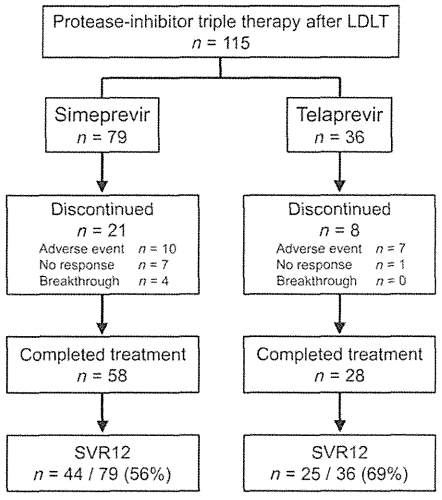

Figure 1 Flow diagram showing the outcomes of 115 patients treated with simeprevir or telaprevir with peginterferon and ribavirin after living-donor liver transplantation (LDLT). The numbers of patients who discontinued the treatment protocol (and their reasons for discontinuation), completed the treatment protocol, and achieved sustained virological response at week 12 (SVR12) after the termination of treatment are shown.

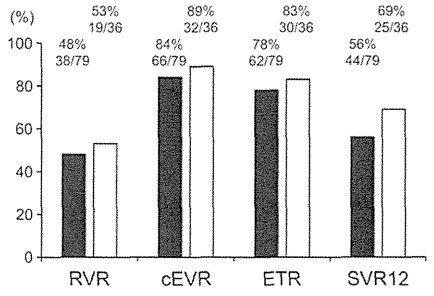

Figure 2 Virological responses in 115 patients treated with simeprevir- or telaprevir-based triple therapy (black and white bars, respectively) after living-donor liver transplantation. Rapid virological response (RVR), complete early virological response (cEVR), and end-of-treatment response (ETR) are defined as undetectable hepatitis C virus RNA in serum at 4 weeks, 12 weeks, and end of treatment, respectively. Sustained virological response at 12 weeks (SVR12) is defined as the absence of hepatitis C virus RNA in the serum for >12 weeks after the termination of treatment.

than telaprevir-based triple therapy, but the differences were not statistically significant.

Safety and tolerability

Adverse events that occurred during the triple therapies are summarized in Table 2. Adverse events occurred in 62% and 72% of patients, including serious adverse events in 11% and 25%, and death in 3% and 3% in the simeprevir- and telaprevir-based triple therapy groups, respectively. Treatment was discontinued due to adverse events in 13% and 19% of patients in the simeprevir and telaprevir groups, respectively. Dose modification of the DAAs, peginterferon, or ribavirin was required in 78 of 79 patients (99%) in the simeprevir group and in all patients (100%) in the telaprevir group. All patients, except for 10 patients (13%) who discontinued the treatment protocol, started receiving simeprevir triple therapy at the standard dose (100 mg/day) and continued the same dose until 12 weeks. Telaprevir was started at a reduced dose (1500 mg/day) in 34 (94%) of 36 patients and was discontinued in 8 patients (22%) until 12 weeks of treatment. The reduced dose of peginterferon at treatment initiation was used in 6 (8%) and 4 (11%) patients, and a reduction from the initial dose during the treatment was required in 22 (28%) and 13 (36%) patients in the simeprevir and telaprevir groups, respectively. A reduced dose of ribavirin compared to the standard dose at treatment initiation was given in 63 (80%) and 35 (97%) patients, and a reduction in the ribavirin dose from the initial dose during treatment was required in 59 (75%) and 30 (83%) patients, including discontinuation in 31 (39%) and 13 (36%) patients, in the simeprevir and telaprevir groups, respectively.

There was no statistically significant difference in the rate of adverse events between the simeprevir and telaprevir groups, although serious adverse events tended to be more frequent in the telaprevir group than in the simeprevir group. The rate of patients who received blood cell transfusion and erythropoietin due to anemia were significantly higher in the telaprevir group than in the simeprevir group. Renal insufficiency, defined when the estimated glomerular filtration rate decreased >30 mL/min/1.73 m^2 from the baseline estimated glomerular filtration rate, was significantly less common in the simeprevir group than in the telaprevir group. Immune-mediated graft dysfunction occurred in 6 patients during simeprevir-based triple therapy, including 3 with acute cellular rejection, 2 with veno-occlusive disease, and 1 with chronic rejection. In

the telaprevir group, IGD occurred in 4 patients, and all had plasma cell hepatitis. Infection was observed in 3 patients in the simeprevir group (2 with a cytomegalovirus infection and 1 with pneumonia), whereas 1 patient had cholangitis in the telaprevir group. In the simeprevir group, 2 patients died of graft failure caused by chronic rejection 5 weeks after the termination of 31 weeks of treatment, and graft failure by infection at 2 weeks of treatment. One patient died of brain hemorrhage at 25 weeks of telaprevir-based triple therapy.

Factors predictive of SVR12 with simeprevir-based triple therapy

Baseline factors that could predict SVR12 with simeprevir-based triple therapy were analyzed by comparing patients in the SVR group ($n = 44$) with those in the non-SVR group ($n = 35$) (Table 3). Three factors, male sex ($P = 0.040$), the presence of prior dual therapy with peginterferon and ribavirin ($P = 0.001$), and non-responders to the prior dual therapy ($P < 0.001$), were identified as significant predictive factors for non-SVR. Associations of prior dual therapy with the efficacy of simeprevir- and telaprevir-based triple therapy are shown in Figure 3. In patients who received simeprevir-based triple therapy, the SVR12 rates were 94% in treatment-naïve patients, and 68%, 67%, and 34% in patients with relapse, withdrawal, and no response to the prior dual therapy, respectively. Differences between treatment-naïve patients and non-responders of prior dual therapy ($P < 0.001$), and between relapsers and non-responders ($P = 0.013$) were statistically significant. The impact of prior dual therapy on the treatment response of triple therapy was observed in both the telaprevir and simeprevir groups, although the difference was not significant in the telaprevir group.

DISCUSSION

IN THE CURRENT study, we showed the efficacy and safety of second-generation NS3/4A inhibitor simeprevir with peginterferon and ribavirin in patients with recurrent hepatitis C after LDLT. The SVR12 rate of simeprevir-based triple therapy was 56% overall, but it was 94% in treatment-naïve patients, indicating that simeprevir-based triple therapy is very effective when patients are selected according to their experience with prior therapy.

The efficacy and safety of first-generation NS3/4A inhibitors telaprevir and boceprevir in liver transplant recipients have been reported mainly in patients after deceased-donor liver transplantation (DDLT).[15–18] Most studies have shown that triple therapy with telaprevir or boceprevir with peginterferon and ribavirin increased the SVR rate, but this resulted in many

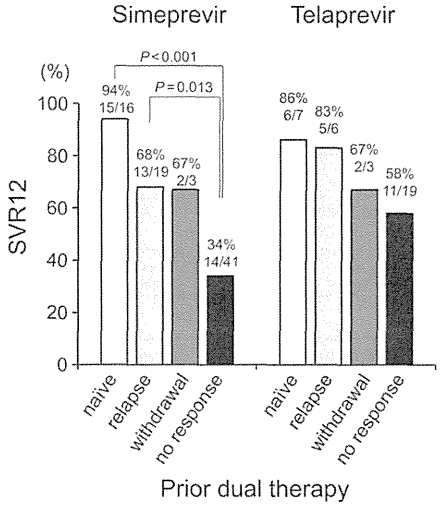

Figure 3 Rate of sustained virological response at 12 weeks (SVR12) after the termination of treatment with simeprevir- or telaprevir-based triple therapy following living-donor liver transplantation in treatment-naïve patients and relapsers, patients who withdrew, and non-responders to the prior dual therapy. *P*-values are shown if the differences are statistically significant ($P < 0.05$).

adverse events. In the present study, the SVR rate of telaprevir-based triple therapy was 69% in patients after LDLT, which is similar to the reported SVR rate of 50–63% in patients after DDLT.[15–18] Severe adverse events, including anemia and renal insufficiency, which have been reported in a previous study on patients after DDLT, also occurred in the present study. These results indicate that the efficacy and safety of telaprevir-based triple therapy in patients after LDLT are similar to those in patients after DDLT.

Compared to telaprevir-based triple therapy, simeprevir-based triple therapy can be more easily administered to transplant recipients because no clinically significant interactions between simeprevir and calcineurin inhibitors have been observed.[20,21] Although the overall incidence of adverse events was not significantly different between telaprevir and simeprevir, simeprevir tended to be associated with fewer serious adverse events than telaprevir. Blood cell transfusion and erythropoietin were less frequently used in the simeprevir group than in the telaprevir group, suggesting that less intensive management for anemia was needed

during simeprevir-based therapy. Moreover, the rate of renal insufficiency was significantly less in the simeprevir group than in the telaprevir group. Furthermore, IGD, including acute cellular rejection, chronic rejection, and plasma cell hepatitis, is one of the major adverse events of interferon-containing therapy in patients after liver transplantation.[13] In this study, there was no difference in the incidence of IGD between the simeprevir and telaprevir groups. Therefore, in terms of safety, simeprevir-based triple therapy is superior to telaprevir-based triple therapy.

The efficacy of simeprevir-based triple therapy was not satisfactory; the SVR rate was 56%. The virological responses, including the RVR, cEVR, ETR, and SVR12, of simeprevir-based triple therapy tended to be lower than those of telaprevir-based triple therapy. To achieve a higher efficacy of simeprevir-based triple therapy in liver transplant recipients, it is necessary to select patients before treatment. An analysis of the predictive factors associated with SVR showed that the presence and efficacy of prior dual therapy are important for predicting the efficacy of simeprevir-based triple therapy. Notably, 94% of treatment-naïve patients achieved SVR12 with simeprevir-based triple therapy, whereas the SVR12 rate in non-responders to prior dual therapy was only 34%. Similar results have been shown in Japanese phase III trials on patients in non-transplant settings; the SVR12 rates of simeprevir-based triple therapy were 88.6%, 95.9%, and 52.8% in treatment-naïve patients, relapsers, and non-responders to prior interferon-based therapy, respectively.[23,36] As the efficacy of dual therapy is determined by multiple factors, including host IL28B genotypes and HCV genomic mutations, these factors may also affect the efficacy of simeprevir-based triple therapy, resulting in the low efficacy of simeprevir-based triple therapy in non-responders to dual therapy. In our study, female patients had a significantly higher SVR12 rate with simeprevir-based triple therapy compared to male patients, although the reason for this difference is unknown. These predictive factors may help in selecting patients before administering simeprevir-based triple therapy, and the efficacy may be higher by selecting patients according to the status of prior dual therapy and sex.

Recent reports have indicated a higher efficacy and safety of interferon-free therapy in liver transplant recipients compared to second-generation NS3/4A inhibitor-based triple therapy clarified in the present study.[29–32] Therefore, interferon-free therapy should be used as first-line therapy for recurrent hepatitis C after liver transplantation, according to the recent recommendation for hepatitis C treatment.[33] The SVR rate of interferon-free therapy in liver transplant recipients is reportedly 70–97%,[29–32] and resistance-associated variants to DAAs were detected in most of the remaining non-SVR patients.[29,31] Second-line therapy with interferon-free therapy for non-SVR patients has not yet been established. As interferon's broad antiviral activity will help clear DAA-resistant HCV, interferon-containing therapy would be one of the choices as second-line therapy for hepatitis C after liver transplantation. Therefore, the efficacy and safety of simeprevir-based therapy clarified in the present study will provide useful information even in the interferon-free therapy era.

In conclusion, simeprevir-based triple therapy for recurrent hepatitis C after LDLT resulted in an SVR rate of 56% and good tolerability. Although this therapy is not recommended for non-responders to prior dual therapy because of low efficacy, simeprevir-based triple therapy may be one of the options for treatment-naïve patients. An individualized treatment strategy that predicts the efficacy and safety of treatment will result in more effective and safer treatment for liver transplant recipients in the DAA era.

ACKNOWLEDGMENTS

THIS RESEARCH WAS supported by the Research Program on Hepatitis from Japan Agency for Medical Research and Development, AMED

REFERENCES

1 Feray C, Caccamo L, Alexander GJ *et al.* European collaborative study on factors influencing outcome after liver transplantation for hepatitis C. European Concerted Action on Viral Hepatitis (EUROHEP) Group. *Gastroenterology* 1999; 117: 619–25.
2 Forman LM, Lewis JD, Berlin JA, Feldman HI, Lucey MR. The association between hepatitis C infection and survival after orthotopic liver transplantation. *Gastroenterology* 2002; 122: 889–96.
3 Prieto M, Berenguer M, Rayon JM *et al.* High incidence of allograft cirrhosis in hepatitis C virus genotype 1b infection following transplantation: relationship with rejection episodes. *Hepatology* 1999; 29: 250–6.
4 Belli LS, Burroughs AK, Burra P *et al.* Liver transplantation for HCV cirrhosis: improved survival in recent years and increased severity of recurrent disease in female recipients: results of a long term retrospective study. *Liver Transpl* 2007; 13: 733–40.
5 Berenguer M, Ferrell L, Watson J *et al.* HCV-related fibrosis progression following liver transplantation: increase in recent years. *J Hepatol* 2000; 32: 673–84.
6 Neumann UP, Berg T, Bahra M *et al.* Fibrosis progression after liver transplantation in patients with recurrent hepatitis C. *J Hepatol* 2004; 41: 830–6.
7 Walter T, Dumortier J, Guillaud O, Hervieu V, Scoazec JY, Boillot O. Factors influencing the progression of fibrosis in patients with recurrent hepatitis C after liver transplantation

under antiviral therapy: a retrospective analysis of 939 liver biopsies in a single center. *Liver Transpl* 2007; **13**: 294–301.
8 Yilmaz N, Shiffman ML, Stravitz RT *et al.* A prospective evaluation of fibrosis progression in patients with recurrent hepatitis C virus following liver transplantation. *Liver Transpl* 2007; **13**: 975–83.
9 Thuluvath PJ, Guidinger MK, Fung JJ, Johnson LB, Rayhill SC, Pelletier SJ. Liver transplantation in the United States, 1999–2008. *Am J Transplant* 2010; **10**: 1003–19.
10 Gordon FD, Kwo P, Vargas HE. Treatment of hepatitis C in liver transplant recipients. *Liver Transpl* 2009; **15**: 126–35.
11 Terrault NA. Hepatitis C therapy before and after liver transplantation. *Liver Transpl* 2008; **14**(Suppl 2): S58–66.
12 Berenguer M. Systematic review of the treatment of established recurrent hepatitis C with pegylated interferon in combination with ribavirin. *J Hepatol* 2008; **49**: 274–87.
13 Levitsky J, Fiel MI, Norvell JP *et al.* Risk for immune-mediated graft dysfunction in liver transplant recipients with recurrent HCV infection treated with pegylated interferon. *Gastroenterology* 2012; **142**: 1132–9 e1.
14 Garg V, van Heeswijk R, Lee JE, Alves K, Nadkarni P, Luo X. Effect of telaprevir on the pharmacokinetics of cyclosporine and tacrolimus. *Hepatology* 2011; **54**: 20–7.
15 Burton JR Jr, O'Leary JG, Verma EC *et al.* A US multicenter study of hepatitis C treatment of liver transplant recipients with protease-inhibitor triple therapy. *J Hepatol* 2014; **61**: 508–14.
16 Coilly A, Roche B, Dumortier J *et al.* Safety and efficacy of protease inhibitors to treat hepatitis C after liver transplantation: a multicenter experience. *J Hepatol* 2014; **60**: 78–86.
17 Faisal N, Yoshida EM, Bilodeau M *et al.* Protease inhibitor-based triple therapy is highly effective for hepatitis C recurrence after liver transplant: a multicenter experience. *Ann Hepatol* 2014; **13**: 525–32.
18 Pungpapong S, Aqel BA, Koning L *et al.* Multicenter experience using telaprevir or boceprevir with peginterferon and ribavirin to treat hepatitis C genotype 1 after liver transplantation. *Liver Transpl* 2013; **19**: 690–700.
19 Fontana RJ, Hughes EA, Bifano M *et al.* Sofosbuvir and daclatasvir combination therapy in a liver transplant recipient with severe recurrent cholestatic hepatitis C. *Am J Transplant* 2013; **13**: 1601–5.
20 Tanaka T, Sugawara Y, Akamatsu N *et al.* Use of simeprevir following pre-emptive pegylated interferon/ribavirin treatment for recurrent hepatitis C in living donor liver transplant recipients: a 12-week pilot study. *J Hepatobiliary Pancreat Sci* 2015; **22**: 144–50.
21 Ueda Y, Kaido T, Uemoto S. Fluctuations in the concentration/dose ratio of calcineurin inhibitors after simeprevir administration in patients with recurrent hepatitis C after liver transplantation. *Transpl Int* 2015; **28**: 251–2.
22 Forns X, Lawitz E, Zeuzem S *et al.* Simeprevir with peginterferon and ribavirin leads to high rates of SVR in patients with HCV genotype 1 who relapsed after previous therapy: a phase 3 trial. *Gastroenterology* 2014; **146**: 1669–79 e3.
23 Hayashi N, Izumi N, Kumada H *et al.* Simeprevir with peginterferon/ribavirin for treatment-naive hepatitis C genotype 1 patients in Japan: CONCERTO-1, a phase III trial. *J Hepatol* 2014; **61**: 219–27.
24 Jacobson IM, Dore GJ, Foster GR *et al.* Simeprevir with pegylated interferon alfa 2a plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-1): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet* 2014; **384**: 403–13.
25 Manns M, Marcellin P, Poordad F *et al.* Simeprevir with pegylated interferon alfa 2a or 2b plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-2): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* 2014; **384**: 414–26.
26 Jacobson IM, McHutchison JG, Dusheiko G *et al.* Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med* 2011; **364**: 2405–16.
27 Kumada H, Toyota J, Okanoue T, Chayama K, Tsubouchi H, Hayashi N. Telaprevir with peginterferon and ribavirin for treatment-naive patients chronically infected with HCV of genotype 1 in Japan. *J Hepatol* 2012; **56**: 78–84.
28 Zeuzem S, Andreone P, Pol S *et al.* Telaprevir for retreatment of HCV infection. *N Engl J Med* 2011; **364**: 2417–28.
29 Charlton M, Everson GT, Flamm SL *et al.* Ledipasvir and sofosbuvir plus ribavirin for treatment of HCV infection in patients with advanced liver disease. *Gastroenterology* 2015; **149**: 649–59.
30 Charlton M, Gane E, Manns MP *et al.* Sofosbuvir and ribavirin for treatment of compensated recurrent hepatitis C virus infection after liver transplantation. *Gastroenterology* 2015; **148**: 108–17.
31 Kwo PY, Mantry PS, Coakley E *et al.* An interferon-free antiviral regimen for HCV after liver transplantation. *N Engl J Med* 2014; **371**: 2375–82.
32 Pungpapong S, Aqel B, Leise M *et al.* Multicenter experience using simeprevir and sofosbuvir with or without ribavirin to treat hepatitis C genotype 1 after liver transplant. *Hepatology* 2015; **61**: 1880–6.
33 Panel AIHG. Hepatitis C guidance: AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus. *Hepatology* 2015; **62**: 932–54.
34 Ohno O, Mizokami M, Wu RR *et al.* New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. *J Clin Microbiol* 1997; **35**: 201–7.
35 Ito K, Higami K, Masaki N *et al.* The rs8099917 polymorphism, when determined by a suitable genotyping method, is a better predictor for response to pegylated alpha interferon/ribavirin therapy in Japanese patients than other single nucleotide polymorphisms associated with interleukin-28B. *J Clin Microbiol* 2011; **49**: 1853–60.
36 Izumi N, Hayashi N, Kumada H *et al.* Once-daily simeprevir with peginterferon and ribavirin for treatment-experienced HCV genotype 1-infected patients in Japan: the CONCERTO-2 and CONCERTO-3 studies. *J Gastroenterol* 2014; **49**: 941–53.

PLOS | GENETICS

RESEARCH ARTICLE

Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution

Ryutaro Uchi[1,2☯], Yusuke Takahashi[1,3☯], Atsushi Niida[4‡*], Teppei Shimamura[4], Hidenari Hirata[1], Keishi Sugimachi[1], Genta Sawada[1,3], Takeshi Iwaya[5], Junji Kurashige[1], Yoshiaki Shinden[1], Tomohiro Iguchi[1], Hidetoshi Eguchi[1], Kenichi Chiba[4], Yuichi Shiraishi[4], Genta Nagae[6], Kenichi Yoshida[7], Yasunobu Nagata[7], Hiroshi Haeno[8], Hirofumi Yamamoto[3], Hideshi Ishii[3], Yuichiro Doki[3], Hisae Iinuma[9], Shin Sasaki[10], Satoshi Nagayama[11], Kazutaka Yamada[12], Shinichi Yachida[13], Mamoru Kato[13], Tatsuhiro Shibata[13], Eiji Oki[14], Hiroshi Saeki[14], Ken Shirabe[14], Yoshinao Oda[15], Yoshihiko Maehara[14], Shizuo Komune[2], Masaki Mori[3], Yutaka Suzuki[16], Ken Yamamoto[17], Hiroyuki Aburatani[6], Seishi Ogawa[7], Satoru Miyano[4], Koshi Mimori[1‡*]

1 Department of Surgery, Kyushu University Beppu Hospital, Beppu, Japan, **2** Department of Otorhinolaryngology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, **3** Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Suita, Japan, **4** Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, **5** Department of Surgery, Iwate Medical University, Morioka, Japan, **6** Genome Science Laboratory, Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan, **7** Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, Japan, **8** Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan, **9** Department of Surgery, Teikyo University School of Medicine, Tokyo, Japan, **10** Department of Surgery, Omori Red Cross Hospital, Tokyo, Japan, **11** Gastroenterological Center, Department of Gastroenterological Surgery, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan, **12** Department of Surgery, Takano Hospital, Kumamoto, Japan, **13** Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan, **14** Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, **15** Department of Anatomical Pathology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, **16** Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa-shi, Chiba, Japan, **17** Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Japan

☯ These authors contributed equally to this work.
‡ AN and KM also contributed equally to this work.
* aniida@ims.u-tokyo.ac.jp (AN); kmimori@beppu.kyushu-u.ac.jp (KM)

OPEN ACCESS

Citation: Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, et al. (2016) Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. PLoS Genet 12(2): e1005778. doi:10.1371/journal.pgen.1005778

Editor: Marco Gerlinger, Institute of Cancer Research, UNITED KINGDOM

Received: April 10, 2015

Accepted: December 8, 2015

Published: February 18, 2016

Data Availability Statement: All sequence data and array data are available from DDBJ under the accession number PRJDB4515.

Funding: The present study was supported in part by the following grants and foundation; CREST, Japan Science and Technology Agency (JST), the Funding Program for Next Generation World-Leading Researchers (LS094), Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research, grant number 25861199, Grants-in-Aid for Scientific Research on Innovative Areas of Ministry of Education, Culture, Sports, Science, and Technology "Systems Cancer Research" (4201), and

Abstract

Understanding intratumor heterogeneity is clinically important because it could cause therapeutic failure by fostering evolutionary adaptation. To this end, we profiled the genome and epigenome in multiple regions within each of nine colorectal tumors. Extensive intertumor heterogeneity is observed, from which we inferred the evolutionary history of the tumors. First, clonally shared alterations appeared, in which C>T transitions at CpG site and CpG island hypermethylation were relatively enriched. Correlation between mutation counts and patients' ages suggests that the early-acquired alterations resulted from aging. In the late phase, a parental clone was branched into numerous subclones. Known driver alterations were observed frequently in the early-acquired alterations, but rarely in the late-acquired alterations. Consistently, our computational simulation of the branching evolution suggests that extensive intratumor heterogeneity could be generated by neutral evolution.

The MEXT Strategic Programs on Innovative Research "Supercomputational Life Science", and the YASUDA Medical Foundation. The funders had no 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.

Collectively, we propose a new model of colorectal cancer evolution, which is useful for understanding and confronting this heterogeneous disease.

Author Summary

Cancer is heterogeneous disease; each tumor in different patients has different cancer genomes. Furthermore, another level of heterogeneity exists: even a single tumor harbors multiple genetically distinct subclones. This intratumor heterogeneity is presumably one of causes of therapeutic difficulty, and its understanding is clinically necessary. In this study, we investigated intratumor heterogeneity in colorectal cancer by analyzing sample obtained from geographically separated regions of 9 colorectal tumors. Our integrated data analyses combined with computational simulation strongly suggest that, after clonally shared alterations were accumulated by aging, numerous subclones were generated by neutral evolution. Importantly, this view can explain the robustness and evolvability of cancer: therapeutic action inducing an environmental change would convert some of the numerous neutral mutations to driver genes that confer therapeutic resistance. We believe that this study not only provides insights into colorectal cancer pathogenesis, but also constitutes a new basis for designing therapeutic strategies.

Introduction

Cancer is a heterogeneous disease. Recent cancer genomics studies have revealed extensive genetic diversity among patients. Moreover, even a clonal tumor in one patient often harbors multiple subclones. This phenomenon is called intratumor heterogeneity (ITH) and is presumably generated by branching clonal evolution of cancer cells. Understanding of ITH is clinically important, since the existence of multiple subclones presumably boosts the evolutionary adaptation of tumors against therapies, constituting a source of resistant clones [1].

Recently, a multiregional sequencing approach, which sequences DNA sampled from geographically separated regions of a single tumor, has revealed branched evolution and ITH. Yachida et al. [2] investigated the genomic evolution of pancreatic cancer, establishing two categories of mutations: "founder" and "progressor" mutations are present in all regions and a subset of regions, respectively. Founder mutations are assumed to appear in the early phase of clonal evolution. We refer to the clone that has accumulated all the founder mutations as the parental clone (or the most recent common ancestor). The parental clone then branches into subclones by accumulating progressor mutations, which shape ITH.

Several studies employing multiregional exome sequencing have revealed the occurrence of branched evolution and ITH in several other types of cancers, including clear cell renal cell carcinomas and non-small cell lung cancers. ITH of clear cell renal cell carcinomas is characterized by parallel evolution, in which the same driver gene is independently mutated in different branches of evolutionary trees [3]. In contrast, no evidence of parallel evolution has been reported for non-small cell lung cancer [4, 5]. In addition to genetic aberrations, epigenetic aberrations are also a hallmark of cancer; as for DNA methylation, a few groups have also performed multiregional epigenomic analyses [6, 7]. However, the types of cancers that have been subjected to multiregional analyses remain limited, and ITH of genomes and epigenomes has been poorly studied in an integrated way.

In this study, we present genetic and epigenetic analysis of ITH in a series of nine colorectal cancers. Following multiregional sampling, we performed exome sequencing and copy number (CN), methylation, and mRNA expression array profiling. Our integrated analysis revealed not only extensive ITH, but also the evolutionary histories of the nine tumors. Finally, we also performed computational simulation of cancer evolution, which suggested a possible evolutionary principle underlying the extensive ITH.

Results

Multiregional exome sequencing unveils extensive ITH and branched evolution

To study ITH in colorectal cancer, we performed genomic analysis of samples from geographically separated regions from nine colorectal tumors (S1 Table). In this study, we referred to the nine patients by the term "case" and to multiregional samples in each case by the term "sample". From each of the nine tumor, we obtained 5–21 multiregional samples, which were 75 samples in total, together with 9 paired normal mucosa samples (S2 Table). For two cases, samples from liver metastases were obtained. Our multiregional exome sequencing of the nine cases found 16857 mutations in total, for an average of 58–1195 mutations per sample (S3 Table). From these values, the mutation rates for each case were estimated to be 1.57–20.2 mutations per megabase. All cases, except for case 9, fall in a range typical for non-hypermutated colorectal cancer [8]. Mutational profiles obtained from the multiregional sequencing demonstrated high genetic ITH for all nine colorectal tumors (Fig 1).

Each of the multiregional mutation profiles harbored founder and progressor mutations; founder mutations are shared by all regions while progressor mutations are not. We further divided progressor mutations into two subcategories: "unique" and "shared" mutations, which are unique to a single specific sample and shared by multiple but not all samples, respectively. Targeted deep sequencing validated 100% (5068/5068), 93.9% (1745/1857) and 95.4% (1362/1427) of founder, shared, and unique mutations, respectively. We can assume that founder, shared, and unique mutations are acquired in this order during cancer evolution. Applying the maximum parsimony method [9] to the multiregional mutation profiles allowed us to depict the evolutionary trees of the nine tumors (Fig 2). Comparison between the evolutionary trees and geographical positions of each of the samples showed that subclones were generally separated in geographically correlated ways, demonstrating that geographical relations are maintained as the evolution of colorectal cancer proceeds. On the other hand, our analysis of the deep sequencing data revealed that some regions in two cases harbor intermixed subclones from separated regions, which confirmed a recent finding by Sottoriva et al. (S2 Fig) [10].

We found that mutations in well-known driver genes such as *APC*, *KRAS*, and *FBWX7* were acquired as founder mutations during the establishment of the parental clones (Fig 3A). Pathway-level analysis also showed that founder mutations disrupted the WNT and RTK/RAS pathways, consistently with their principal roles in colorectal tumorigenesis (S3 Fig). Once the parental clones were established, these clones branched into subclones by accumulating progressor mutations. We found that mutations in *PIK3CA* recurrently occurred as progressor mutations, suggesting that *PIK3CA* mutations are a late event in the evolution of colorectal cancer. On the other hand, we did not find any evidence of parallel evolution, as has been observed in studies of clear cell renal cell carcinomas [3].

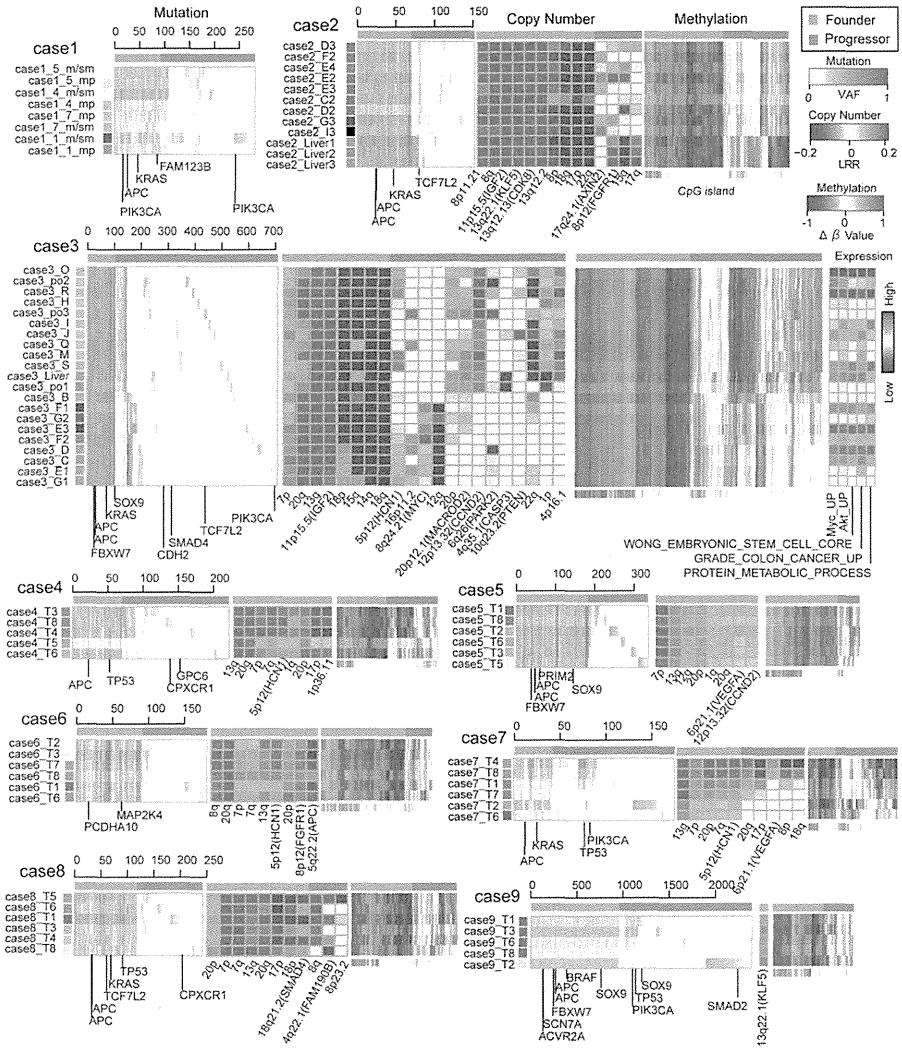

Fig 1. An integrated view of ITH in the 9 colorectal tumors. Multiregional profiles of mutations, CN and methylation alterations were visualized as heat maps. Orange and green bars indicated founder and progressor alterations, respectively. Colored labels for each sample were prepared so that color similarity represents similarity between mutation profiles. For case3, activities of expression signatures were also provided.

doi:10.1371/journal.pgen.1005778.g001

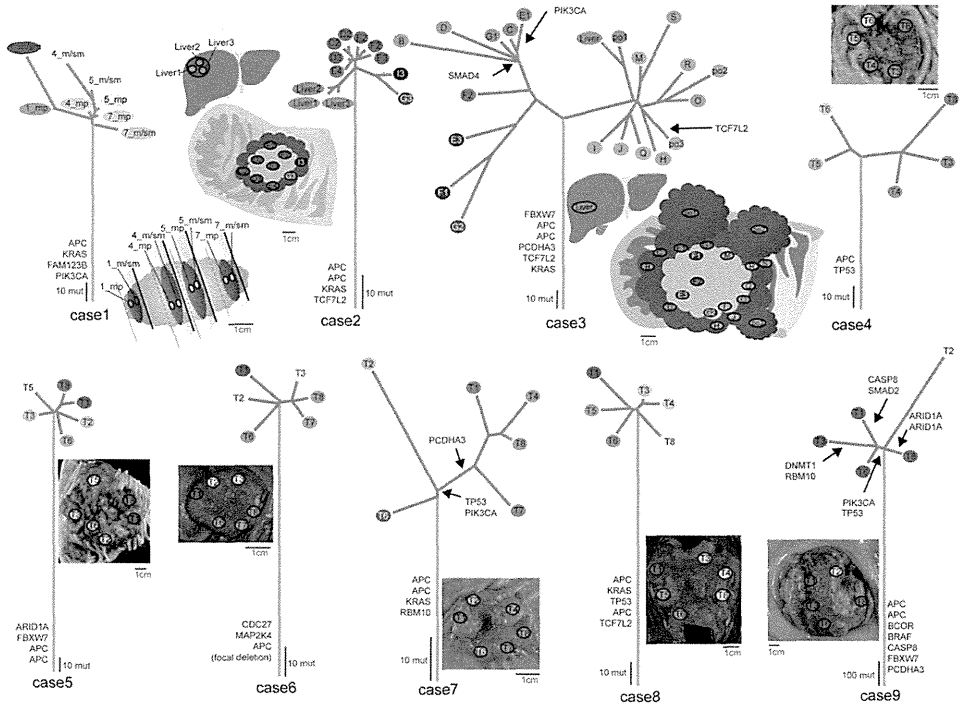

Fig 2. Evolutionary trees of the 9 colorectal tumors. Evolutionary trees inferred from the multiregional mutation profiles have orange trunks, green branches and variously colored leaves, which correspond to founder, progressor mutations and samples, respectively. The leaves were colored based on the color-coding scheme used in Fig 1. Mutation timings of reported driver genes in colorectal cancer were indicated along the trees, and schemas or photos of multiregionally sampled tumors were also provided. Red and blue scales measure tumor size and tree size based on the number of mutations, respectively.

doi:10.1371/journal.pgen.1005778.g002

Analysis of genetic ITH suggests that early-acquired mutations results from aging

We counted each category of mutation and then identified a correlation between the number of founder mutations and the age of the patients (Fig 3B, S4 and S5 Figs). Our findings are consistent with the model of founder mutations accumulating from aging, and similar correlations have also been observed in other types of solid tumors, such as pancreatic cancer and clear cell renal cell carcinomas (S5 Fig). To investigate the temporal signatures embedded in the mutations, we then compared mutational signatures between founder mutations and progressor mutations (Fig 3C). Our analysis showed that C > T transitions at CpG sites are more prominent in founder mutations than in progressor mutations. Next, we calculated the fraction of cancer cells harboring each category of mutation from variant allele frequency (VAF), read depth and CN data (Fig 3D). We found that the cancer cell fractions decreased while proceeding from founder to shared and unique mutations; that is, founder and progressor mutations

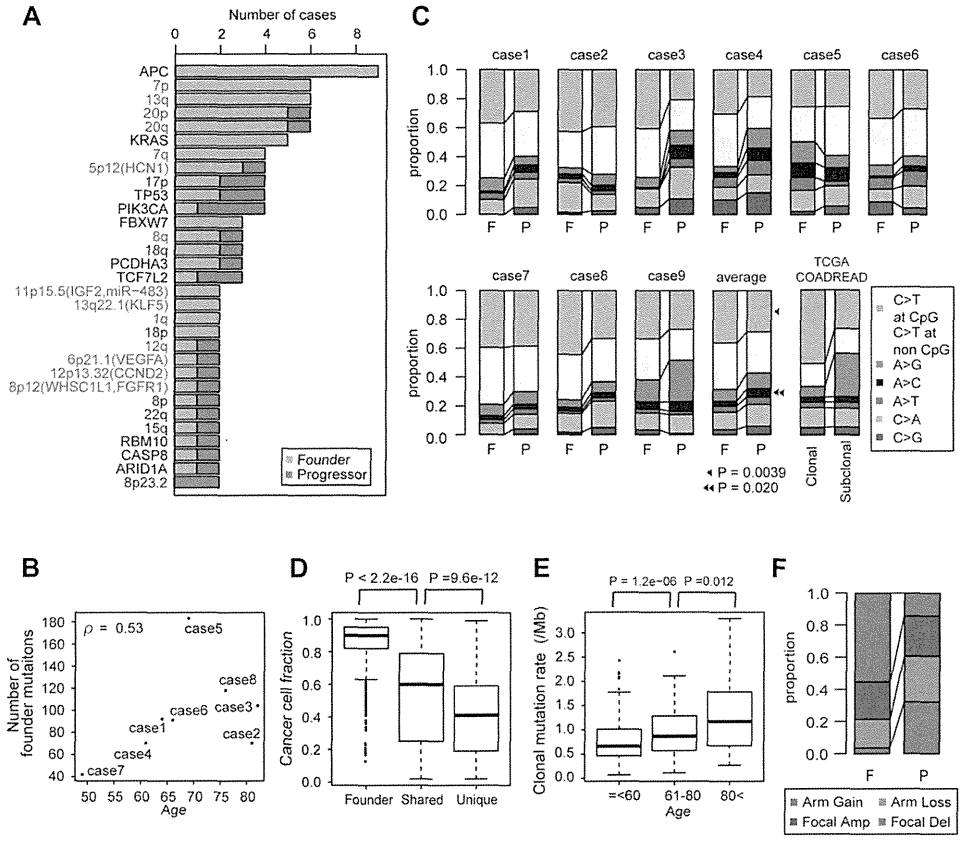

Fig 3. Analysis of genetic ITH. (**A**) The number of samples having mutation (black letters), CN gains (red letters) and losses (blues letters) were counted for each of the founder and progressor categories. The APC item counts one sample subjected to focal deletion. (**B**) Correlation between founder mutation counts and patients' ages in our 8 cases. Case 9 was excluded due to a hypermutation phenotype. ρ is Spearman's correlation coefficients. (**C**) Mutational signatures were calculated from founder (F) and progressor (P) mutations in our 9 cases, and also from clonal and subclonal mutations in non-hypermutated TCGA samples. P-values were calculated by Wilcoxon signed-rank test on the 9 cases. (**D**) Distribution of cancer cell fraction in which founder, shared and unique mutations occur. P-values were calculated by The Wilcoxon rank-sum test. (**E**) Correlation between clonal mutation rates and patients' ages in TCGA non-hypermutated samples. P-values were calculated by The Wilcoxon rank-sum test. (**F**) Proportion of arm-level gain, loss, focal amplification and deletion was calculated for founder and progressor CN alterations in the 8 cases subjected to CN profiling.

doi:10.1371/journal.pgen.1005778.g003

tend to exist as clonal and subclonal mutations in each sample, respectively. In our multiregional sampling, we estimated that the cell population sizes of each sample are about 10^6 from the amount of DNA, while those of each case as a whole are from 10^9 to 10^{10} based on tumor size. This observation suggests that ITH is extremely extensive and the resolution of our multiregional sampling remains insufficient to reveal its totality.

This result also prompted us to estimate founder and progressor mutations in single sample sequencing data (S6 Fig); from The Cancer Genome Atlas (TCGA) colon and rectum adenocarcinoma exome sequencing data [8], we obtained clonal and subclonal mutations as surrogates of founder and progressor mutations, respectively. Using this data, we confirmed that clonal mutations are increased with patients' ages and that they have a higher proportion of C>T transitions at CpG sites than do subclonal mutations (Fig 3C and 3E and S7 Fig). A recent pan-cancer analysis reported that C>T transitions at CpG sites are positively correlated with patients' ages [11]. Consistently, we confirmed that C>T transitions at CpG sites are increased with patients' ages in the TCGA data (S7 Fig). Taken together, the TCGA data analysis supported the hypothesis that founder mutations enriched with C>T transitions at CpG sites were accumulated during aging.

In addition to the exome sequencing, we also performed single-nucleotide polymorphism (SNP) array-based CN profiling for all cases except case 1 (Fig 1 and S9 Fig). The multiregional CN profiles showed that amplifications of 7p, 13q, 10q, 20p, and 20q frequently occurred across all samples in multiple tumors, namely as founder CN alterations (Fig 3A). We also found that, compared with the degrees of mutational ITH, those of CN ITH were variable among cases. For example, cases 2, 3, and 7 showed relatively high CN ITH, which was acquired in a manner that correlated with mutational ITH, as shown by cluster analysis (S9 Fig). In addition to founder CN alterations, we identified CN alterations that occurred along the mutation-based evolutionary trees as progressor CN alterations. We found that arm-level gains tended to occur in founder CN alterations, while focal deletions tended to occur as progressor CN alterations (Fig 3F).

Analysis of epigenetic ITH shows that CpG island hypermethylation occurs early in the evolution

To examine epigenetic ITH, we obtained DNA methylation array data for eight cases. Cluster analysis of the multiregional methylation profiles revealed tight clustering of each case, indicating that intertumor heterogeneity (i.e., heterogeneity among cases) is dominant over ITH (Fig 4A). However, we did observe substantial ITH for each case and, to analyze epigenetic ITH, we focused on variance in methylation levels of each probe in multiregional methylation profiles. Note that the total variance can be decomposed into variance among cases and within cases (hereafter referred as to inter- and intratumor variance, respectively). Notably, we found that different categories of methylation probes contribute differently to the two types of variance (Fig 4B, S10 Fig). We categorized probes based on positional information: relative distances to CpG islands and whether the probes are located in promoter regions. We found that probes within CpG islands tended to show higher intertumor variance than those outside CpG islands. This is consistent with the fact that CpG island hypermethylation marks epigenetic subtypes in colorectal cancer [12]; in our study, cases 5 and 9 appeared to fall into the CpG island methylator phenotype (CIMP) subtype as shown by cluster analysis combined with TCGA samples (S11 Fig). On the other hand, methylation probes outside CpG islands tended to show higher intratumor variance than those within CpG islands. This observation possibly reflects the unstable nature of methylations outside CpG islands, and indicates that methylation alterations outside CpG islands are a main contributor to epigenetic ITH. We also performed similar analysis using classification of chromosomal regions based epigenetic status in normal colon tissue (S12 Fig)[13]. We found that the bivalent domains, which are marked by the H3 lysine 4 and H3 lysine 27 methylation and reportedly involved in cellar differentiation, were associated with high intertumor variance. On the other hand, no clear association existed between any specific region category and intratumor variance, suggesting no functionality of epigenetic ITH.

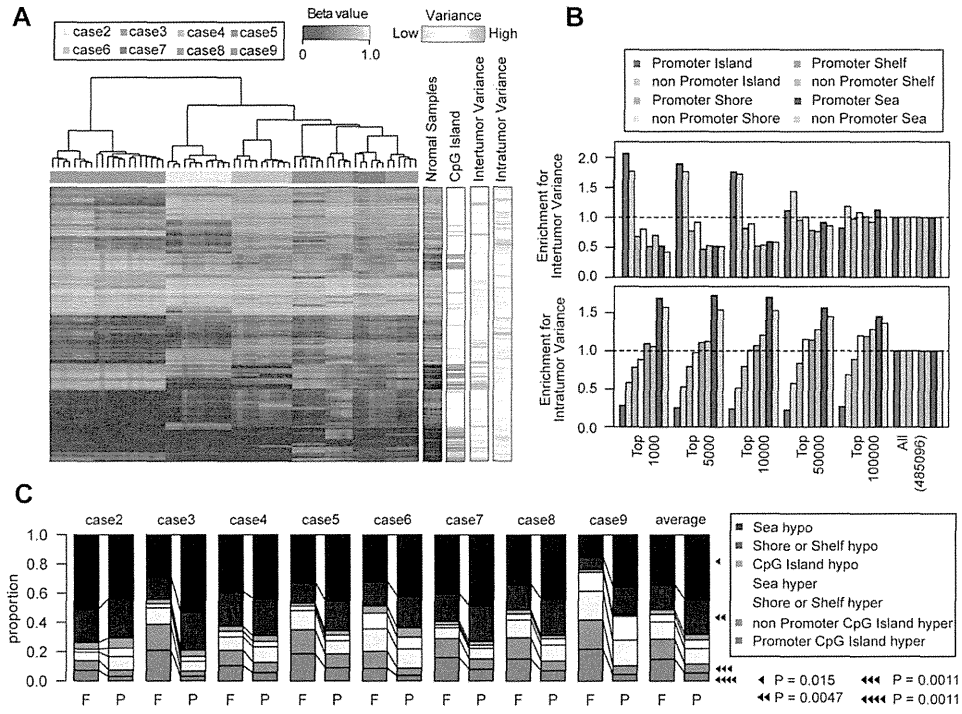

Fig 4. Analysis of epigenetic ITH. (**A**) A heat map of multiregional methylation profiles of the 8 cases. (**B**) Differential contribution of different categories of probe to intratumor and intertumor variance. According to intratumor and intertumor variance, Probes were ranked to obtain each indicated number of top-ranked probes. Probes were categorized based on their genomic positions and enrichment of each category in the top-ranked probes was measured. (**C**) Proportion of indicated types of methylation alterations was calculated for founder (F) and progressor (P) methylation alterations in the 8 cases. P-values were calculated by Wilcoxon signed-rank test on the 8 cases.

doi:10.1371/journal.pgen.1005778.g004

To further investigate ITH and evolution in the colorectal cancer epigenomes, we identified founder and progressor methylations, as done for genetic alterations (S13 Fig). We also differentiated hyper- and hypomethylations by reference to methylation levels in paired normal mucosa samples. Founder methylations are defined as hyper- or hypomethylations that commonly occur across all samples within each case. As expected, we found that the loss of epigenetic gatekeepers such as SFRPs, GATA4, and GATA5 [14] is incurred by founder hypermethylation in many cases (S14 Fig). We also identified hyper- or hypomethylations that occurred along the mutation-based evolutionary trees as progressor methylations. We then deduced temporal signatures from the multiregional methylation data by combining the three types of categorizations: founder and progressor methylation; hyper- and hypomethylation; and positional categories of methylation probes (Fig 4C). Our data showed that hypermethylations in CpG islands were more prominent in founder methylations than in progressor methylations, suggesting that CpG island hypermethylations mainly occur in the early phase of

colorectal cancer evolution. Intriguingly, connections between epigenetic aberration and aging have been reported [15], and our analysis of TCGA data also demonstrated that patients' ages were correlated with the number of hypermethylated probes, but not with that of hypomethylated probes (S15 Fig). Collectively, our results suggest that CpG island hypermethylations, as early events in the evolution of colorectal cancer, result from aging, consistent with our findings regarding mutations. On the other hand, enrichment of hypomethylations in progressor methylations suggests that global hypomethylation starts at a relatively late stage in the evolution, and shapes a substantial part of epigenetic ITH.

An integrated view of ITH tells a cancer's life history

By combining mutation, CN, and methylation profiles, we obtained integrated views of ITH in a series of colorectal cancer samples (Fig 1). For case 3, we additionally obtained mRNA expression profiles. From these integrated views, as well as from evolutionary trees, we can envision the life history of each tumor. Here, we describe that of case 3 as an example. In the founder phase, the parental clone accumulated founder mutations together with CN gain and loss and hyper- and hypomethylation. The founder mutation contains driver mutations represented by mutations in *APC*, *KRAS*, and *FBWX7*. In the progressor phase, the parental clone divided into two major subclones. One major subclone had focal *MYC* amplification, suggesting that this major subclone was shaped by positive natural selection. Although not having a clear driver alteration, the other major subclone had several shared CN alterations, such as 20p amplification and 1p deletion. Then, each of the two major subclones branched into minor subclones, while accumulating many progressor alterations. During this process generating ITH, mutation accumulations, CN alterations, and methylation alterations appeared to occur in a correlated manner. We also found ITH in the transcriptome; notably, the major subclone harboring *MYC* amplification showed upregulation of the *MYC* expression signature together with other signatures related to cancer malignancy. The case 3 tumor also contains a sample from liver metastasis, and the evolutionary tree suggests that the liver metastasis occurred late in the evolution, from a polypoid-like part containing the po1 sample. The metastatic sample of case 3 is contained by the major subclone showing a lower activity of the *MYC* expression signature, which is unexpected if we assume that metastasis results from the acquisition of a malignant phenotype during the evolution. In case 2, which also contains liver metastatic samples, the metastatic samples branched out early in the evolution. Although we need more cases to form a general rule regarding metastasis, our data demonstrate that the multiregional approach is effective to obtain information about the manner in which metastatic clones evolve.

Simulation of cancer evolution suggests that extensive ITH is generated by neutral evolution

As described so far, our genomic analysis revealed a heterogeneous evolution of colorectal cancer. We found that *PIK3CA* mutations and *MYC* amplification occurred in the progressor phase, suggesting that a fraction of ITH is generated by positive natural selection. However, most of the branches in the evolutionary trees had no clear evidence of such positive natural selection, and our clonality analysis of mutations suggests that ITH exists even in each of the multiregional samples (Fig 3D and S6 Fig). To clarify the principle underlying the extensive ITH, we performed computer simulation of a branching evolutionary process (BEP) in cancer evolution (S16 Fig)[16]. In our BEP simulation, cells proliferate while accumulating random mutations in multiple genes. Among the genes, we assume the existence of driver genes whose mutations confer a growth advantage to cells. In the course of tumor growth, cells having

mutations in driver genes are evolutionarily selected and, depending on parameter settings, cells accumulate different combinations of mutations to reproduce ITH. Through a parameter fitting analysis, we found that the BEP simulation can generate heterogeneous mutation profiles similar to the real experimental data, if a high mutation rate, a sufficient number and sufficient strength of driver genes are assumed. (S17 and S18 Figs). We simulated tumor evolution with such a parameter setting and then performed multiregional sequencing of the simulated tumor in silico. Similarly to those of our 9 cases, the simulated multiregional mutation profile harbored founder, shared, and unique mutations while the heterogeneity were well correlated with geographical positions (Fig 5A–5C). Moreover, the VAF of each type of mutation tended to decrease while proceeding from founder to shared and unique mutations (Fig 5D). Note that the VAF is equal to the cancer cell fraction in which the mutation occurs, since the simulated tumors have the haploid genome and no contamination of normal cells. Namely, this result reproduced the local ITH within each of the multiregional samples (Fig 3D). Most importantly, our BEP simulation identified a possible evolutionary principle underlying the extensive ITH. The simulated multiregional mutation profile demonstrated that, while founder mutations occurred in most of the driver genes, progressor mutations rarely occurred in driver genes; namely, most of the progressor mutations were neutral mutations that do not confer a growth advantage (Fig 5E). Taken together with our observation that most of the branches in the evolutionary trees lacked apparent driver alteration, our data suggest that most of the ITH is generated by neutral evolution. It should be noted that our BEP simulation could explain an origin of intertumor heterogeneity. For a probabilistic nature of the model, independent simulation trials even with the same parameter setting generated different multiregional mutation profiles, which remind us of the multiregional mutation profiles unique to each of the nine cases (Fig 1 and S19 Fig).

Discussion

In this study, our integrated multiregional analysis revealed the ITH and evolutionary history of a series of nine colorectal tumors. In particular, by focusing on founder and progressor mutations, we identified clues for decoding the life history of the tumors. For example, we found that founder mutations included established driver mutations such as *APC*, *KRAS*, and *FBWX7*, and their counts correlated with the ages of patients, suggesting that accumulation of alterations in the early phase results from aging.

It is a well-accepted dogma that cancer results from aging [17]. Moreover, associations between somatic mutations and aging have been studied recently. Welch *et al.* [18] found that acute myeloid leukemia (AML) genomes accumulate mutations as a function of age; furthermore, they also confirmed age-dependent mutation accumulation in hematopoietic stem/progenitor cells. Other recent studies report that somatic mosaicism in blood increases in an age-dependent way, and it also has a positive association with cancer risk [19, 20]. Although the association between somatic mutations and aging has been poorly studied in the context of solid tumors, our findings indicate that an association between somatic mutations and patients' ages exists in colorectal cancer. During aging, a colorectal stem/progenitor cell presumably accumulates somatic mutations, some of which could unfortunately be driver mutations that transform the normal cell to a parental clone. This view is also consistent with a recent report that a high division rate of colorectal stem/progenitor cells well explains a high lifetime risk of colorectal cancer [21].

Through mutational signature analysis, we also found that CpG transitions at CpG sites more frequently occur in founder mutations than in progressor mutations. This mutational signature is related to spontaneous deamination of 5-methyl-cytosine at CpG dinucleotides

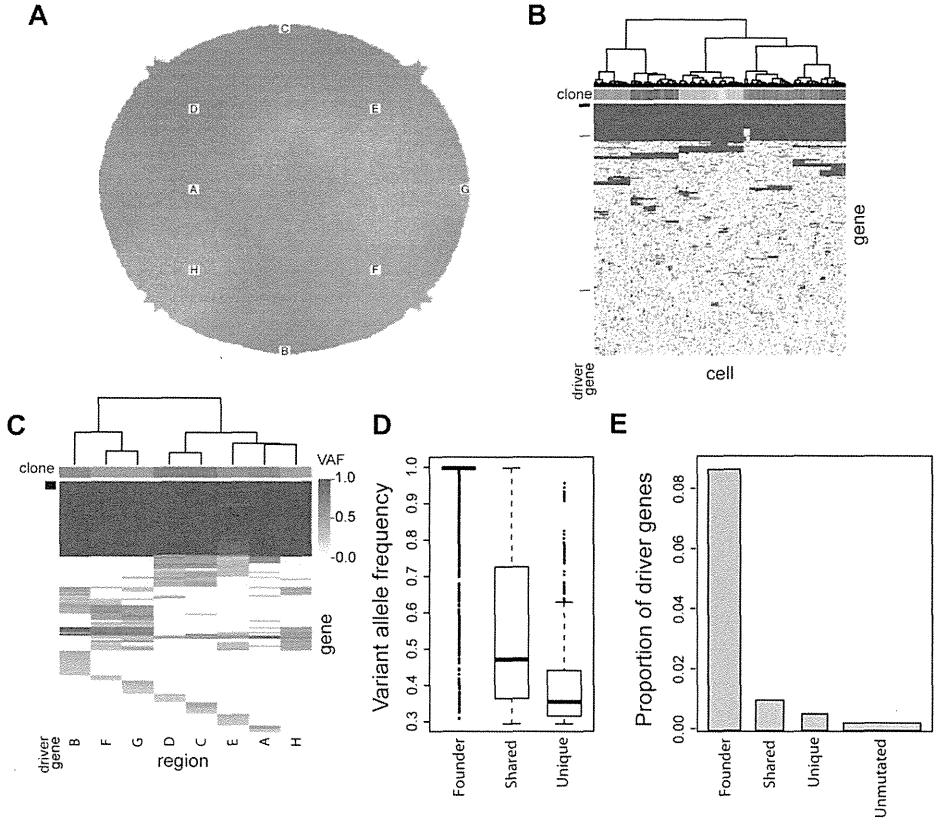

Fig 5. Simulation of cancer evolution. (**A**) A simulated tumor. Different colors represent different clones. White rectangles labeled with alphabets indicate regions subjected to multiregional sampling. (**B**) A simulated single-cell mutation profile matrix. Columns represent 500 cells sampled from the simulated tumor, and the top colored bars label each clone. Rows represent mutated genes and driver genes are indicated by left blue bars. (**C**) A simulated multiregional mutation profile matrix. VAFs of each gene were calculated for cell subpopulations from the 8 regions indicated in (A). (**D** and **E**) Distribution of VAFs (D) and Proportion of driver genes (E) in different categories of mutations. The mutations were obtained from 20 multiregional mutation profile matrices generated by independent simulation trials. In (E), the width of each bar is proportional to the count of each category of mutations. Therefore, the area of each bar is proportional to the count of driver genes that belong to each category of mutations.

doi:10.1371/journal.pgen.1005778.g005

and is most predominantly observed in various cancer types. A recent pan-cancer analysis [11] and our TCGA data analysis showed that this mutational signature is positively correlated with patients' ages, which is consistent with our finding that founder mutations marked by this signature increased with patients' ages. As for DNA methylation, hypermethylation in CpG islands was more prominent in founder methylation than in progressor methylation. We also found that the number of hypermethylated probes is correlated with patients' ages in TGCA samples. Taken together, we speculate that CpG island hypermethylation incurred by aging

also predisposes a colorectal stem/progenitor cell to tumorigenesis in collaboration with somatic mutations.

Thus, genetic and epigenetic alterations are accumulated during aging, and some of them act as driver alterations that transform the normal cell to a parental clone. Once the parental clone is established, it undergoes branched evolution in a geographically consistent way. In addition to ITH of mutations and CN alterations, we found that epigenetic ITH marked by global hypomethylations is prevalent. Our integrated analysis also showed that the genetic and epigenetic ITH are correlated with each other. In contrast to founder alterations, progressor alterations appeared not to have any known driver alterations, with the exception of a few examples such as *PIK3CA* mutation and *MYC* amplification. There also existed no parallel evolution, which is conspicuous in clear cell renal cell carcinomas [3]. Namely, we found little evidence that positive natural selection shaped the extensive ITH, similar to the findings of recent non-small cell lung cancer studies [4, 5]. Moreover, our clonality analysis of mutations suggested that subclones existed even in each of the multiregional samples. It should be noted that such local ITH is consistent with a recent breast cancer study in which single-cell sequencing identified subclonal mutations occurring at low frequencies [22].

In pursuit of the unknown principles generating such extensive ITH, we performed the BEP simulation. Intriguingly, our simulation suggests that neutral evolution can shape extensive ITH as observed in our multiregional mutation profiles. Notably, our simulation also well explained the local ITH within each of the multiregional samples. Although a single-cell mutation profile showed that a simulated tumor actually harbored numerous subclones, snapshots of the simulated evolution suggested that "macroscopic" subclones, which can be captured by the resolution of multiregional sequencing, were generated by genetic drift in the course of the neutral evolution (S20 Fig). A possible mechanism that boosts the neutral mutations is a high mutation rate, as assumed in our simulation. We speculate that genetic instability is incurred and the mutation rate increases before the branched evolution, which is also indicated by the temporal change of mutational signatures. Our computational analysis also suggests that a cancer stem cell hierarchy can boost the neutral evolution [16]. Most importantly, our view that a tumor harbors numerous neutral mutations can explain the robustness and evolvability of cancer. A therapeutic action induces an environmental change, which would convert some of the numerous neutral mutations to driver genes that confer therapeutic resistance. Consistent with this idea, it has recently been reported that resistance to some targeted cancer drugs may result from the outgrowth of preexisting low-frequency subclones [23].

Collectively, this work presents a new model of colorectal cancer evolution; aging leads to the accumulation of genetic and epigenetic alterations in the early phase, while neutral evolution shapes extensive ITH in the late phase (Fig 6). Colorectal cancer has been an attractive subject for studying cancer evolution and its evolution have been addressed from various viewpoints [24–28]. Recently, Sottoriva et al. have also proposed that ITH is mainly shaped by neutral evolution, based on uniformly high ITH, subclonal mixing in distant sites and a power-law distribution of VAFs [10, 29]. Along with these works, this study is unique in that it not only unveiled the extensive ITH, but also explained the underlying principle. We believe that our model not only provides insights into colorectal cancer pathogenesis, but also constitute a new basis for designing therapeutic strategies.

Materials and Methods

Sample collection and preparation

Nine patients who provided written informed consent were enrolled in this study. Detailed information about participants is provided in S1 Table. The study protocol was reviewed and

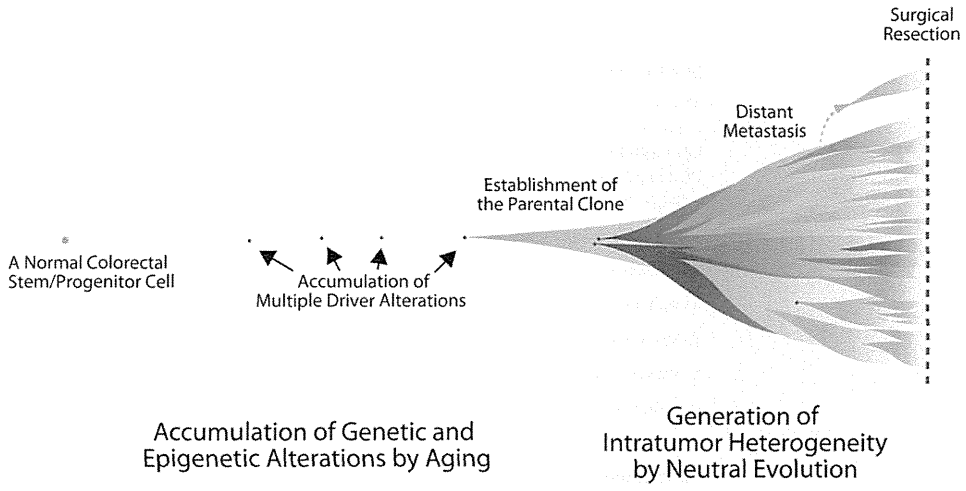

Fig 6. Our model of colorectal cancer evolution. First, founder alterations containing a set of drive alterations are accumulated in the genome and epigenome as a result of aging. After establishment of a parental clone, extensive ITH is generated by neutral evolution, although a few driver alterations are acquired as progressor alterations. Note that this illustration is based on the evolutionary tree of case 3 (Fig 2). However, an actual tumor should harbor numerous subclones, as suggested by the local ITH data (Fig 5D) and simulated single-cell mutation profile (Fig 5B).

doi:10.1371/journal.pgen.1005778.g006

approved by Kyushu University and cooperative institutes. All samples were obtained during surgery from patients who underwent radical resection of primary and/or liver metastatic tumors at Kyushu University Beppu Hospital, Kyushu University Hospital and Osaka University Hospital. DNA and RNA were extracted from fresh frozen multiregional tumor samples and adjacent normal intestinal mucosa with AllPrep DNA/RNA Mini Kit (Qiagen, Hiden, Germany). For case 1–3, high-purity tumor samples were obtained using a laser microdissection system (Leica Laser Microdissection System; Leica Microsystems, Wetzlar, Germany). A detailed protocol of sample preparation was described previously [30].

Ethics statement

The study design was approved by the institutional review boards and ethics committees of hospitals that made the practice of patient (Kyushu University Hospital Institutional Review Board: Protocol Number 486–01, Osaka University Institutional Review Board: Protocol Number 12–27). The study was conducted according to the principles expressed in the Declaration of Helsinki. We obtained written informed consent from all the parents in this study. There was no animal experiment in the study.

Whole exome sequencing (WES)

Whole exome capture was performed with The SureSelct Human All Exon 50Mb kit (Agilent technologies) was used for all samples. The captured targets were subjected to massive

sequencing using HiSeq 2000 (Illumina, San Diego, CA, USA) with the pair end 75–100 bp read option. Information of read depth is provided in S1 Fig and S2 Table.

Mutation calling

The sequence data were processed through an in-house pipeline (http://genomon.hgc.jp/exome/). The sequencing reads were aligned to the NCBI Human Reference Genome Build 37 hg19 with BWA version 0.5.10 using default parameters (http://bio-bwa.sourceforge.net/). PCR duplicate reads were removed with Picard (http://www.picard.sourceforge.net). Mutation calling was performed using the EBcall algorithm [31] with following parameters:

1. Mapping Quality score ≥ 30
2. Base Quality score ≥ 15
3. Both the tumor and normal depths ≥ 8
4. Variant reads in tumors ≥ 4
5. VAFs in tumor samples ≥ 0.08
6. VAF in normal samples ≤ 0.1

Identification of founder and progressor mutations

For each case, the variants that are called in any samples and whose positions have read depths ≥ 10 in all the samples were obtained and their VAFs in all the samples were calculated by mpileup of samtools-0.1.18. Then, variants whose VAF > 0.05 were finally obtained as somatic mutations. This step rescued mutations that were presumably shared among samples, but missed by EBcall due to disagreement with the above parameter conditions, and also filtered out variants that potentially have false negatives in any samples due to low coverage. This procedure was applied for each case to obtain a multiregional mutation profile, from which we identified mutations shared by all the samples as founder mutations, and others as progressor mutations. Progressor mutations were further divided into shared mutations, which were shared by a subset of samples, and unique mutations, which were unique to a single sample. The mutations were annotated by ANNOVAR (http://www.openbioinformatics.org/annovar/). Information of reported driver genes was based on the TCGA colon and rectum adenocarcinoma (COADREAD) study [8]. Information of all the mutations is provided in S3 Table. The multiregional mutation profile obtained for each case is visualized as a heat map whose intensities represent VAFs. In the heat map, founder mutations were ordered along chromosomal positions, shared mutations were ordered by a hierarchical clustering, and unique mutations were sorted for samples and VAFs. From multiregional mutation profiles, maximum parsimony trees were constructed using the maximum likelihood algorithm in the MEGA6 package [9].

Color-coding schemes of sample colors

From the multiregional mutation profile of each case, we also deduced a color-coding scheme to prepare color labels of samples. The multiregional mutation profile were regarded as an $n \times m$ matrix, whose n columns and m rows indexed n mutational positions and m samples, respectively. We applied principle component analysis to the multiregional mutation profile and obtained the first, second and third loading vectors. By multiplying these loading vectors, n-dimensional vectors representing mutational profiles of each sample were reduced into three-dimensional vectors. RGB colors used for sample labels are finally papered by mixing red, green

and blue proportionally to the three vector elements. In a color-coding scheme deduced by this approach, color similarity reflects similarity of mutation profiles between samples.

Validation of the mutations by targeted deep sequencing

We validated WES-derived mutations by targeted deep sequencing. Preamplified cDNA library prepared for WES were captured by a custom-designed SureSelect bait library, which included:

1. All progressor mutations in case2-8.
2. At most 100 nonsynonymous mutations randomly selected from founder mutations in each of case2-9.

Enriched targets were sequenced and Sequencing reads were aligned to the NCBI Human Reference Genome Build 37 as done for WES. After the reads that had either mapping quality of <25, base quality of < 30, or ≥ 5 mismatched bases were excluded, mutation calling was performed using following criteria:

1. Both the tumor and normal depths ≥ 100
2. Fisher's exact test P values < 0.01

Results of the targeted deep sequencing are provided in S3 Table.

CN profiling

DNA was processed and hybridized to the HumanOmniExpress BeadChip Kit (Illumina). Illumina's GenomeStudio software was used to obtain B allele frequencies (BAF) and log R ratios (LRR) from the raw output data. BAF and LRR were input into the ASCAT algorithm [32] to estimate purity and allele-specific absolute CN, which are used for calculation of CCF. Segmented LRR was also obtained from ASCAT and used for subsequent analyses after the median was shift to 0.

Identification of founder and progressor CN alterations

To obtain founder and progressor CN alterations, we focused on chromosomal regions subjected to arm-level and focal alterations recurrent among patients, which were reported by the TCGA study [8]. For all the samples in each case, we obtained LRR averaged along each of the chromosomal regions. We assumed that chromosomal regions subjected to founder CN alterations have $|LRR| > 0.12$ at least in one sample and $|LRR| > 0.06$ in all the samples. To identify progressor CN alterations, we searched for differentially altered chromosomal regions among every pair of sample groups divided by mutation-based evolutionary tree. The groups were prepared by focusing on branching points in the tree. Note that, except for the first branching point that joins the trunk and branches, each branching point divided samples into two groups: those branching out from the point and the others. For each of the chromosomal regions, we obtained difference of mean LRR between every pair of such groups, and the maximum difference as a statistic, ΔLRR. We then obtained chromosomal regions whose $|\Delta LRR| > 0.06$ as those subjected to progressor CN alterations. For founder and progressor CN alterations identified in this manner, if a chromosomal region subjected to a focal CN alteration is contained by that subjected to an arm-level alteration of the same category, we discarded the former.

Calculation of cancer cell fraction

Cancer cell fraction (CCF) in which each mutation occurs was estimated by modifying a previously reported approach [33]. Consider a mutation observed in m of n sequencing reads on a

locus of allele-specific absolute CN of q_a and q_b in a sample of purity α. The expected allele-fraction f of a mutation present in p copy in a fraction c of cancer cells is calculated by $f(c, p, \alpha, q_a, q_b) = \alpha cp(2(1-\alpha)+\alpha(q_a+q_b))$, with $c \in [0.01, 1]$. Then $P(c|n, m, \alpha, q_a, q_b) \propto Binom(m|n, f(c, p, \alpha, q_a, q_b))P(p|q_a, q_b)P(c)$. For calculating $P(p|q_a, q_b)$, we obtained every possible configuration of p from q_a, and q_b, and assumed that they are equally probable. For example, when $q_a = q_b = 1$, $P(p = 1) = 1$. When $q_a = 1$ and $q_b = 2$, $P(p = 1) = 2/3$ and $P(p = 2) = 1/3$, because a mutation could occur in one copy of allele a, one copy or two copies of allele b. $P(c)$ is a prior probably of c and we assumed a uniform prior. $P(c|n, m, \alpha, q_a, q_b)$ was then obtained by calculating these values over a regular grid of 100 c values and reported after normalization. We defined the median of $P(c|n, m, \alpha, q_a, q_b)$ as CCF.

In this study, m ad n were obtained from WES data while α, q_a and q_b were obtained by ASCAT analysis on SNP array data. For deducing Fig 3D, CCFs were calculated for mutations which occur in sample of which $\alpha > 0.6$. CCF is also used for obtaining clonal and subclonal mutations from single sample sequencing data; we assumed mutations of which CCF exceed 0.8 as clonal mutations, and others as subclonal mutations.

Mutation analysis using TCGA data

TCGA WES data were processed by the same pipeline used for our WES data. Because only BAM files were available for TCGA data, we first convert BAM files to fastq files. BAM files for COADREAD paired tumor/normal samples were downloaded from https://cghub.ucsc.edu/ with the accession no. 26594–6, which were first converted back to fastq files, where the 'read1' and 'read2' entries in fastq files were reconstructed based on pair-read information. For reverse strand reads, we generated complemented bases and assigned BaseQuality accordingly. In this process, the reads that had Flag of "not primary alignment", "read fails platform/vendor quality checks" or "supplementary alignment" were discarded using samtools-0.1.18. Then, the fastq file were input into our WES analysis pipeline to call mutations.

SNP 6.0 array data of COADREAD paired tumor/normal samples were also obtained from https://tcga-data.nci.nih.gov/tcga/. The probe-level signal intensities were converted to BAF and LLR as follows: BAF = $I_b/(I_a + I_b)$ and LLR = $I_a + I_b$ where I_a and I_b were signal intensities for A and B alleles. Then, BAF and LLR were input into the ASCAT algorithm to estimate purity and allele-specific absolute CN. In total, 295 COAD and 110 READ TCGA patient samples were available for both the mutation and CN data.

From these data, CCFs were calculated for each mutation to obtain clonal and subclonal mutations. In the analyses of mutational signatures and correlation between patients' ages and mutation rates, 53 COAD and 5 READ samples whose mutation rates $\geq$ 10 Mb were separately analyzed, assuming that they belong to the hypermutated subtype.

Methylation profiling

Genomic DNA was bisulfite treated using the EZ-96 DNA Methylation Kit (Zymo Research Corporation, Orange, CA) and genome-wide DNA methylation were profile with the Illumina HumanMethylation450 BeadChip (Illumina). A methylation level of each probe was then obtained as a β-value. In this study, we focused on only the probes designed for autosomal sequences. In analyses of methylation variance and signature, Positional categories were based on probe information of the HumanMethylation450 BeadChip.

Identification of founder and progressor methylations

For each probes, differential methylation was quantified with each tumor samples and matched normal tissue by the difference of β-values (Δβ = tumor β value–normal β value). For each case,