], Wei Qu [b,2], Kenshiro Oshima [b,2], Katsushi Tokunaga [c], Sumio Sugano [d,3], Masahira Hattori [b,2], Shinichi Morishita [b,2], Kouji Matsushima [a,*,1]

[a] Department of Molecular Preventive Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561 Japan
[c] Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[d] Department of Medical Genome, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561 Japan

ARTICLE INFO

Article history:
Received 11 January 2011
Accepted 8 July 2011
Available online 29 July 2011

Keywords:
DNA methylation
Second generation sequencing technology
Differentially methylated regions
Colon cancer

ABSTRACT

Global changes in DNA methylation correlate with altered gene expression and genomic instability in cancer. We have developed a methylation-specific digital sequencing (MSDS) method that can assess DNA methylation on a genomic scale. MSDS is a simple, low-cost method that combines the use of methylation-sensitive restriction enzymes with second generation sequencing technology. DNA methylation in two colon cancer cell lines, HT29 and HCT116, was measured using MSDS. When methylation levels were compared between the two cell lines, many differentially methylated regions (DMRs) were identified in CpG island shore regions (located within 2 kb of a CpG island), gene body regions and intergenic regions. The number of DMRs in the vicinity of gene transcription start sites correlated with the level of expression of TACC1, CLDN1, and PLEKHC1 (FERMT2) genes, which have been linked to carcinogenesis. The MSDS method has the potential to provide novel insight into the functional complexity of the human genome.

1. Introduction

DNA methylation is an important component of the epigenetic regulation of gene expression in eukaryotic cells. Methylation of CpG dinucleotides within transcriptional regulatory sequences often results in reduced expression or silencing of adjacent genes [1–3]. Global hypomethylation has been associated with chromosomal instability, while hypermethylation of the promoters of certain tumor suppressor genes is an important event in malignant transformation [4–8].

There are three predominant methods for investigating DNA methylation. The first method is an antibody-based, affinity purification procedure that captures methylated DNA [9–11]. The second method is bisulfite sequencing, which converts unmethylated cytosines to uracils that are subsequently recognized as thymines [12,13]. The third method makes use of methylation-sensitive restriction enzymes, which only cleave DNA when their recognition sites are unmethylated [14]. These three approaches can be combined with electrophoresis, tag based sequencing [15], microarray analysis [9,11,16–21] or high-throughput sequencing [22–25].

High-throughput sequencing can report the sequence of millions of DNA fragments in parallel. This technology has revolutionized the analysis of genome-wide histone modifications [26] and the re-sequencing of the human genome [27], as well as the analysis of transcription start sites (TSSs) [28], full-length RNAs [29], small RNAs [30], single nucleotide polymorphisms (SNPs) [31], and transcription factor combination domains [32].

The development of new technology has driven new discoveries in the area of DNA methylation. Meissner et al. showed that alterations in histone modification can reduce DNA methylation [22]. Irizarry et al. demonstrated that most tissue-specific differentially methylated regions (T-DMRs) and cancer-specific DMRs (C-DMRs) localize to CpG island shore regions [21,33]. Moreover, it has been established that DNA methylation in both the promoter and the gene body can affect gene expression [11,16,34,35]. Evolutionarily conserved regions (ECRs) within the gene body have been reported to act as alternative promoters under the control of DNA methylation [36]. On the other hand, DNA methylation occurs not only on CpG dinucleotides but also in non-CG contexts (mCHG and mCHH, where H = A, C or T) in embryonic stem cells [37].

Each method of methylation detection has its advantages and limitations. The use of antibodies allows determination of the methyl cytosine density within a region of interest. The use of methylation-sensitive restriction enzymes is only useful for the analysis of

* Corresponding author. Fax: +81 3 5684 2297.
E-mail addresses: oisho@m.u-tokyo.ac.jp (K. Ogoshi), hashimot@m.u-tokyo.ac.jp (S. Hashimoto), nakatani@cb.k.u-tokyo.ac.jp (Y. Nakatani), quwei@cb.k.u-tokyo.ac.jp (W. Qu), oshima@cb.k.u-tokyo.ac.jp (K. Oshima), tokunaga@m.u-tokyo.ac.jp (K. Tokunaga), ssugano@k.u-tokyo.ac.jp (S. Sugano), hattori@k.u-tokyo.ac.jp (M. Hattori), moris@cb.k.u-tokyo.ac.jp (S. Morishita), koujim@m.u-tokyo.ac.jp (K. Matsushima).
[1] Fax: +81 3 5841 3393.
[2] Fax: +81 4 7136 3977.
[3] Fax: +81 4 7136 4080.

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.07.003

methylation of an enzyme's specific recognition sites. Analysis of the global DNA methylation profile of the genome at the single-base level requires bisulfite sequencing. Whole genome bisulfite sequencing has been accomplished in Arabidopsis and humans by using second generation sequencing technology [23,24,37]. However, the widespread application of this method is currently limited due to the vast quantity of data it generates and its huge cost. Here, we report a simple, low-cost method to detect global DNA methylation, which combines the use of methylation-sensitive restriction enzymes together with second generation sequencing technology.

2. Methods

2.1. Cell culture and sample preparation

Tumor samples examined in the present study comprised colon cancer cell lines HT29 and HCT116 obtained from the ATCC. Cells were cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum. Genomic DNA was isolated using a QIAamp DNA mini kit (QIAGEN) according to the manufacturer's protocol.

2.2. Generation of methylation-specific digital sequencing (MSDS) libraries

Genomic DNA (1 μg) was digested sequentially with the methylation-sensitive enzymes, SacII, EagI and BssHII (Table 1). DNA was first digested by SacII (NEB) at 37 °C for 3 h, then by EagI (NEB) at 37 °C for 3 h, and finally by BssHII (NEB) at 50 °C for 3 h. Digested samples were then purified by phenol-chloroform extraction and the fragments were ligated to biotinylated linkers containing synthetic restriction sites using T4 DNA ligase (Invitrogen), as follows: (BssHII-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATG-3′, BssHII-B, 5′-pho-CGCGCATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′, EagI-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATC-3′, EagI-B, 5′-pho-GGCCGATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′, SacII-A, 5′-bio-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATCCGC-3′, SacII-B, 5′-pho-GGATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′). The ligated DNA was then fragmented randomly using an ultrasonic device and 110–160 bp fragments were isolated by PAGE. Non-methylated ends were enriched using streptavidin beads. The ends of the DNA were repaired using the End-it DNA End repair kit (Epicentre) so that they were blunt-ended. Repaired fragments were then ligated to the following blunt-ended linkers using T4 DNA ligase: (blunt-linker-A, 5′-CTGCCCCGGGTTCCTCATTCTCT-3′, blunt-linker-B, 5′-AGAGAATGAGGAACCCGGGGCAGTT-3′). The ligated DNA was amplified by PCR using linker-specific primers (primer-A, 5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′, primer-B, 5′-CTGCCCCGGGTTCCTCATTCTCT-3′) and sequenced in 35 bp reads on the Life Technologies' SOLiD sequencer (Fig. 1A).

2.3. Data analysis

Sequencing data (30/35 bp per sequencing read) was analyzed using Corona Light software. Human genome sequence and mapping information (Mar. 2006, hg18) was downloaded from the University of California Santa Cruz Genome Bioinformatics Site and used to construct a virtual BssHII, EagI and SacII tag library on the basis of the genome sequence. Genes neighboring each restriction enzyme site were also identified in order to determine the effect of methylation on their expression.

Each enzyme site contributes two tags to the sequence library: a plus-strand library tag and a minus-strand library tag. The methylation level of an enzyme site can be inferred by either of the two sequence tag counts; we selected the tag that generated the greater number of counts. When the distance between two successive restriction enzyme sites is less than about 100 bp, the tags cannot be sequenced, because the tags are deleted by the PAGE selection. CpGs localized in close proximity to one another are often co-methylated [37], such that the level of methylation of one enzyme site can usually be inferred from the level of methylation of other sites localized within the same area, when the distance between sites is too small. In situations where the distance between these enzymes sites was <100 bp, the sites were considered together as a group, and we again selected the tag that generated the greatest number of counts within the group. If the tag sequence matched a site within a repeated sequence or at more than two sites within the whole human genome, the tag was excluded. Excluding such sequences, 86,897 sites were analyzed (Table 1).

2.4. Bisulfite sequencing

Bisulfite modification of genomic DNA was performed using the EpiTect Bisulfite Kit (Qiagen). Primers were designed using Methyl Primer Express software (Life Technologies). Bisulfite-treated DNA was amplified by PCR. The PCR products were cloned into the pCR2.1-TOPO Vector, then transformed into One Shot TOP10 Competent Cells (Invitrogen). At least 24 clones were sequenced using an ABI3730 Sequencer (Life Technologies). The data was analyzed using the Quantification Tool for Methylation Analysis (Riken Institute of Physical and Chemical Research).

2.5. Gene expression data

Gene expression data for the HT29 cells was generated as described previously [28], while for the HCT116 cells, this data was obtained using 5'SOLiD technology [28]. Analysis of gene expression data was based on the 5′ SOLiD method [28].

Table 1
The number of restriction sites for the methylation-sensitive enzymes BssHII, EagI and SacII, in silico data.

Enzyme	Recognition site		No. of total sites in human genome	No. of genes containing the recognition site within the promoter region	No. of CpG island containing the recognition site
BssHII	G^CGCGC	6 bp	72,899	8787	13,363
EagI	C^GGCCG	6 bp	90,190	9089	14,023
SacII	CCGC^GG	6 bp	66,312	9793	15,317
BssHII + EagI + SacII			229,401	13,978	21,164

Defined site data

	No. of defined sites within human genome	No. of genes containing the defined site within the promoter region	No. of CpG island containing the defined site	No. of CpG island shore containing the defined site
MSDS method	86,897	11,217	19,592	8522

Promoter region is defined as the region 500 bp either side of the TSS.

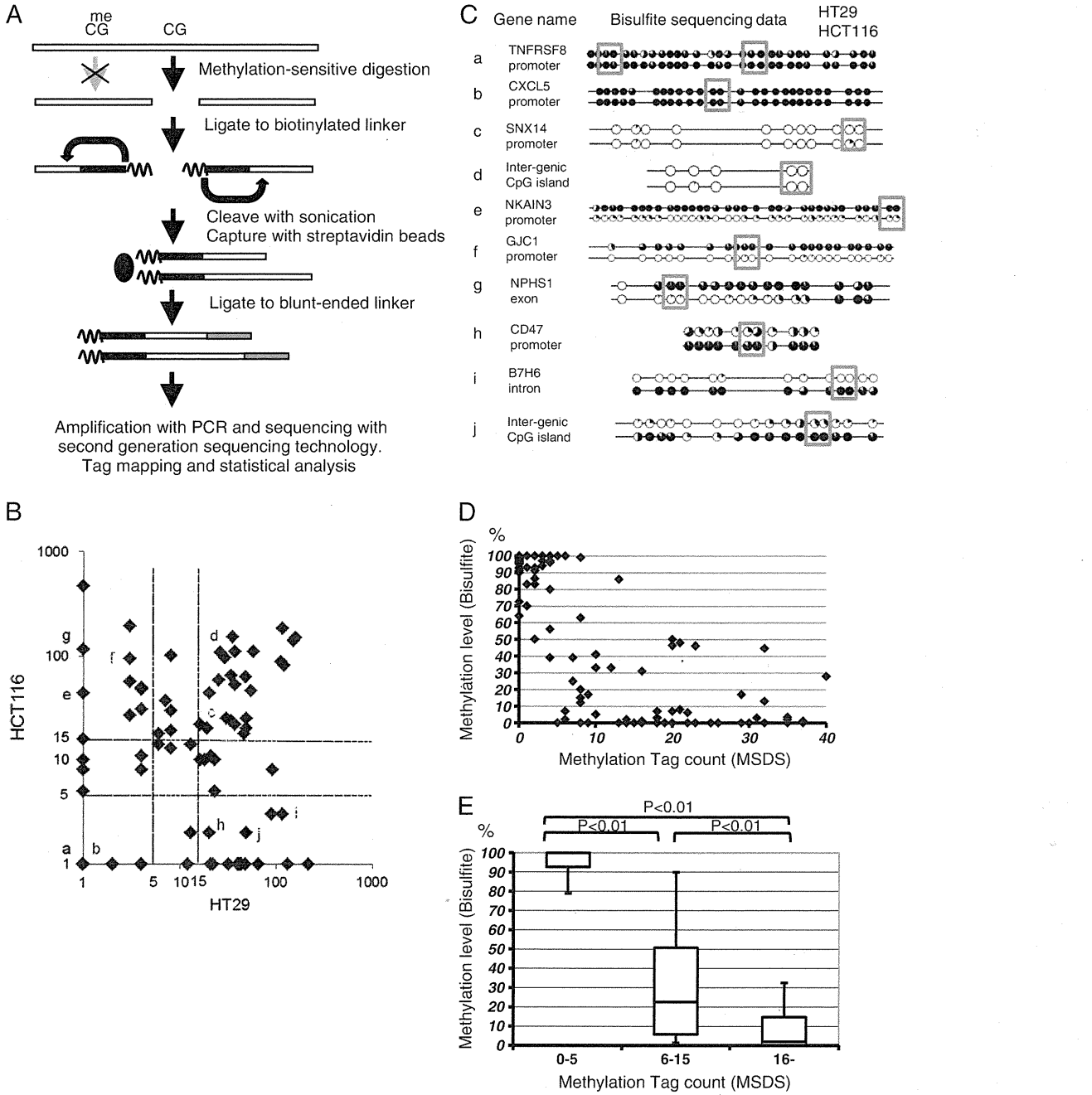

Fig. 1. Flow chart schematic of the methylation-specific digital sequencing (MSDS) method and the correlation between methylation level as assessed by bisulfite sequencing and the MSDS methylation tag count. A. Genomic DNA was digested with methylation-sensitive restriction enzymes and ligated to biotinylated linkers. The ligated sample was then cleaved by sonication and the fragments were captured with streptavidin beads. The captured DNA fragments were ligated to blunt-end linkers, amplified by PCR, and sequenced on the Life Technologies' SOLiD sequencer. B. Scatter diagram of MSDS tag counts corresponding to 100 regions selected at random that were also subjected to bisulfite sequencing. Sample labels a–d and e–j indicate sets of genes whose pattern of methylation was similar (a–d) or different (e–j) between HT29 and HCT116 cells. C. DNA methylation of genes a-j as analyzed by bisulfite sequencing. Gray squares represent CpG sites within restriction enzyme sites. Circles represent potential methylation sites (CpG) where the shading intensity indicates the frequency at which the site was found to be methylated amongst the clones analyzed (0–100%). a. TNFRSF8 (tumor necrosis factor receptor superfamily, member 8) gene promoter; b. CXCL5 (chemokine (C–X–C motif) ligand 5) gene promoter; c. SNX14 (sorting nexin 14) gene promoter; d. intergenic region within CpG island; e. NKAIN3 (Na^+/K^+ transporting ATPase interacting 3) gene; f. GJC1 (gap junction protein, gamma 1, 45kDa) gene promoter; g. NPHS1 (nephrosis 1, congenital, Finnish type (nephrin)) gene exon; h. CD47 (CD47 molecule) gene promoter; i. B7H6 (B7 homolog 6) gene intron; j. intergenic region within CpG island. D. and E. Correlation between methylation level as assessed by bisulfite sequencing and methylation tag count as assessed by MSDS. Horizontal bars represent methylation as determined by bisulfite sequencing. In E, boxes represent the quartiles and whiskers mark the 5th and 95th percentiles.

3. Results

3.1. MSDS analysis

We have developed a simple method, which we call MSDS, to detect global DNA methylation. The method makes use of three methylation-sensitive restriction enzymes (BssHII, EagI, and SacII) and second generation sequencing technology. There are two CpG dinucleotides within the recognition sequence of each enzyme (GCGCGC, CGGCCG, and CCGCGG, respectively). Short sequence tag fragments derived from the sites cleaved by these restriction enzymes can be mapped to unique sites within the genome. In silico

analyses indicate that there are 72,899, 90,190 and 66,312 recognition sites for BssHII, EagI and SacII within the human genome, respectively (Table 1). These restriction sites cover 64% (13,978) of unique gene promoters (within 500 bp of the TSS) and 75% (21,164) of the CpG islands within the genome. This combination of restriction enzymes allowed the most comprehensive coverage of CpG islands without using a greater number of enzymes. After excluding certain uninformative sites as described in Section 2.3, we analyzed a total of 86,897 sites. These defined sites cover 51% (11,217) of unique gene promoters (Supplemental Table 1), 69% (19,592) of CpG islands, and 8522 CpG island shores within the genome (Table 1).

Two colon cancer cell lines (HT29 and HCT116) were utilized to confirm the utility of this method. Using the SOLiD platform, 2,397,132 and 2,971,226 tags were mapped to the human genome in HT29 and HCT116 cells, respectively. After excluding uninformative sites as noted above, 1,493,950 tags (62.3% of total tags) in HT29 cells and 1,886,310 tags (63.5% of total tags) in HCT116 cells were matched to the remaining informative sites (Table 2, Supplementary Fig. 1). There were 35,309 (40.6%), and 48,141 (55.4%) zero tag count or un-hit sites in HT29 and HCT116 cells, respectively (Table 3).

3.2. Correlation between methylation level (bisulfite sequencing) and MSDS tag counts

To validate the accuracy of the MSDS method, MSDS tag counts were compared with the results of direct DNA methylation analysis, which were generated by bisulfite sequencing at more than 100 random CpG sites (Fig. 1B). In both cell lines, CpG sites that generated the highest number of tags corresponded to those that also exhibited a low level of methylation by bisulfite sequencing analysis, such as those within the SNX14 (sorting nexin 14) gene promoter. CpG sites that generated relatively fewer tags in both cell lines corresponded to those that were highly methylated in both cell lines according to the bisulfite sequence analysis, such as those within the TNFRSF8 (tumor necrosis factor receptor superfamily, member 8) gene promoter. Some sites appeared to be differentially methylated between the two cell lines. CpG sites that gave rise to many tags in HT29 cells alone were poorly methylated only in these cells, such as those within the promoter of the CD47 (CD47 molecule) gene. Similarly, CpG sites that gave rise to many tags in HCT116 cells alone were poorly methylated only in these cells, such as those within the promoter of the NKAIN3 (Na^+/K^+ transporting ATPase interacting 3) gene (Fig. 1C). Thus, the methylation levels determined by bisulfite sequencing correlated inversely with the MSDS tag count in both cell lines. Based on these results (Fig. 1D), we classified the genes into three arbitrary groups, based on the number of tag counts at each restriction enzyme site. Sites represented by ≥16 tags were considered to be poorly methylated (<20% methylated), whereas sites represented by ≤5 tags were considered to be highly methylated (>80% methylated), and sites represented by 6–15 tags were considered to be methylated at an intermediate level (21–79% methylated) (Fig. 1E).

Table 2
Summary of the results of sequence analysis for methylation tags, MSDS data.

Enzyme	No. of tags mapped to the human genome (% of total tags)		No. of defined tags	
	HT29	HCT116	HT29	HCT116
BssHII	648,117 (27.0%)	889,793 (29.9%)	415,643	563,851
EagI	1,077,327 (44.9%)	1,141,883 (38.4%)	660,181	740,139
SacII	671,688 (28.0%)	939,550 (31.6%)	418,126	582,320
BssHII + EagI + SacII	2,397,132	2,971,226	1,493,950	1,886,310

Table 3
Level of DNA methylation at CpG sites in HT29 and HCT116 cells.

Methylation level	No. of CpG sites			
	HT29	%	HCT116	%
Low	27,478	31.60%	25,240	29.00%
Middle	10,112	11.60%	5485	6.30%
High	49,307	56.70%	56,172	64.60%
0 tag count sites	35,309	40.63%	48,141	55.40%
Total CpG sites	86,897	100%	86,897	100%

Low level corresponds to 0–20% methylation, by bisulfite sequencing analysis, and corresponds to a tag count of greater than 16, whereas the high level corresponds to 80%~methylation and is correlated with tag counts between zero and five.

3.3. Genome-wide DNA methylation

The results of the genome-wide DNA methylation analysis are presented in Table 3. 56.7% (49,307) and 64.6% (56,172) of unique CpG sites were highly methylated, and 31.6% (27,478) and 29.0% (25,240) were poorly methylated in HT29 and HCT116 cells, respectively, indicating that a majority of CpG sites were either highly or poorly methylated, which is consistent with what has been reported previously [37,38]. As CpGs in close proximity are often co-methylated [37], we considered the tag with the greatest number of counts within such promoter regions as representative of the methylation level of those regions. Some gene promoters overlap CpG islands, whereas others do not. 48% of the promoters evaluated in both cell lines were poorly methylated (Fig. 2B-1); those that did not overlap CpG islands were methylated to a greater extent than those that contained a CpG island within the promoter region ($P<0.01$) (Figs. 2B-2, 3). Within gene body regions, 63% of enzyme sites were highly methylated in both cell lines (Figs. 2C-1) and the methylation level of CpG islands in such regions was lower than in other regions of the gene body ($P<0.01$) (Figs. 2C-2, 3). In promoter and gene body, the extent of methylation within CpG islands was lower than that in regions outside of the CpG islands ($P<0.01$). In addition, there was a difference in the extent of methylation between CpG islands within the promoter and CpG islands within the gene body. Methylation within CpG islands has previously been reported to be dependent upon the position of the island within the promoter region or within the gene body [36].

3.4. Comparison of MSDS methylation with gene expression

There is a strong relationship between gene expression and DNA methylation. Therefore, we analyzed gene expression in the HT29 and HCT116 cell lines and compared the results to the methylation data from the same cell lines. The genes were divided into three groups: high expression (≥100 copies), low expression (≤1 copy), and moderate expression (1<copies <100) [28]; and the level of expression of the genes within these groups was compared to the level of methylation. Highly and moderately expressed genes exhibited a low level of methylation within the promoter and a high level of methylation within the gene body. Genes with a low expression level exhibited a moderate level of methylation within both the gene promoter and the gene body. These results are consistent with those of previous reports [35,36] (Fig. 3).

3.5. Differentially methylated regions between HT29 and HCT116

Next, we compared the extent of methylation of the HT29 and HCT116 cells. Of the sites measured, 77% were methylated to a similar extent in the two cell lines: 52% of the sites measured were poorly methylated, and 23% of the sites measured were highly methylated (Figs. 2A-1). DMRs between the two cell lines were defined as enzyme sites that were highly methylated in one cell line and poorly methylated in the other cell line. Such sites made up 9% of the sites

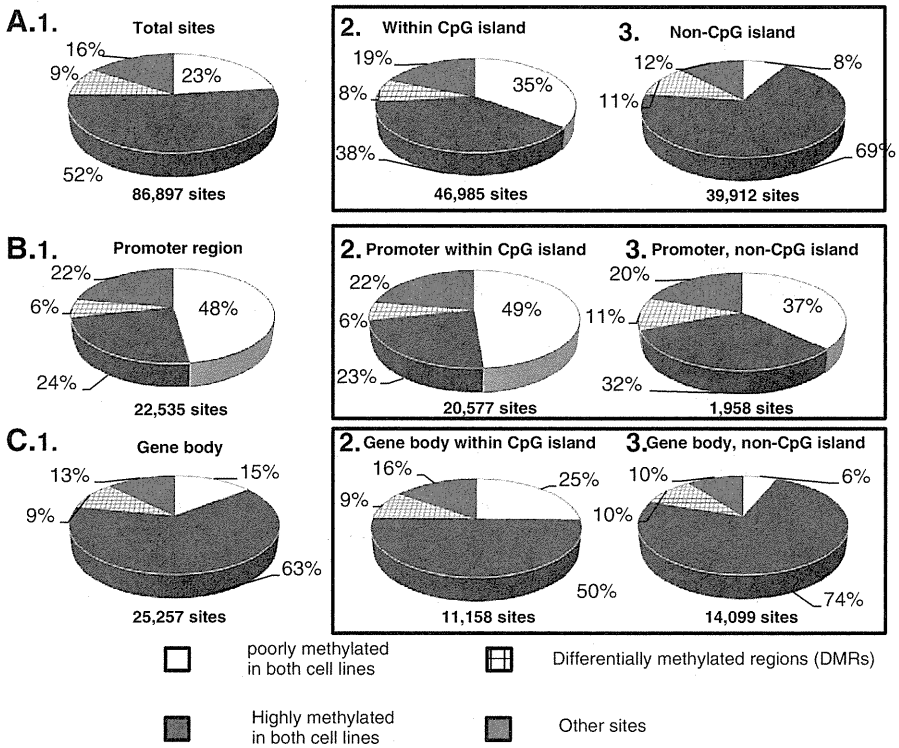

Fig. 2. DNA methylation in the vicinity of the TSSs and gene bodies. The number under each panel represents the total number of CpG sites for each pie chart. A. Pie charts represent the distribution of restriction enzyme sites, which are divided into four categories: poorly methylated in both cell lines, highly methylated in both cell lines, DMRs and other sites. B. The chart on the left represents methylation within 500 bp of the TSS. The charts on the right represent methylation within or outside of the CpG islands that are within 500 bp of the TSS. C. The chart on the left represents methylation within the gene body. The charts on the right represent methylation within or outside of CpG islands in the gene body.

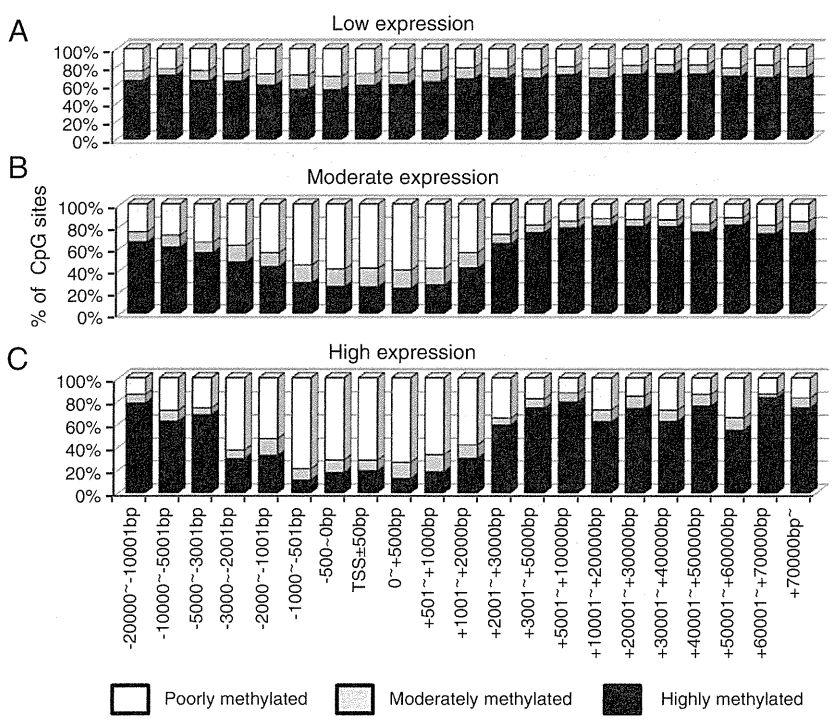

Fig. 3. Correlation of the extent of methylation assessed by MSDS relative to the distance from the TSS with the level of gene expression assessed by 5′-end SOLiD transcriptome analysis. Genes were categorized into three groups based on the 5′-end SOLiD transcriptome data: low expression, ≤1copy; high expression, ≥100 copy; moderate expression, 1<copies<100.

measured. Two percent of the sites measured were highly methylated in HT29 cells and poorly methylated in HCT116 cells, while 7% of the sites measured were highly methylated in HCT116 cells and poorly methylated in HT29 cells. As shown in Table 4, many DMRs were observed not in the gene promoter region, but instead in the gene body, intergenic or CpG island shore regions.

Finally, we investigated the correlation between the extent of methylation in the vicinity of the TSS and the expression of the associated gene. We evaluated the expression of several genes localized to DMRs within 1000 bp of the TSS and whose expression level differed at least five-fold between HT29 and HCT116 cells. Of these genes, we analyzed those linked to carcinogenesis. These included SMPD3 (sphingomyelin phosphodiesterase 3, neutral membrane (neutral sphingomyelinase II)), VASH1 (vasohibin 1), and TACC1 (transforming acidic coiled-coil containing protein 1), which are involved in the cell cycle; RIPK3 (receptor-interacting serine-threonine kinase 3), and TNFSF9 (tumor necrosis factor (ligand) superfamily, member 9), which are involved in apoptosis; and CLDN1 (claudin 1), NELL1 (NEL-like 1 (chicken)), ROBO1 (roundabout, axon guidance receptor, homolog 1 (Drosophila)), ENG (endoglin), PLEKHC1 (pleckstrin homology domain containing, family C) (FERMT2 (fermitin family member 2)), which are involved in cell adhesion. In these genes, the DMRs localize to several positions within the gene, in some cases only upstream of the TSS (e.g. VASH1 and TACC1); in some cases only downstream of the TSS (e.g. SMPD3, RIPL3, TNFSF9, CLDN1, NELL1 and ENG); and in some cases, both upstream and downstream of the TSS (e.g. ROBO1 and PLEKHC1 (FERMT2)) (Fig. 4). In TNFSF9, NELL1 and ENG genes, CpG sites closer to the TSS than the DMR were methylated to a similar level in both cell lines (Fig. 4). This result demonstrates that these DMRs are not localized solely in the vicinity of the TSS.

4. Discussion

Diverse methods for the measurement of DNA methylation have been developed [9,11,16–24]. Moreover, second generation sequencing technology has facilitated genome-wide analyses of DNA methylation. However, this technology generates too much data and is too expensive to be accessible to most researchers in the field. Therefore, we have developed a method for detecting global DNA methylation that makes use of methylation-sensitive restriction enzymes combined with second generation sequencing technology. The methylation level can be determined by monitoring the number of times a particular recognition site is sequenced within the genomic pool. This method optimizes data handling by limiting the data generated to only the restriction enzyme sites.

Table 4
Differentially methylated regions between the HT29 and HCT116 cell lines.

Gene position	No. of sites in each gene position	No. of differentially methylated sites	*P* value
Tags mapped to genome	86,897	7903	
CpG island	46,985	3682	1.67E−44
CpG island shore[a]	9895	1219	2.12E−32
Other region	30,017	3002	1.47E−11
Gene body			
−5000~2001 bp	2216	240	0.00399
−2000~1001 bp	1642	187	0.001
−1000~501 bp	2503	264	0.01
−500~1 bp	9178	605	1.16E−18
TSS ± 50 bp	3562	187	3.65E−16
1st exon	9168	651	2.21E−12
Intron	23,164	2111	0.908
Exon	4191	314	0.0002172
Last exon	2090	180	0.437
Intergenic	32,730	3348	1.55E−19

[a] CpG island shore; 2000 bp within CpG island.

Changes in DNA methylation profiles have been observed during normal development and tumorigenesis. However, differences in the genome-wide methylation pattern between related cancer cell lines derived from related tumors have not been examined previously. We found many sites, localized predominantly to CpG island shore and intergenic regions, which are differentially methylated in HT29 and HCT116 cells. Differences between the HT29 and HCT116 cell lines may have arisen in the process of transforming from normal tissue to cancer tissue, or may be due to differences in the cells before they became cancerous. Regions of the genome that are differentially methylated may provide insight into the variety of regions of the genome where differential gene expression is linked to the development of cancer. By searching regions of the genome that are differentially methylated between cell lines from similar tissue, it may be possible to reveal differences in the initiation of cancer and/or in the types of cells that cause the malignant transformation: from a basal cell or layer cell, for example.

The following genes have been reported to exhibit tumor suppressor activity and to be regulated by methylation within their promoter regions: SMPD3 [39], VASH1 [40], RIPK3 [41], TNFSF9 [42], NELL1 [43,44], ROBO1 [45–47], ENG [43,48]. In this study, we identified additional genes characterized by a link between DNA methylation and expression: TACC1, CLDN1, and PLEKHC1 (FERMT2). These genes may also be important in carcinogenesis.

Analysis of DNA methylation using microarrays suffers from a bias towards gene promoter regions and CpG islands. The microarray method can only provide relative data, and therefore requires a reference standard. Methods based on restriction enzymes possess inferior sensitivity to bisulfite sequencing, but do allow a quantitative comparison of methylation patterns. Bisulfite sequencing allows analysis of DNA methylation at the single nucleotide level, but at a very high cost for a genome the size of the human genome [37].

Established methods that make use of methylation-sensitive restriction enzymes include methylation-specific digital karyotyping (MSDK) [15], modified methylation-specific digital karyotyping (MMSDK) [25], methyl-sensitive cut counting (MSCC) [35] and others. The MMSDK method is similar to the MSDS method that we have developed. Both methods are based on digital karyotyping technology. The process of digestion of genomic DNA with a methylation-sensitive restriction enzyme, and the use of a fragmenting enzyme with which to generate short sequence tags is essential to both methods. The key difference between MSDS and MMSDK is the location of the short sequence read using second generation sequencing technology. The MMSDK method targets the nearest NlaIII sites to the methylation-sensitive enzymes, while the MSDS method targets the methylation-sensitive enzyme sites themselves. This difference simplifies the construct preparation process. It results in a shortening of the duration of the procedure and reduces loss of the sample. Furthermore, it requires less PCR cycles, and therefore minimizes PCR bias. During data analysis, the directly identifiable measurement positions make it easy to perform direct sequencing of the cleavage site. However, because direct sequencing is employed, when a restriction enzyme site exists within a repeated sequence, these tags are considered unreliable and excluded. Accordingly, MSDS method is not able to measure repeated sequences. The difference between MSCC and MSDS is the combination of methylation-sensitive enzymes used. MSCC uses the HpaII enzyme, while MSDS uses three enzymes (BssHII, EagI and SacII). Methods that make use of methylation-sensitive restriction enzymes are limited to profiling the methylation of these enzyme's recognition sites, suggesting that use of enzymes that cleave at a greater number of sites is advantageous. The HpaII and MspI enzymes are four base encoding enzymes, and the recognition site of each enzyme contains one CpG dinucleotide within an identical recognition sequence, CCGG. HpaII cleaves only at unmethylated sites, whereas MspI cleaves both methylated and unmethylated sites. In silico analyses reveal that there are approximately two million recognition sites for these enzymes

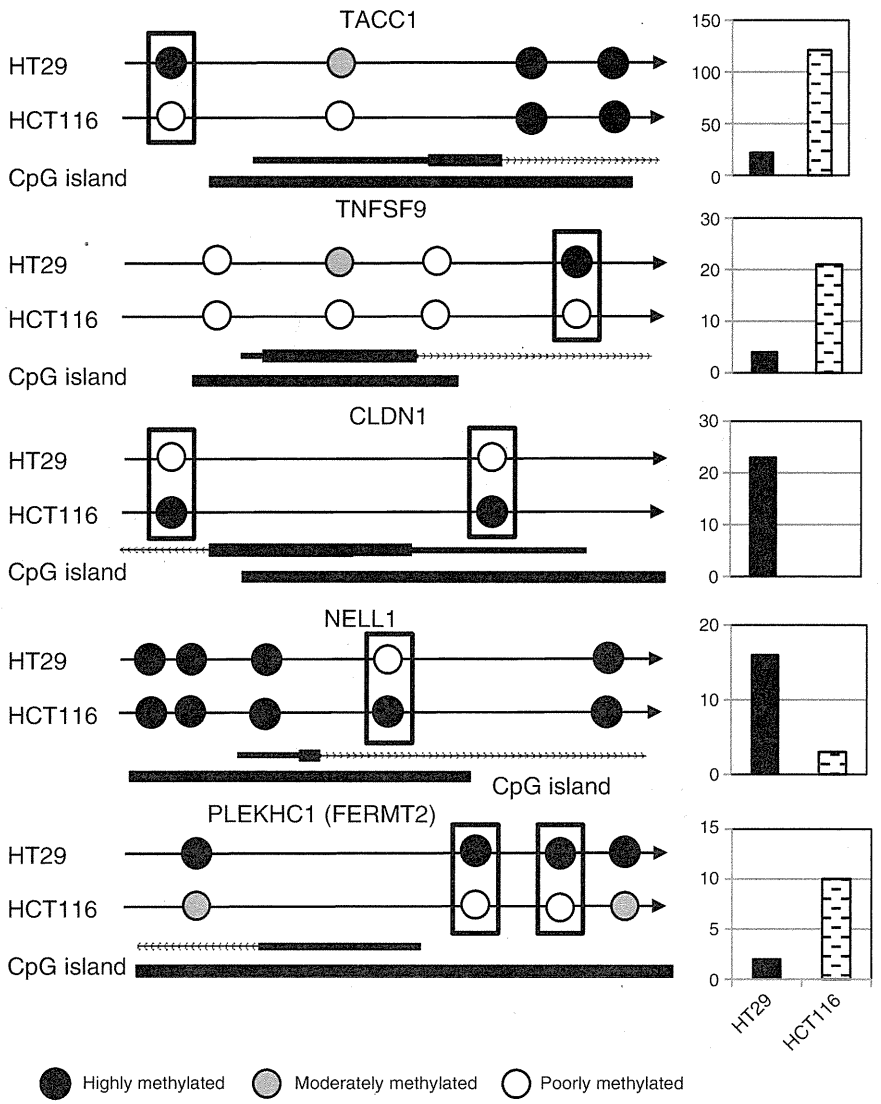

Fig. 4. Correlation between the extent of methylation within DMRs and the level of expression of the associated gene. We selected five genes containing DMRs in close proximity to the TSSs: TACC1, TNFSF9, CLDN1, NELL1, and PLEKHC1 (FERMT2). The panel to the left shows relative MSDS tag counts (circles) at the relevant sites within each gene in each cell line. A black box around the two circles represents the position of the DMRs that were selected. The panel on the right shows the relative expression of each gene in each cell line.

within the human genome. The MSDS method reported here makes data analysis simpler because it results in analysis of fewer sites and therefore fewer tags.

Within the data obtained via the MSDS methods, zero tag count or un-hit sites made up about 40–50% of the uninformative sites. These sites may be fully methylated or just false negative. The use of enzymes that recognize six-base target sequences obviates the need for a comparison target, as is the case when evaluating MspI against HpaII. To verify this, we performed bisulfite sequencing at 44 CpG sites that gave rise to a MSDS tag count of zero. We found that 41 of the 44 sites were highly methylated. This finding suggests that the bias due to PCR or cleavage efficiency of the restriction enzymes was small, and that the number of sequence tag counts was enough to cover the total restriction enzyme sites. At the same time, this false negative rate is considered low. Under the MSDS method, we determined the cut-off values for poorly, moderately and highly methylated sites, based on the results of bisulfite sequencing. The depth of the MSDS sequence affects the cut-off values, and deeper sequencing coverage should improve accuracy, as is the case with the MSCC method [35]. Quantitative information was based on the comparison between bisulfite sequencing data and MSDS tag counts in Fig. 1E. The reliability of these data might be affected by imperfect cleavage by BssHII, EagI and SacII enzymes or a bias derived from the PCR amplification used in MSDS. However, the cutting efficiency of these enzymes was verified, and only a limited number of PCR cycles were used to make the construct.

The MSDS method was able to detect the difference between highly methylated and poorly methylated sites. In the process of the carcinogenesis, DNA methylation is changed in many ways. MSDS was unable to detect minor changes (10–20%) in DNA methylation. However, the MSDS method was able to detect large changes in DNA methylation (e.g. a change from 0–20% methylation to 80–100% methylation).

The MSDS method is a robust means of detecting global DNA methylation. This method has the potential to provide novel insight into the functional complexity of the human genome, and may also serve as a basis for the diagnosis of diseases such as cancer.

5. Accession number

MSDS tags have been deposited in the NCBI Short Read Archive under the project accession SRA028119. Expression tags for HT29 and

HCT116 cell lines have been deposited in NCBI Short Read Archive under the project accessions SRA002659 and SRA028119, respectively.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.07.003.

Acknowledgments

This work was partly supported by a Grant-in Aid for Scientific Research (C) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by Core Research for Evolutional Science and Technology (CREST) funding from the Japan Science and Technology Agency (JST). We thank M. Sarashi, H. Hanaoka, O. Takahashi, K. Azuma, W. Suda, E. Iioka, K. Furuya, M. Ogawa, S. Tomura, F. Shand, Y. Okajima, K. Harukawa, and S. Hatanaka for technical assistance.

References

[1] T.H. Bestor, The DNA methyltransferases of mammals, Hum. Mol. Genet. 9 (2000) 2395–2402.
[2] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.
[3] W. Reik, Stability and flexibility of epigenetic gene regulation in mammalian development, Nature 447 (2007) 425–432.
[4] P.W. Laird, The power and the promise of DNA methylation markers, Nat. Rev. Cancer 3 (2003) 253–266.
[5] A.P. Feinberg, The epigenetics of cancer etiology, Semin. Cancer Biol. 14 (2004) 427–432.
[6] S.E. Ozanne, M. Constancia, Mechanisms of disease: the developmental origins of disease and the role of the epigenotype, Nat. Clin. Pract. Endocrinol. Metab. 3 (2007) 539–546.
[7] P.A. Jones, S.B. Baylin, The epigenomics of cancer, Cell 128 (2007) 683–692.
[8] M. Esteller, Epigenetics in cancer, N. Engl. J. Med. 358 (2008) 1148–1159.
[9] I. Keshet, et al., Evidence for an instructive mechanism of de novo methylation in cancer cells, Nat. Genet. 38 (2006) 149–153.
[10] S.H. Cross, J.A. Charlton, X. Nan, A.P. Bird, Purification of CpG islands using a methylated DNA binding column, Nat. Genet. 6 (1994) 236–244.
[11] T.A. Rauch, X. Wu, X. Zhong, A.D. Riggs, G.P. Pfeifer, A human B cell methylome at 100-base pair resolution, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 671–678.
[12] M. Frommer, et al., A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 1827–1831.
[13] S.J. Clark, J. Harrison, C.L. Paul, M. Frommer, High sensitivity mapping of methylated cytosines, Nucleic Acids Res. 22 (1994) 2990–2997.
[14] A.P. Bird, E.M. Southern, Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis, J. Mol. Biol. 118 (1978) 27–47.
[15] M. Hu, et al., Distinct epigenetic changes in the stromal cells of breast cancers, Nat. Genet. 37 (2005) 899–905.
[16] A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome, Science 315 (2007) 1141–1143.
[17] B. Khulan, et al., Comparative isoschizomer profiling of cytosine methylation: the HELP assay, Genome Res. 16 (2006) 1046–1055.
[18] X. Zhang, et al., Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis, Cell 126 (2006) 1189–1201.
[19] D. Zilberman, M. Gehring, R.K. Tran, T. Ballinger, S. Henikoff, Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription, Nat. Genet. 39 (2007) 61–69.
[20] M. Bibikova, et al., High-throughput DNA methylation profiling using universal bead arrays, Genome Res. 16 (2006) 383–393.
[21] R.A. Irizarry, et al., The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores, Nat. Genet. 41 (2009) 178–186.
[22] A. Meissner, et al., Genome-scale DNA methylation maps of pluripotent and differentiated cells, Nature 454 (2008) 766–770.
[23] R. Lister, et al., Highly integrated single-base resolution maps of the epigenome in Arabidopsis, Cell 133 (2008) 523–536.
[24] S.J. Cokus, et al., Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning, Nature 452 (2008) 215–219.
[25] J. Li, et al., An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines, BMC Genomics 10 (2009) 223.
[26] A. Barski, et al., High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.
[27] L.W. Hillier, et al., Whole-genome sequencing and variant discovery in *C. elegans*, Nat. Methods 5 (2008) 183–188.
[28] S. Hashimoto, et al., High-resolution analysis of the 5′-end transcriptome using a next generation DNA sequencer, PLoS One 4 (2009) e4108.
[29] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. 18 (2008) 1509–1517.
[30] M. Hafner, et al., Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing, Methods 44 (2008) 3–12.
[31] K.E. Holt, et al., High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi, Nat. Genet. 40 (2008) 987–993.
[32] J.D. McGhee, et al., ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult, Dev. Biol. 327 (2009) 551–565.
[33] S. Yagi, et al., DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression, Genome Res. 18 (2008) 1969–1978.
[34] Y.J. Shann, et al., Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines, Genome Res. 18 (2008) 791–801.
[35] M.P. Ball, et al., Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, Nat. Biotechnol. 27 (2009) 361–368.
[36] A.K. Maunakea, et al., Conserved role of intragenic DNA methylation in regulating alternative promoters, Nature 466 (2010) 253–257.
[37] R. Lister, et al., Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322.
[38] F. Eckhardt, et al., DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 38 (2006) 1378–1385.
[39] B. Demircan, et al., Comparative epigenomics of human and mouse mammary tumors, Genes Chromosomes Cancer 48 (2009) 83–97.
[40] C. Lu, et al., Regulation of tumor angiogenesis by EZH2, Cancer Cell 18 (2010) 185–197.
[41] M. Fukasawa, et al., Microarray analysis of promoter methylation in lung cancers, J. Hum. Genet. 51 (2006) 368–374.
[42] S. Yamashita, Y. Tsujino, K. Moriguchi, M. Tatematsu, T. Ushijima, Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2′-deoxycytidine treatment and oligonucleotide microarray, Cancer Sci. 97 (2006) 64–71.
[43] Y. Mori, et al., A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer, Gastroenterology 131 (2006) 797–808.
[44] Z. Jin, et al., Hypermethylation of the nel-like 1 gene is a common and early event and is associated with poor prognosis in early-stage esophageal adenocarcinoma, Oncogene 26 (2007) 6332–6340.
[45] A. Dallol, et al., Tumour specific promoter region methylation of the human homologue of the Drosophila Roundabout gene DUTT1 (ROBO1) in human cancers, Oncogene 21 (2002) 3020–3028.
[46] J. Xian, et al., Targeted disruption of the 3p12 gene, Dutt1/Robo1, predisposes mice to lung adenocarcinomas and lymphomas with methylation of the gene promoter, Cancer Res. 64 (2004) 6432–6437.
[47] S. Ghosh, et al., Alterations of ROBO1/DUTT1 and ROBO2 loci in early dysplastic lesions of head and neck: clinical and prognostic implications, Hum. Genet. 125 (2009) 189–198.
[48] V.C. Wong, et al., Identification of an invasion and tumor-suppressing gene, Endoglin (ENG), silenced by both epigenetic inactivation and allelic loss in esophageal squamous cell carcinoma, Int. J. Cancer 123 (2008) 2816–2823.

OPEN ACCESS Freely available online

PLoS one

Highly Parallel and Short-Acting Amplification with Locus-Specific Primers to Detect Single Nucleotide Polymorphisms by the DigiTag2 Assay

Nao Nishida[1,2]*, Yoriko Mawatari[1,2], Megumi Sageshima[1], Katsushi Tokunaga[1]

1 Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **2** Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan

Abstract

The DigiTag2 assay enables analysis of a set of 96 SNPs using Kapa 2GFast HotStart DNA polymerase with a new protocol that has a total running time of about 7 hours, which is 6 hours shorter than the previous protocol. Quality parameters (conversion rate, call rate, reproducibility and concordance) were at the same levels as when genotype calls were acquired using the previous protocol. Multiplex PCR with 192 pairs of locus-specific primers was available for target preparation in the DigiTag2 assay without the optimization of reaction conditions, and quality parameters had the same levels as those acquired with 96-plex PCR. The locus-specific primers were able to achieve sufficient (concentration of target amplicon $\geq$5 nM) and specific (concentration of unexpected amplicons $<$2 nM) amplification within 2 hours, were also able to achieve detectable amplifications even when working in a 96-plex or 192-plex form. The improved DigiTag2 assay will be an efficient platform for screening an intermediate number of SNPs (tens to hundreds of sites) in the replication analysis after genome-wide association study. Moreover, highly parallel and short-acting amplification with locus-specific primers may thus facilitate widespread application to other PCR-based assays.

Citation: Nishida N, Mawatari Y, Sageshima M, Tokunaga K (2012) Highly Parallel and Short-Acting Amplification with Locus-Specific Primers to Detect Single Nucleotide Polymorphisms by the DigiTag2 Assay. PLoS ONE 7(1): e29967. doi:10.1371/journal.pone.0029967

Editor: Javier S. Castresana, University of Navarra, Spain

Received September 26, 2011; **Accepted** December 9, 2011; **Published** January 13, 2012

Funding: This work was supported by a KAKENHI [grant number 22710191] Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and the Miyakawa Memorial Research Foundation. Partial support by the SENTAN program, Japan Science and Technology Agency, is also acknowledged. The funders had no direct role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nishida-75@umin.ac.jp

Introduction

Polymerase chain reaction (PCR) is a commonly used technique in molecular biology. Several previously developed methods have employed multiplexed PCR in order to analyze genomic variations such as microsatellites or short tandem repeats (STRs), single nucleotide polymorphisms (SNPs) and insertions/deletions [1–3]. Multiplexed preparation of DNA templates in a single reaction is cost-effective, saving starting materials and run-time, while requiring careful optimization of assay conditions. The optimization process is highly empirical and time consuming, and depending on the combinations of markers, may or may not lead to successful assay development. For the conventional design of multiplex PCR, optimization of reaction conditions and careful pre-selection of targets are required in order to prevent excessive off-target priming by the numerous primers in the reaction. Moreover, the risk of generating errors in multiplex PCR, such as insufficient amplification, biased amplification and considerable primer-dimer formation within primers, tends to increase roughly as the square of the number of added primer pairs [4].

There are several approaches to resolving these drawbacks, including solid-phase assay formats (glass slide arrays, microbeads), oligonucleotides containing locked nucleic acid (LNA) residues and circularized amplification. Primers immobilized on the surface of the solid phase appear to markedly increase product yield on solid supports and may avoid the need for target pre-selection with a modification to enrich the input genomic DNA via a crude solution-phase multiplex PCR [5,6]. LNA pentamers showed high priming efficiency to achieve small biased priming in multiplex PCR [7]. Circularized amplification avoids generating artifacts associated with conventional multiplex PCR where two primers are used for each target [8]. This procedure was shown to perform a 96-plex amplification of an arbitrary set of specific DNA sequences. The arrayed primer extension-based genotyping method (APEX-2) allows efficient homogeneous 640-plex DNA amplification with locus-specific primers [9]. These approaches show effective consequences for multiplex amplification, however, a small number of approaches are practically used in the field of molecular genetics, presumably due to its cost and time consuming steps in preparation.

We developed the DigiTag2 assay for multiplex SNP typing as a simple and cost effective approach by combining multiplex PCR to enrich genetic regions including the target SNPs and an oligonucleotide ligation assay to encode all of the SNP genotypes into well-designed oligonucleotides designated DNA coded numbers (DCNs) [10]. For an effective primer design for multiplex PCR, there are several important physical properties for primer sequences, including melting temperature, Gibbs energy of duplex between primer and template, and interactions between primers and PCR amplicons. The DNA polymerase enzyme used in a multiplex PCR is one of the important factors for a successful unbiased amplification.

The DigiTag2 assay is a suitable approach to analyze an intermediate number of SNPs (tens to hundreds of locus) in the replication study after genome wide association study [11–12]. However, the most time consuming step for the DigiTag2 assay in a total running time of 13 hours is multiplex PCR for target preparation (5.5 hours). Here, we report an improved protocol for the DigiTag2 assay with a short-acting multiplex PCR through the use of Kapa 2GFast HotStart DNA polymerase, which reduces total running time and increases assay throughput. In this study, we also validate the applicability of the 192-plex PCR with locus specific primers to amplify the target regions from genomic DNA, which leads to save genomic DNA samples.

Methods

DNA samples

Genomic DNA samples from 96 unrelated healthy donors were obtained from the Japan Health Science Research Resources Bank (Osaka, Japan). All donors provided written informed consent and samples were anonymized. One microgram of purified genomic DNA was dissolved in 100 µl of TE buffer (pH 8.0) (Wako, Osaka, Japan), followed by storage at −20°C until use.

Primer design

A total of 192 pairs of primer were designed using the Visual OMP software version 7.1.0.0 (DNA software, Ann Arbor, MI, USA) with relatively long length (35–45-mer; average, 39.5-mer) to give amplicon sizes between 312 bp and 995 bp (average, 589 bp), each of which had an SNP site (Table S1). Prediction of DNA melting temperature was calculated using nearest-neighbor thermodynamic models. To avoid spurious amplification products, we employed a two-step protocol (denature and extension steps) using specifically designed primer pairs with an extension temperature at 68°C. The specificity of primer sequences was verified by Blat search in order to predict its location(s) on the human genome (GRCh37), and to confirm no unexpected SNP(s) within the primer sequence. The specificity of primer pairs was verified using MFE primer software, which can predict potential amplicon(s) generated from the human genome (GRCh37, up to 5 kb in amplicon size) [13]. All oligonucleotides (de-salted, 100 pmol/µl in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA)) were purchased from Life Technologies (Carlsbad, CA, USA), and were stored at −20°C.

Multiplex PCR with Kapa 2GFast HotStart DNA polymerase

Multiplex PCR mix had a final volume of 10 µl, including 10 ng of genomic DNA, 25 nM each primer, 1.5× KAPA2G Buffer (including 2.25 mM Mg^{2+}), an additional 2.25 mM Mg^{2+} (final concentration of Mg^{2+}: 4.5 mM), 0.2 mM dNTPs and 0.4 U of Kapa 2GFast HotStart DNA polymerase (Kapa Biosystems, Woburn, MA, USA). PCR amplification was conducted using a TGradient (Biometra, Göttingen, Germany) or PTC-225 (MJ Research, Waltham, MA, USA) as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and 68°C for 2 min. When necessary, the fragment length of PCR products was confirmed by capillary electrophoresis (Agilent 2100 Bioanalyzer, Agilent, Santa Clara, CA, USA) in order to evaluate PCR efficiency. The total running times for multiplex PCR with Kapa 2GFast HotStart DNA polymerase using TGradient and PTC-225 were 1 h 48 min 55 s and 2 h 6 min 59 s, respectively.

Multiplex PCR with QIAGEN Multiplex PCR Kit

Multiplex PCR mix had a final volume of 10 µl, including 10 ng of genomic DNA, 25 nM each primer, 1× Multiplex PCR Buffer (including 3.0 mM Mg^{2+}), 0.2 mM dNTPs and HotStar-Taq DNA polymerase (QIAGEN Multiplex PCR Kit; QIAGEN, Valencia, CA, USA). PCR amplification was conducted using a TGradient or PTC-225 as follows: 95°C for 15 min, followed by 40 cycles of 95°C for 30 s and 68°C for 6 min. The total running times for multiplex PCR with QIAGEN Multiplex PCR Kit using TGradient and PTC-225 were 5 h 27 min 53 s and 5 h 46 min 39 s, respectively.

96-plex genotyping by the DigiTag2 assay

The DigiTag2 assay performs multiplex SNP typing by encoding all of the SNP genotypes into well-designed oligonucleotides, designated DNA coded numbers (Figure 1, DCNs: D1_i, ED-1 and ED-2) [10]. The DCNs are assigned to the target SNPs in an unconstrained manner; therefore, the DNA chips prepared to read out the types of DCNs are universally available for any type of SNP without optimization of assay conditions. The DigiTag2 assay proceeds in four steps; target preparation, encoding, labeling and detection.

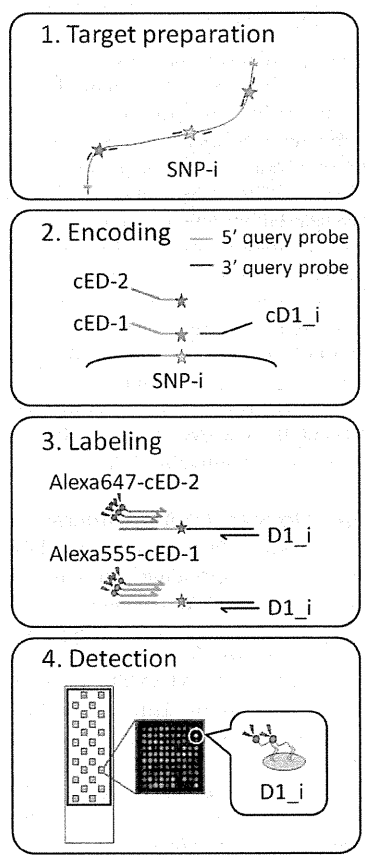

Figure 1. Schematic representation of the DigiTag2 assay. The assay has four steps: target preparation, encoding, labeling and detection. SNP genotypes are encoded into well-designed oligonucleotides, designated DNA coded numbers (DCNs: D1_i, ED-1 and ED-2). D1_i is a variable sequence assigned to each SNP. Reverse complement sequences are written by attaching the character 'c' before the sequence name.
doi:10.1371/journal.pone.0029967.g001

The encoding reactions had a final volume of 15 µl, including 0.5 µl of multiplex PCR products, 20 mM Tris-HCl, pH 7.6, 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1 mM NAD, 0.1% Triton X-100 (1× Taq DNA ligase buffer) with 0.33 nM of each probe and 5 U Taq DNA ligase (New England BioLabs, Ipswich, MA, USA). Encoding reactions were conducted using a TGradient or PTC-225 under the following conditions: 95°C for 5 min, followed by 58°C for 15 min. The reaction was stopped by holding the temperature at 10°C.

The labeling reactions had a final volume of 12 µl, including 6 µl of ligation products, 0.5 µM each labeled primer (Alexa555-cED-1 and Alexa647-cED-2), 2.5 nM each D1 primer (D1_i), 50 mM KCl, 2 mM Mg^{2+}, 0.1 mM DTT, 0.2 mM each dNTP (N = A, G, C), 0.1 mM [^{3}H]-dTTP, 0.25 mg/ml activated salmon sperm DNA (1× *Ex Taq* Buffer) and 0.05 U of *Ex Taq*™ polymerase (TaKaRa, Shiga, Japan). Labeling reactions were conducted using a TGradient or PTC-225 under the following conditions: first held at 95°C for 1 min, followed by 30 cycles of 95°C for 30 s, 55°C for 6 min and 72°C for 30 s. The reaction was stopped by holding the temperature at 10°C. Total running times for labeling using TGradient and PTC-225 were 3 h 49 min 48 s and 4 h 8 min 48 s, respectively.

In the detection step, a hybridization mixture was prepared by mixing 6.25 µl of labeling products with 8.75 µl of hybridization buffer containing 0.5× SSC, 0.1% SDS, 15% formamide, 1 mM EDTA and 3.125 fmol of hybridization control (Alexa555-labeled D1_100 and Alexa647-labeled D1_100). The hybridization control was prepared for ensuring the hybridization step. Ten microliters of hybridization mixture was applied to each block on the universal DNA chip. Hybridization was carried out for 30 min at 37°C in a hybridization oven (ThermoStat plus; Eppendorf, Ham, Germany). After hybridization, glass slides were washed in washing buffer (0.1× SSC, 0.1% SDS) by shaking at 60 rpm for 3 min. Glass slides were consecutively washed in distilled water by shaking at 60 rpm for 1 min and then dried up by centrifugation at 500× g for 1 min. Hybridization images were scanned at photomultiplier voltages of 400 V for Alexa555 and 480 V for Alexa647 using a commercially available DNA chip scanner and fluorescence image analysis was performed using commercially available software (GenePix 4000B unit and GenePix Pro 4.1 software package; Molecular Devices, Sunnyvale, CA, USA).

Labeling with Kapa 2GFast HotStart DNA polymerase

The labeling reactions with Kapa 2GFast HotStart DNA polymerase had a final volume of 12 µl, including 6 µl of ligation products, 0.5 µM each labeled primer (Alexa555-cED-1 and Alexa647-cED-2), 2.5 nM each D1 primer (D1_i), 1.5× KAPA2G Buffer (including 2.25 mM Mg^{2+}), an additional 2.25 mM Mg^{2+} (final concentration of Mg^{2+}: 4.5 mM), 0.2 mM dNTPs and 0.4 U of Kapa 2GFast HotStart DNA polymerase. Labeling reactions were conducted using a TGradient or PTC-225 under the following conditions: first held at 95°C for 1 min, followed by 30 cycles of 95°C for 15 s, 55°C for 120 s and 72°C for 5 s. The reaction was stopped by holding the temperature at 10°C. The total running times for labeling using TGradient and PTC-225 were 1 h 29 min 48 s and 1 h 48 min 34 s, respectively.

Results

Singleplex PCR using 192 pairs of locus-specific primers

Singleplex PCR was conducted under the same reaction condition with multiplex PCR using 25 ng of genomic DNA to ensure target amplicon detection and to confirm the emergence of extra bands (unexpected amplicons). Singleplex PCR with 192 pairs of locus-specific primers revealed that most of the primer pairs are able to achieve sensitive detection (concentration of target amplicon ≥5 nM) and specific amplification without extra bands (concentration of unexpected amplicons <2 nM) except for 14 pairs of primers; low sensitivity (<5 nM) for 5 pairs of primers (61, 99, 102, 189 and 191) and low specificity with extra bands (≥2 nM) for 9 pairs of primers (40, 56, 62, 70, 91, 106, 149, 173 and 174) (Figure 2 and Table S2). Five pairs among the 9 low-specific primer pairs with extra bands (62, 70, 149, 173 and 174) resulted from heteroduplex formation of target amplicons during polyacrylamide gel electrophoresis. Despite the presence of extra bands, the remaining 4 pairs of low-specific primers had a target amplicon with a detectable concentration ≥5 nM.

Validation of efficacy of 192-plex PCR by 96-plex genotyping with the DigiTag2 assay

The DigiTag2 assay enables the simultaneous analysis of 96 target SNPs in: (1) multiplex PCR with locus-specific primers to amplify target genomic regions including target SNPs; (2) multiple oligonucleotide ligation assay with locus-specific probes to determine the genotype of each SNP; and (3) hybridization to the universal DNA chip tethered with probe sequences identical to D1_i (23-mer) (Figure 1) [10]. The validity of 192-plex PCR was assessed with 96 individual DNAs (population control samples) by comparing two sets of 96-plex genotype calls acquired from 96-plex PCR with those from 192-plex PCR (Table 1).

Conversion rate shows the proportion of successfully genotyped SNPs with fewer than 3 undetected samples after excluding low-quality genotyping data, which had more than 5 undetected SNPs in a total of 96 SNPs. However, the composition of failed SNPs in genotyping was not identical, and the conversion rate showed no differences between 192-plex PCR and 96-plex PCR. For the 1st set of 96 SNPs, 7 SNPs among 10 failed SNPs were matched between 192-plex PCR and 96-plex PCR, and for the 2nd set, 8 SNPs among the 9 failed SNPs were matched. The average call rate for successfully genotyped SNPs was over 99.79% for both sets of 96-plex genotyping, even if 192-plex PCR products were adopted for target preparation. Reproducibility was determined by independent genotyping with 96 individuals twice. As a consequence, four discordant genotype calls were observed in the duplicated genotyping data. Concordance of genotype calls between 192-plex PCR and 96-plex PCR was determined using 6,290 genotype calls for the 1st set and 7,884 genotype calls for the 2nd set. Consequently, 14,171 out of 14,174 genotype calls were matched by comparison with 83 SNPs for the 1st set and 86 SNPs for the 2nd set. In total, 3 discordant genotype calls were observed (Figure 3).

Short-acting multiplex PCR by use of Kapa 2GFast HotStart DNA polymerase

Kapa 2GFast HotStart DNA polymerase was employed to perform multiplex PCR with the locus-specific primers for target preparation in genotyping with the DigiTag2 assay. To optimize reaction conditions with Kapa 2GFast HotStart DNA polymerase, singleplex PCR was conducted using 25 ng of genomic DNA with three randomly chosen pairs of locus-specific primers. The designed amplicon sizes for the three pairs of primers were 501 bp, 671 bp and 492 bp. We performed singleplex PCR using a two-step protocol (denature and extension steps) with varied extension periods (15 s, 30 s, 60 s and 120 s) and with varied Mg^{2+} concentrations (3.0 mM and 4.5 mM) (Figure 4). The most sensitive detection and highest levels of amplification for the three

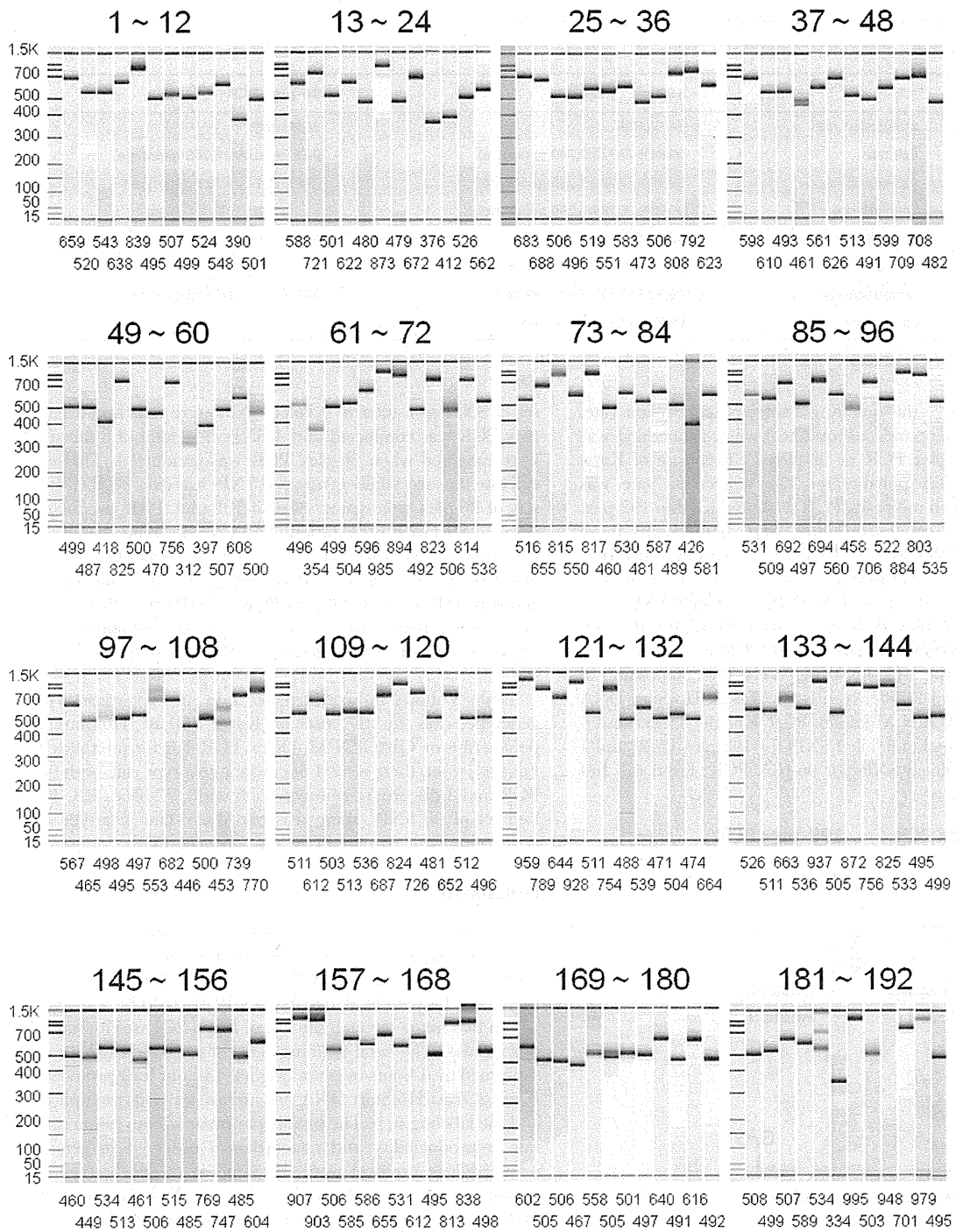

Figure 2. Electropherogram of singleplex PCR products with 192 pairs of locus-specific primers. The designed amplicon size is depicted below each lane.
doi:10.1371/journal.pone.0029967.g002

pairs of primers were observed with 120 s for the extension period and 4.5 mM for the Mg^{2+} concentration. The total running time for multiplex PCR with locus-specific primers was less than 2 hours, which is about 3 h 30 min shorter than the previous protocol (see MATERIALS AND METHODS).

The total running time of the DigiTag2 assay was markedly reduced when the labeling step was also conducted using Kapa 2GFast HotStart DNA polymerase instead of *Ex Taq* polymerase. When the DigiTag2 assay was conducted with Kapa 2GFast HotStart DNA polymerase for multiplex PCR and labeling step, the total running time of the assay was about 7 hours, which is about 6 hours shorter than the previously used protocol in combination with QIAGEN Multiplex PCR Kit for multiplex PCR and *Ex Taq* polymerase for the labeling step.

Table 1. Validation of efficacy of 192-plex PCR by 96-plex genotyping.

		192-plex PCR	96-plex PCR
1st set	Conversion rate	86/96 SNP	86/96 SNP
	Call rate	99.84% (7,728/7,740 genotype)	99.81% (6,695/6,708 genotype)
	reproducibility	99.99% (7,288/7,289 genotype)	100% (6,121/6,121 genotype)
	concordance	99.98% (6,289/6,290 genotype)	
2nd set	Conversion rate	87/96 SNP	87/96 SNP
	Call rate	99.79% (8,074/8,091 genotype)	99.79% (8,161/8,178 genotype)
	reproducibility	99.97% (7,792/7,794 genotype)	99.99% (7,712/7,713 genotype)
	concordance	99.97% (7,882/7,884 genotype)	

doi:10.1371/journal.pone.0029967.t001

Table 2 summarizes the quality parameters (conversion rate, call rate, reproducibility and concordance) when genotyping was conducted with 192-plex PCR or 96-plex PCR by use of Kapa 2GFast HotStart DNA polymerase. The conversion rate was slightly decreased when multiplex PCR was conducted in 192-plex form. However, the conversion rates were better than those observed when multiplex PCR was conducted with the QIAGEN Multiplex PCR Kit. The composition of failed SNPs in genotyping was not consistent for the 1st set of 96 SNPs, in which 4 SNPs were matched between 192-plex PCR and 96-plex PCR. For the 2nd set, a total of 8 failed SNPs in the 96-plex PCR were completely matched to those in the 192-plex PCR. When the composition of failed SNPs were compared between Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit, the 1st set had 5 matched SNPs in a total of 8 failed SNPs for 192-plex PCR, and 4 matched SNPs in 5 failed SNPs for 96-plex PCR. From the 2nd set, 5 SNPs in a total of 9 failed SNPs were matched when 192-plex PCR was conducted and 4 SNPs in a total of 8 failed SNPs were matched when 96-plex PCR was conducted. The average call rate for successfully genotyped SNPs was over 99.76% for both sets of 96-plex genotyping, even if 192-plex PCR products were adopted for target preparation. The reproducibility was 100% for the 2nd set; however, three discordant genotype calls were observed for the 1st set. With regard to the concordance of genotype calls between 96-plex PCR and 192-plex PCR, only one discordant genotype call was observed in the comparison for the 1st set, and no discordant genotype calls were observed in the 2nd set.

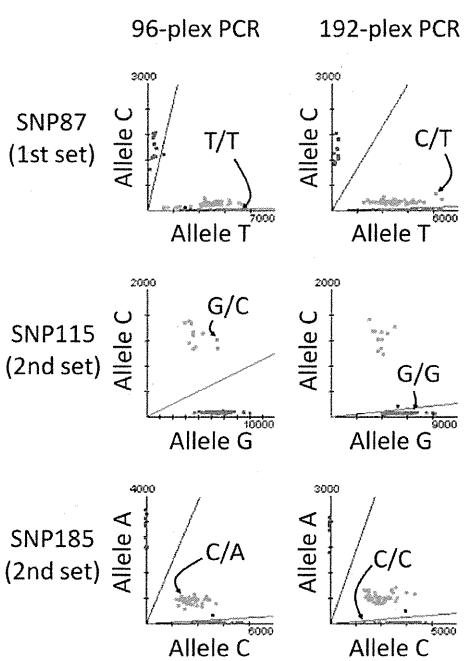

Figure 3. Scatter plots for three SNPs with 3 discordant genotypes. Scatter plots in genotyping with 192-plex PCR and 96-plex PCR are depicted side-by-side. The genotypes of discordant samples are indicated in the scatter plots by arrows.
doi:10.1371/journal.pone.0029967.g003

Table 3 shows the concordance rate in comparison with the genotype calls by the use of Kapa 2GFast HotStart DNA polymerase or QIAGEN Multiplex PCR Kit for multiplex PCR. For the 1st set, there were 4 discordant genotype calls with 96-plex PCR and 8 discordant genotype calls with 192-plex PCR. For the 2nd set of 96 SNPs, there was one discordant genotype call in genotyping with 96-plex PCR and 192-plex PCR.

Discussion

The locus specific primers sufficiently worked in a multiplex form under the same reaction conditions without any optimization processes, either 96-plex PCR or 192-plex PCR. We also found that either 96-plex PCR or 192-plex PCR could be accomplished within two hours through the use of Kapa 2GFast HotStart DNA polymerase. The total running time of the DigiTag2 assay was shortened by 6 hours over the original 13-hour long protocol using Kapa 2GFast HotStart DNA polymerase for both multiplex PCR and the labeling step. The quality parameters (conversion rate, call rate, reproducibility and concordance) observed in genotyping with the new protocol were the same as those observed in the original protocol using QIAGEN Multiplex PCR Kit for multiplex PCR and *Ex Taq* polymerase for the labeling step. The DigiTag2 assay worked with a conversion rate of over 93.2% (179 / 192 SNPs), average call rate of over 99.80% (16,789/16,823 genotypes) and reproducibility of over 99.99% (16,135/16,136 genotypes) using 96-plex PCR under the new protocol. The composition of successfully genotyped SNPs was different when the genotype calls were acquired using the different polymerases (Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit), which would result from a varying amplification bias in multiplex PCR. We also found that 192-plex PCR with locus-specific primers worked in 96-plex genotyping with the DigiTag2 assay, giving the same quality parameter data as those observed in genotyping with 96-plex PCR. However, the

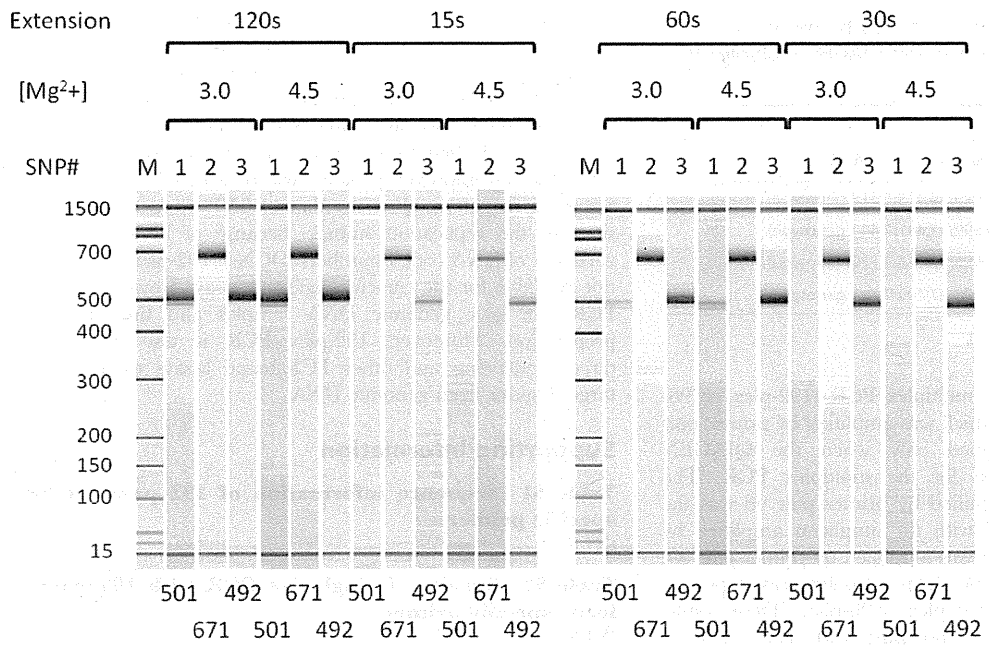

Figure 4. Electropherogram of singleplex PCR products using Kapa 2GFast HotStart DNA polymerase. Singleplex PCR was performed with varied extension periods (15 s, 30 s, 60 s and 120 s) and with varied Mg^{2+} concentrations (3.0 mM and 4.5 mM) using three pairs of locus-specific primers. The designed amplicon size is depicted below each lane.
doi:10.1371/journal.pone.0029967.g004

composition of successfully genotyped SNPs was not consistent between 192-plex PCR and 96-plex PCR, which may be explained by changing the interactions between primer pairs in 192-plex PCR and in 96-plex PCR. The composition of successful SNPs was not consistent when using different polymerases or multiplex systems in the multiplex PCR, which casts some shadows on the reliability of the assay. Regardless of the existing shadows, indeed, 96-plex and 192-plex PCR work with a high conversion rate in genotyping over 93.2%. To clear the existing shadows, it is necessary to continuously accumulate genotyping data.

In this study, fifteen discordant genotype calls were in total observed in the comparison of genotype calls with: i) duplicated genotyping data; ii) genotyping data by use of 192-plex PCR and 96-plex PCR; and iii) genotyping data with different types of polymerases (Table S3). Table S3 shows the genotype calls acquired 8 times under different conditions. All fifteen discordant genotype calls were analyzed with direct sequencing, of which 13 genotype calls were determined. In 8 of 15 discordant genotype calls, the genotype calls were completely different depending on the type of polymerase. The genotype calls acquired using Kapa 2GFast HotStart DNA polymerase were 100% concordant (6 of 6) with those acquired by direct sequencing. This suggests that SNP allelic bias in PCR amplification readily occurred with the QIAGEN Multiplex PCR Kit; however, the error rate in genotyping was only 0.04% (6 out of 14,886 genotypes). The remaining 7 discordant genotype calls were randomly observed in 1 out of 8 different conditions. This shows that the random error rates were almost equal in the genotype data acquired with both types of polymerases (4 out of 62,227 genotypes for QIAGEN Multiplex PCR Kit and 3 out of 66,008 genotypes for Kapa 2GFast HotStart DNA polymerase).

Among the five low-sensitivity primer pairs found on singleplex PCR (61, 99, 102, 189 and 191), no amplicons were detected by primer pair 189 and low concentrations (<5 nM) of amplicon were detected by the 4 other primer pairs (Table S2). Therefore, the SNP189 failed in genotyping, independently of the type of

Table 2. Validation of efficacy of 192-plex and 96-plex PCR with Kapa 2GFast HotStart DNA polymerase.

		192-plex PCR	96-plex PCR
1st set	Conversion rate	88/96 SNP	91/96 SNP
	Call rate	99.84% (8,259/8,272 genotype)	99.76% (8,443/8,463 genotype)
	reproducibility	99.97% (8,069/8,071 genotype)	99.99% (8,339/8,340 genotype)
	concordance	99.99% (7,982/7,983 genotype)	
2nd set	Conversion rate	87/96 SNP	88/96 SNP
	Call rate	99.91% (8,171/8,178 genotype)	99.83% (8,346/8,360 genotype)
	reproducibility	100% (7,705/7,705 genotype)	100% (7,796/7,796 genotype)
	concordance	100% (8,161/8,161 genotype)	

doi:10.1371/journal.pone.0029967.t002

Table 3. Concordance of genotype calls between Kapa 2GFast HotStart DNA polymerase and QIAGEN Multiplex PCR Kit.

		Kapa 2G	QIAGEN
1st set	96-plex PCR	99.94% (6,513/6,517 genotype)	
	192-plex PCR	99.89% (7,441/7,449 genotype)	
2nd set	96-plex PCR	99.99% (7,778/7,779 genotype)	
	192-plex PCR	99.99% (7,700/7,701 genotype)	

doi:10.1371/journal.pone.0029967.t003

polymerase and multiplicity in multiplex PCR (192-plex or 96-plex). However, the SNP191, which was amplified by primer pair 191, was successfully genotyped only when the QIAGEN Multiplex PCR Kit was used for the multiplex PCR. The concentration of amplicon amplified by primer pair 99 was the same as the 2.8 nM observed with the amplicon amplified by primer pair 191. SNP99, which was amplified by primer pair 99, was successfully genotyped independently of polymerase type and multiplicity in multiplex PCR (192-plex or 96-plex). These results suggest that the sensitivity in genotyping with Kapa 2GFast HotStart DNA polymerase was lower than the previously used protocol with QIAGEN Multiplex PCR Kit. These results would be explained by a biased amplification with the shortened protocol using Kapa 2GFast HotStart DNA polymerase, which tends to lead to a consequent biased genotyping. However, the investigated number of primer pairs would not be sufficient to decide the sensitivity in genotyping; therefore, it is necessary to continuously accumulate genotyping data. As the investigated number of primer pairs was only 192 (384 primers) in this study, melting temperature of each primer and the number of potential amplicons predicted by the MFE primer software were strongly associated with low sensitivity and low specificity in an amplification, respectively (multiple regression analysis, $P=1.26\times10^{-37}$ and $P=1.52\times10^{-21}$, respectively).

Through the use of Kapa 2GFast HotStart DNA polymerase, the genotype calls for 96 SNPs can be acquired in about 7 hours by the DigiTag2 assay. The genotyping platform with high conversion rate plays an important role for the replication studies to identify the disease associated genes from candidate loci found in the GWAS (genome-wide association study). The DigiTag2 assay with an improved protocol will be an efficient platform for screening an intermediate number of SNPs (tens to hundreds of sites) in the replication studies. Because of limitations in the variation of DNA coded numbers (DCNs), 192-plex genotyping is not available for the current DigiTag2 assay. However, 192-plex PCR can save genomic DNA samples and time for target preparation. Moreover, 192-plex PCR is also available for direct-sequencing and other PCR-based assays to amplify the target regions from genomic DNA.

Supporting Information

Table S1 Sequence information of 192 pairs of locus specific primer.
(XLSX)

Table S2 Results of singleplex PCR with 192 pairs of locus specific primer.
(XLSX)

Table S3 The 15 discordant genotype calls in 8 different conditions.
(XLSX)

Acknowledgments

We would like to thank M. Takasu for technical support, and H. Adachi, N. Tabei and J. Fujimiya (Dynacom Co., Ltd.) for assistance with primer and probe design.

Author Contributions

Conceived and designed the experiments: NN KT. Performed the experiments: YM MS. Analyzed the data: NN YM MS. Contributed reagents/materials/analysis tools: NN YM MS. Wrote the paper: NN KT.

References

1. Deter J, Gala M, Charbonnel N, Cosson JF (2009) Characterization and PCR multiplexing of polymorphic microsatellite loci in the whipworm *Trichuris arvicolae*, parasite of arvicoline rodents and their cross-species utilization in *T. muris*, parasite of murines. Mol Biochem Parasitol 167: 144–146.
2. Hosseini-Maaf B, Hellberg Å, Chester MA, Olsson ML (2007) An extensive polymerase chain reaction–allele-specific polymorphism strategy for clinical ABO blood group genotyping that avoids potential errors caused by null, subgroup, and hybrid alleles. Transfusion 47: 2110–2125.
3. Goguet de la Salmonière YO, Kim CC, Tsolaki AG, Pym AS, Siegrist MS, et al. (2004) High-throughput method for detecting genomic-deletion polymorphisms. J Clin Microbiol 42: 2913–2918.
4. Landegren U, Nilsson M (1997) Locked on target: strategies for future gene diagnostics. Ann Med 29: 585–590.
5. Pemov A, Modi H, Chandler DP, Bavykin S (2005) DNA analysis with multiplex microarray-enhanced PCR. Nucleic Acids Res 33: e11.
6. Meuzelaar LS, Lancaster O, Pasche JP, Kopal G, Brookes AJ (2007) MegaPlex PCR: a strategy for multiplex amplification. Nat Methods 4: 835–837.
7. Sun Z, Chen Z, Hou X, Li S, Zhu H, et al. (2008) Locked nucleic acid pentamers as universal PCR primers for genomic DNA amplification. Plos One 3: e3701.
8. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M (2005) Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. Nucleic Acids Res 33: e71.
9. Krjutškov K, Andreson R, Mägi R, Nikopensius T, Khrunin A, et al. (2008) Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. Nucleic Acids Res 36: e75.
10. Nishida N, Tanabe T, Takasu M, Suyama A, Tokunaga K (2007) Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay. Anal Biochem 364: 78–85.
11. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41: 1105–1109.
12. Miyagawa T, Kawashima M, Nishida N, Ohashi J, Kimura R, et al. (2009) Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. Nat Genet 40: 1324–1328.
13. Qu W, Shen Z, Zhao D, Yang Y, Zhang C (2009) MFEprimer: multiple factor evaluation of the specificity of PCR primers. Bioinformatics 25: 276–278.