nature genetics

Recurrent gain-of-function mutations of *RHOA* in diffuse-type gastric carcinoma

Miwako Kakiuchi[1,2], Takashi Nishizawa[3], Hiroki Ueda[1], Kengo Gotoh[1], Atsushi Tanaka[4], Akimasa Hayashi[4], Shogo Yamamoto[1,5], Kenji Tatsuno[1,5], Hiroto Katoh[6], Yoshiaki Watanabe[7], Takashi Ichimura[4], Tetsuo Ushiku[4], Shinichi Funahashi[3], Keisuke Tateishi[2], Ikuo Wada[8], Nobuyuki Shimizu[8], Sachiyo Nomura[8], Kazuhiko Koike[2], Yasuyuki Seto[8], Masashi Fukayama[4], Hiroyuki Aburatani[1,5] & Shumpei Ishikawa[1,4,6]

© 2014 Nature America, Inc. All rights reserved.

Diffuse-type gastric carcinoma (DGC) is characterized by a highly malignant phenotype with prominent infiltration and stromal induction. We performed whole-exome sequencing on 30 DGC cases and found recurrent *RHOA* nonsynonymous mutations. With validation sequencing of an additional 57 cases, *RHOA* mutation was observed in 25.3% (22/87) of DGCs, with mutational hotspots affecting the Tyr42, Arg5 and Gly17 residues in RHOA protein. These positions are highly conserved among RHO family members, and Tyr42 and Arg5 are located outside the guanine nucleotide–binding pocket. Several lines of functional evidence indicated that mutant RHOA works in a gain-of-function manner. Comparison of mutational profiles for the major gastric cancer subtypes showed that *RHOA* mutations occur specifically in DGCs, the majority of which were histopathologically characterized by the presence of poorly differentiated adenocarcinomas together with more differentiated components in the gastric mucosa. Our findings identify a potential therapeutic target for this poor-prognosis subtype of gastric cancer with no available molecularly targeted drugs.

Gastric cancer is one of the most frequent causes of cancer-related mortality in the world[1], with the highest incidence occurring in East Asia, Central and Eastern Europe, and South Africa in particular[2]. Among various types of gastric cancer, DGC represents a subtype with poor prognosis[3–6]; as a result, DGC has gained substantial public attention worldwide. DGC is histologically characterized as a poorly differentiated adenocarcinoma in which single isolated cancer cells or small collective masses of cancer cells massively infiltrate into adjacent tissue in a highly invasive manner with prominent scirrhous stromal reactions. So far, a couple of molecularly targeted drugs have been developed against gastric cancer, including HER2 antagonists. In the intestinal type of gastric cancer, the most common type, approximately 30% of tumors show positivity for HER2 protein expression; however, only fewer than 10% of DGC tumors show positivity for HER2[7,8]. Thus, DGC constitutes a poor-prognosis subgroup of gastric cancer with no known effective molecularly targeted therapies. Recent genomic characterization of gastric cancer by whole-exome sequencing showed that a large number of known cancer-related genes are frequently mutated in gastric malignancies: Wang *et al.* reported frequent mutations in *TP53*, *PTEN*, *ARID1A*, *APC*, *CTNNB1*, *CDH1* and other genes[9], and Zang *et al.* identified somatic mutations in *TP53*, *PI3KCA*, *CTNNB1*, *ARID1A*, *KMT2C* and *FAT4*, among other genes[10]. However, most of the sequenced tumors in these studies were intestinal-type gastric cancers (IGCs). Therefore, DGC-specific molecular carcinogenesis mechanisms and druggable gene targets remain to be elucidated.

To explore driver mutations in DGC development, we performed whole-exome sequencing on 30 clinical samples of DGC that exhibited poorly differentiated cancer cells and/or signet ring cells with prominent scirrhous stroma in histological analysis. To enrich the tumor fraction relative to the dominant stromal component and other normal cells, we performed macrodissection in which areas that had relatively predominant tumor features under histopathological examination were scraped from tissue sections, eliminating normal epithelium components. Even after these procedures, our analysis estimated that the tumor content ratio ranged from 65% at the maximum to less than 20% at the minimum (**Supplementary Fig. 1**). We extracted genomic DNA from cancer tissues and corresponding normal components, and coding exon fragments were enriched by exon capture (SureSelect, Agilent Technologies) according to the manufacturer's protocol. Whole-exome sequencing was conducted using the Illumina HiSeq platform. Sequencing depths for tumors and corresponding normal components were 102× and 99× on averages, respectively, depths that are substantially greater than in typical exome studies, meaning that sufficient numbers of reads for mutant alleles could be obtained,

[1]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [2]Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Forerunner Pharma Research Co., Ltd., Tokyo, Japan. [4]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [5]Translational Systems Biology and Medicine Initiative (TSBMI), The University of Tokyo, Tokyo, Japan. [6]Department of Genomic Pathology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan. [7]Kamakura Research Laboratories, Chugai Pharmaceutical Co., Ltd., Kanagawa, Japan. [8]Department of Gastrointestinal Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. Correspondence should be addressed to S.I. (sish.gpat@mri.tmd.ac.jp).

Received 28 June 2013; accepted 14 April 2014; published online 11 May 2014; doi:10.1038/ng.2984

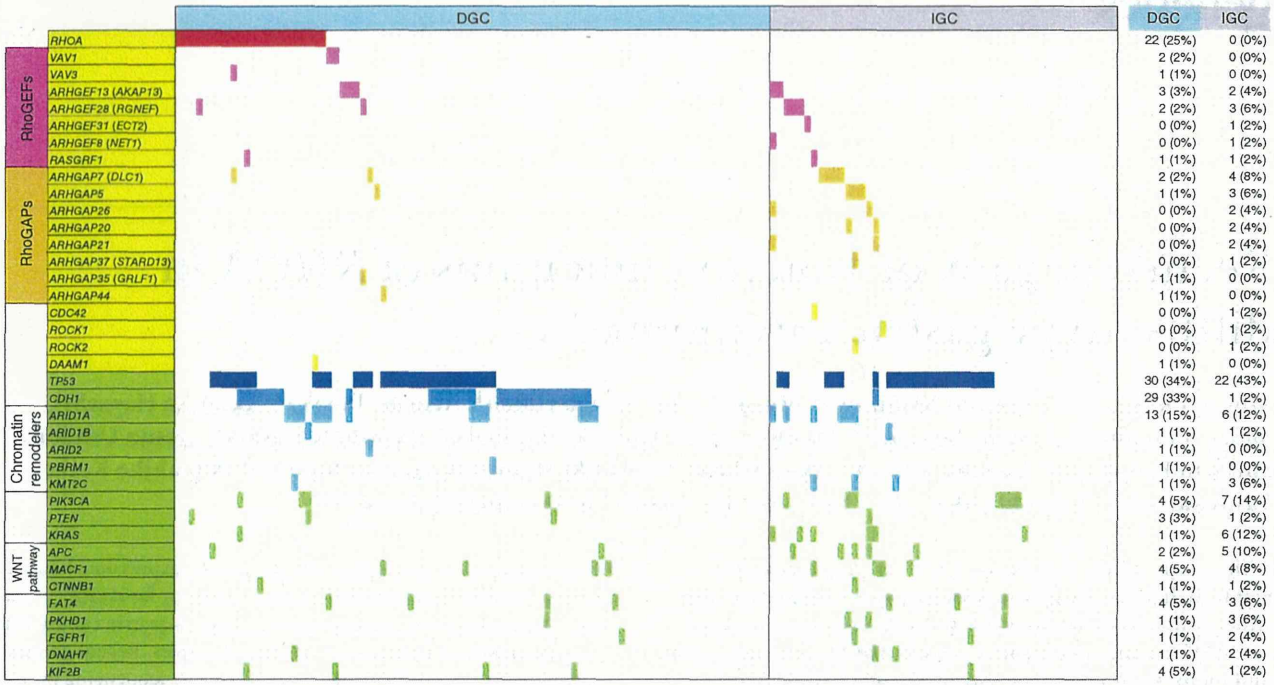

Figure 1 Landscape of genetic changes in two major subtypes of gastric cancer. Overall view of the somatic mutations in 87 DGC and 51 IGC samples. Genes closely linked to *RHOA* (yellow) or frequently mutated in the discovery screen and in previous reports[9,10] (light green) are shown. Colored cells represent somatically mutated genes: for example, *RHOA* (red), RhoGEF genes (pink) and RhoGAP genes (orange). The total numbers and percentages of the mutant cases in each subtype are shown to the right.

even from such stroma-rich background (**Supplementary Fig. 2**). We identified 6,616 somatic mutations (5,359 single-nucleotide variations (SNVs) and 1,257 indels) in the 30 DGC specimens, of which 1,838 occurred as nonsynonymous mutations (1,680 SNVs and 158 indels), with averages of 56 SNVs and 5 indels per case. (Detailed somatic mutation data in this report are available in **Supplementary Data Set 1**.) These somatic mutation frequencies in DGC are comparable to those described for IGC in previous reports. Profiles of mutations and indels suggested that none of the DGCs analyzed in this study had the MSI-H (microsatellite instability high) phenotype (**Supplementary Fig. 1**). Mutational signature analyses showed that the sequence contexts for DGC mutations were distinct from those in IGC (**Supplementary Fig. 3**), and the most characteristic T>G substitution pattern (GTG>GGG or CTT>CGT substitution was relatively frequent) was similar to a known cancer mutation signature reported by a global cancer study[11] (**Supplementary Fig. 3**), although its related etiology is still unclear. Previously known mutations[9,10] were detected in the *TTN* and *TP53* genes in 11 of 30 DGC samples, and *ARID1A* gene mutations were detected in 5 of the DGC samples. We also found nonsynonymous somatic mutations in five DGC cases in the *CDH1* gene, germline mutations of which are known to cause familial DGC[12]. Furthermore, four cases had mutations in genes encoding WNT pathway molecules such as *APC*, *MACF1* (refs. 13,14) and *CTNNB1* (**Supplementary Table 1**).

We observed recurrent somatic mutations in 23% (7/30) of DGC samples in the *RHOA* gene encoding a small GTPase, a finding not yet reported in gastric cancers (**Fig. 1**). Notably, mutation encoding p.Tyr42Cys was detected in four of the seven mutated cases. This biased distribution of mutations implies that these mutations work in a gain-of-function manner. To further investigate the precise frequency and distribution of somatic mutations in *RHOA* and 26 selected *RHOA*-related genes in DGC, we performed targeted sequencing for a validation sample set consisting of 57 independent DGC cases and the 30 cases from the discovery set. Furthermore, to more comprehensively investigate the mutational profiles of the major gastric cancer subtypes, we included 51 IGC cases in our validation screen as well as 19 additional genes (46 genes in total) that were frequently mutated in our discovery screen and have been reported to be altered in previous exome studies (Online Methods and **Supplementary Table 2**). Deep resequencing of the validation set with mean sequencing depths of 4,082× and 4,489× in cancer and corresponding normal components, respectively, identified nonsynonymous somatic mutations in the *RHOA* gene in 25.3% (22/87) of DGC cases (**Fig. 1**). These somatic mutations were unevenly distributed across the *RHOA* gene: mutational hotspots affected the Tyr42, Arg5 and Gly17 residues in the RHOA protein (**Fig. 2a**). Nucleotide substitutions causing p.Gly17Glu and p.Arg5Gln alterations were present in the dbSNP database (rs11552761 and rs11552758, respectively), but we confirmed that they were somatic in nature in our cohort (**Supplementary Fig. 2**). Other known recurrent mutations were also found in *TP53* in 34% (30/87) of cases, in *CDH1* in 32% (28/87) of cases and in *ARID1A* in 15% (13/87) of cases.

Interestingly, we found that some cases harbored nonsynonymous mutations in RhoGAP (GTPase-activating protein) or RhoGEF (GDP/GTP-exchange factor) genes (**Fig. 1**), suggesting that RHOA in combination with its regulatory molecules is frequently mutated in DGC samples (36%; 31/87 cases in our cohort). Mutations of RhoGAP or RhoGEF genes were not specific to DGC and were also observed in IGC cases. In our comparative mutational analysis of DGC and IGC samples, many of the frequent mutations, such as those in *TP53*, *PIK3CA*, WNT pathway genes (*CTNNB1* and *APC*) and chromatin-remodeling and SWI/SNF complex components (*ARID1A* and *KMT2C*), were commonly found in both DGCs and

© 2014 Nature America, Inc. All rights reserved.

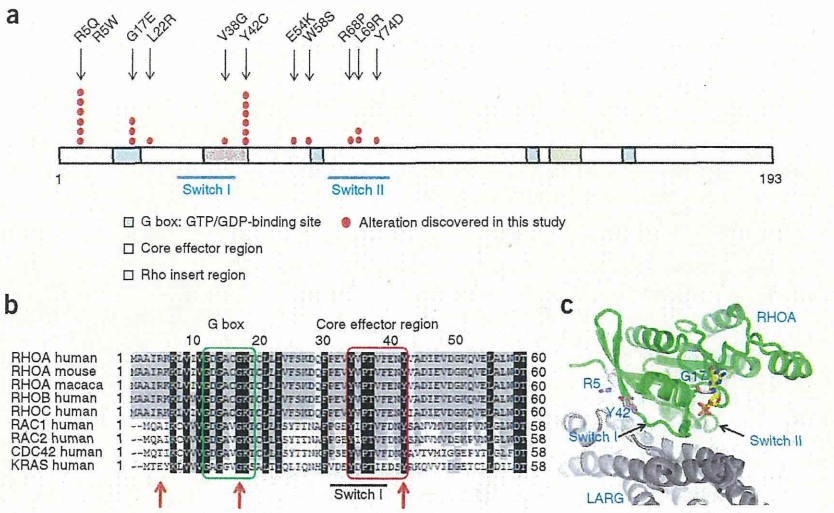

Figure 2 Distribution of RHOA alterations in DGC. (**a**) Map of RHOA functional regions and the sites of the amino acid substitutions discovered in this study. Recurrent alterations are indicated by multiple circles. (**b**) Amino acid alignment of RHOA and RHO family proteins (amino acids 1–60). The G box and core effector region are highly conserved across other mammalian RHOA and human RHO families, as highlighted by green and red boxes, respectively. Red arrows indicate the most frequently mutated positions—Arg5, Gly17 and Tyr42. (**c**) Structure of RHOA and one of its representative RhoGEF proteins, LARG, in its GDP-bound form. RHOA and LARG are shown in green and gray, respectively (Protein Data Bank (PDB) 1X86). Tyr42, one of the most frequently mutated residues in this study, is located on an interaction surface of RHOA with RhoGEFs.

IGCs. *KRAS* mutation was relatively frequent in IGCs. In sharp contrast, the somatic mutations in *RHOA* and *CDH1* were highly specific to the DGC subtype (**Fig. 1**). Frequently affected by mutation, the CDH1 and RHOA signaling pathways are reported to be functionally linked with each other[15,16], indicating that dysregulation around these two signaling pathways would contribute to DGC development (**Supplementary Fig. 4**). Immunohistochemical analyses showed that, of the *RHOA*-mutant DGC cases tested, only a small fraction (4.5%; 1/22) had positivity for HER2 staining (**Supplementary Table 3**), suggesting the possibility of using RHOA as a therapeutic target in DGCs that are not responsive to HER2-targeted therapy. The RHO family of small GTPase proteins consists of approximately 20 members, including RHOA[17], and some family members such as RAC1 have been reported to be mutated in cancers other than gastric cancer[18,19]. The RHO family is clustered as a branch of the KRAS small GTPase superfamily including RAS. The Tyr42, Arg5 and Gly17 residues in RHOA, which were identified as the most frequent sites of alteration in this study, are highly conserved among RHO family proteins (**Fig. 2**). The Tyr42 residue is located within a region called the core effector domain, an important functional domain for physical interaction with effector molecules and/or RhoGEFs and RhoGAPs in the RHOA signaling pathway (**Fig. 2b,c**).

Known biological functions of RHOA are related to actin organization[20], cell migration[21], cytokinesis and the cell cycle[22]. Although somatic mutations of *RHOA* in malignant tumors have not been reported to occur frequently, RHOA overexpression has been found in various cancers, and RHOA is also known to have a key role in tumorigenesis and in tumor cell invasion in various malignancies[23]. To investigate the consequences of the *RHOA* mutations found in DGC tumors, we performed three-dimensional cell growth assays using the OE19 cell line derived from adenocarcinoma of the gastric cardia, which harbors a nucleotide substitution in the *RHOA* gene affecting Tyr42, and the SW948 colon cancer cell line and the BT474 breast cancer cell line, which harbor *RHOA* mutations affecting a conserved Gly17 position located near a region of the protein that is important for its GTPase activity (**Fig. 2c**) (Cancer Cell Line Encyclopedia (CCLE); see URLs). Growth rates for cells from all three cell lines were obviously suppressed in three-dimensional culture by small interfering RNA (siRNA)-mediated knockdown of *RHOA* expression, implying that the *RHOA* mutations provide a strong growth advantage in DGC progression. In sharp contrast, no obvious growth suppression was observed in the AGS and MKN74 gastric cancer cell lines that have wild-type *RHOA* genes (**Fig. 3a,b**). Furthermore, we conducted gain-of-function rescue experiments using siRNA-resistant wild-type or mutant *RHOA* constructs in combination with endogenous *RHOA* knockdown. For these experiments, we used the *RHOA*-mutated SW948 cell line whose growth was strongly inhibited by knockdown of endogenous *RHOA*. Expression of siRNA-resistant Gly17Glu RHOA strongly nullified the inhibitory effect from knockdown of endogenous *RHOA* in SW948 cells, and these cells showed growth comparable to that of control cells (**Fig. 3c**). This phenomenon was replicated by introducing expression of the Tyr42Cys RHOA mutant, indicating that the p.Tyr42Cys and p.Gly17Glu substitutions, although affecting different amino acid positions, have a similar effect on DGC biology. In contrast, expression of siRNA-resistant wild-type RHOA failed to reverse the inhibitory effect from knockdown of endogenous *RHOA* (**Fig. 3c**), indicating that it is specifically the RHOA mutants that have growth-promoting effects. We did not observe similarly clear effects in the AGS and MKN74 cell lines with wild-type *RHOA* genes and in non-transformed HEK293 cells (data not shown), and we therefore speculate that the RHOA mutants work in a context-dependent manner, most likely coupled with other cancer-specific signaling, a hypothesis that is well exemplified by the *KRAS* oncogene[24]. In cancer cells with wild-type *RHOA*, the mutants would have a different mutational context, in which unidentified RHOA-related pathway(s) are possibly activated by different mechanisms. Taken together, our data demonstrate that mutant RHOA, by working in a gain-of-function manner, has a key role in DGC biology (**Supplementary Note**).

In histopathological examination of our cohort of DGC samples, 16 of 22 *RHOA*-mutant cases were classified as advanced gastric cancer with tumor cells invading deeper than the muscle layer, although this invasive feature was also observed in cases with wild-type *RHOA*. Most of these advanced cancers (75%; 12/16) were macroscopically defined as having an appearance consistent with Borrmann type 3 (ref. 25). Histologically, these tumors were defined by the presence of both poorly differentiated cancer components with extensive stromal reactions under submucosal layers and more differentiated tubular components in the mucosa (**Supplementary Fig. 5a–c**), with this heterogeneity possibly reflecting stepwise progression in DGC[26]. To examine the relationship between *RHOA* mutations and cancer progression, we performed microscope-based dissection of histologically distinct features—for example, diffuse and tubular components—from each of these tumors, and extracted DNA was sequenced

© 2014 Nature America, Inc. All rights reserved.

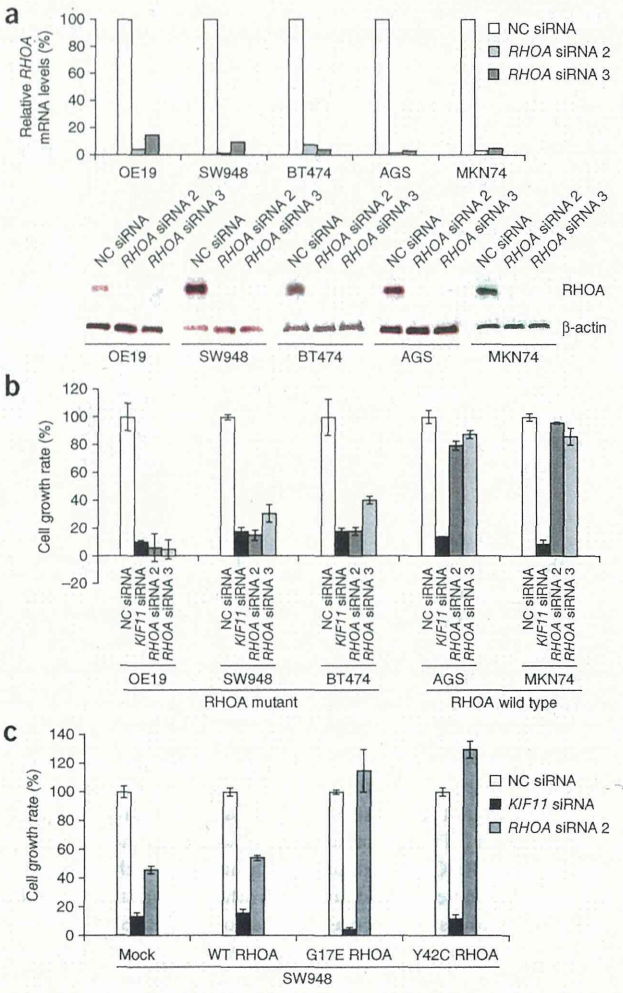

Figure 3 Functional assays for mutant RHOA proteins. (**a**,**b**) Effect of knockdown on mutant RHOA. (**a**) Efficient knockdown of RHOA mRNA and protein in all cell lines used. Top, *RHOA* knockdown was confirmed by quantitative RT-PCR. Levels of *RHOA* mRNA were normalized to *RPS18* mRNA levels (set as 100% for the negative control (NC) baseline states). Bottom, RHOA protein blots showing the efficacy of siRNA in OE19, SW948, BT474, AGS and MKN74 cells; β-actin was used as a loading control. (**b**) Growth inhibition mediated by *RHOA* knockdown in three-dimensional culture. In *RHOA*-mutant OE19, SW948 and BT474 cells, cell growth was inhibited as much as in the positive control cells treated with *KIF11* siRNA, whereas growth inhibition was hardly observed in AGS and MKN74 cells with wild-type *RHOA*. Cell growth inhibition assays were performed in triplicate; data are shown as means ± s.d. Percentages are relative to cells treated with negative control siRNA. (**c**) Gain-of-function rescue experiments for mutant RHOA proteins. SW948 cells were infected with an episome encoding either siRNA-resistant wild-type (WT), Gly17Glu or Tyr42Cys RHOA in combination with siRNA-mediated knockdown of endogenous *RHOA*. Expression of the Gly17Glu and Tyr42Cys mutants in SW948 cells nullified the growth inhibitory effects from knockdown of endogenous *RHOA*. In contrast, expression of plasmid alone (mock) or wild-type RHOA in SW948 cells did not yield such effects. Assays were performed in triplicate; data are shown as means ± s.d. Percentages are relative to cells treated with negative control siRNA.

(**Supplementary Fig. 5d–g**). In all of the samples analyzed, *RHOA* mutations were observed in both classes of tissue, suggesting that *RHOA* mutations might have a key role beginning in the initial stages of cancer progression (**Supplementary Table 4**). In our cohort, we did not find any significant difference in clinical features (stage, lymph node status or extent of metastasis) between DGCs with mutant and wild-type *RHOA*, although this finding should be confirmed by further large-scale investigations.

Previous sequencing analyses of gastric cancer have not reported *RHOA* mutations so far. This may, however, be owing to the paucity of DGC cases in the discovery sample sets of previous studies: with only two or three DGCs in a sample set, it would have been statistically difficult to identify *RHOA* mutations by screening. Furthermore, it is possible that the low tumor content in DGC tissues caused under-representation of mutant sequences in conventional analysis thus far (**Supplementary Figs. 1** and **2**).

In summary, using whole-exome sequencing and subsequent targeted deep sequencing, we herein report recurrent mutations of *RHOA* in DGC (occurring in 25.3% of the cases in our cohort). Furthermore, several functional assays implicated RHOA mutants as oncogenic drivers in DGC progression. The frequency of DGC varies across geographic areas, but DGC comprises around 30–50% of all gastric cancers according to recent reports from European countries and Japan[4,27–29]. However, despite such a relatively high frequency, *RHOA*-mutant gastric cancers cannot be optimal therapeutic applications of any known molecularly targeted antitumor drugs; for example, almost all of the DGC cases with *RHOA* mutation were HER2 negative in our cohort. Taking into consideration its structural characteristics (**Fig. 2**), RHOA could be a druggable oncogenic protein, as it has various targetable domains such as binding pockets for GTP[30] and structural regions of protein-protein interaction with effectors, RhoGAPs and RhoGEFs[31]. By specifically interfering with mutant RHOA and its oncogenic pathways, it might be possible to develop novel therapeutic strategies against otherwise devastating DGCs that undergo massive invasion and metastasis.

URLs. Cancer Cell Line Encyclopedia (CCLE), http://www.broadinstitute.org/ccle/; karkinos genotyper, http://sourceforge.net/projects/karkinos/; BioGRID, http://thebiogrid.org/; Human Protein Reference Database (HPRD), http://www.hprd.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank K. Shiina, K. Sakuma, K. Nakano, A. Nishimoto and R. Sato for technical assistance. This study was supported by Grants-in-Aid for Scientific Research in Innovative Areas to the "Systems Cancer Project" (grant 22134003) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (S.I.), by JSPS (Japan Society for the Promotion of Science) KAKENHI grant 24221011 (H.A.), by a Health Labor Sciences Research Grant (H.A.) and by a grant for the Translational Systems Biology and Medicine Initiative (TSBMI) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (S.Y., K. Tatsuno and H.A.). The supercomputing resource was provided by the Human Genome Center (University of Tokyo).

AUTHOR CONTRIBUTIONS

S.I. and H.A. designed the study. M.K. processed samples and supervised exome sequencing. H.U., K.G., K. Tatsuno and S.Y. performed computational analyses. Y.S., M.F., I.W., N.S., A.H. and S.N. coordinated sample acquisition. H.K., T.I. and T.U. carried out pathological review and analysis. A.T. performed dissection analysis. T.N. and S.F. performed functional experiments. Y.W. carried out structural analysis. S.I. wrote the manuscript. H.A., M.F., K. Tateishi and K.K. were involved in critical review and discussion.

© 2014 Nature America, Inc. All rights reserved.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Hohenberger, P. & Gretschel, S. Gastric cancer. *Lancet* **362**, 305–315 (2003).
2. Guggenheim, D.E. & Shah, M.A. Gastric cancer epidemiology and risk factors. *J. Surg. Oncol.* **107**, 230–236 (2013).
3. Kong, X., Wang, J.L., Chen, H.M. & Fang, J.Y. Comparison of the clinicopathological characteristics of young and elderly patients with gastric carcinoma: a meta analysis. *J. Surg. Oncol.* **106**, 346–352 (2012).
4. Warneke, V.S. *et al.* Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *J. Clin. Oncol.* **29**, 2364–2371 (2011).
5. Chiaravalli, A.M. *et al.* Histotype-based prognostic classification of gastric cancer. *World J. Gastroenterol.* **18**, 896–904 (2012).
6. Stiekema, J. *et al.* Surgical treatment results of intestinal and diffuse type gastric cancer. Implications for a differentiated therapeutic approach? *Eur. J. Surg. Oncol.* **39**, 686–693 (2013).
7. Tanner, M. *et al.* Amplification of HER-2 in gastric carcinoma: association with Topoisomerase IIα gene amplification, intestinal type, poor prognosis and sensitivity to trastuzumab. *Ann. Oncol.* **16**, 273–278 (2005).
8. Bang, Y. *et al.* Pathological features of advanced gastric cancer (GC): relationship to human epidermal growth factor receptor 2 (HER2) positivity in the global screening programme of the ToGA trial. *J. Clin. Oncol.* **27**, 4556 (2009).
9. Wang, K. *et al.* Exome sequencing identifies frequent mutation of *ARID1A* in molecular subtypes of gastric cancer. *Nat. Genet.* **43**, 1219–1223 (2011).
10. Zang, Z.J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44**, 570–574 (2012).
11. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
12. Guilford, P. *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature* **392**, 402–405 (1998).
13. Chen, H.J. *et al.* The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev.* **20**, 1933–1945 (2006).
14. Salinas, P.C. Modulation of the microtubule cytoskeleton: a role for a divergent canonical Wnt pathway. *Trends Cell Biol.* **17**, 333–342 (2007).
15. Anastasiadis, P.Z. *et al.* Inhibition of RhoA by p120 catenin. *Nat. Cell Biol.* **2**, 637–644 (2000).
16. Carothers, A.M. *et al.* Deficient E-cadherin adhesion in C57BL/6J-Min/+ mice is associated with increased tyrosine kinase activity and RhoA-dependent actomyosin contractility. *Exp. Cell Res.* **312**, 387–400 (2006).
17. Wennerberg, K. & Der, C.J. Rho-family GTPases: it's not only Rac and Rho (and I like it). *J. Cell Sci.* **117**, 1301–1312 (2004).
18. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic *RAC1* mutations in melanoma. *Nat. Genet.* **44**, 1006–1014 (2012).
19. Kawazu, M. *et al.* Transforming mutations of *RAC* guanosine triphosphatases in human cancers. *Proc. Natl. Acad. Sci. USA* **110**, 3029–3034 (2013).
20. Mitchison, T.J. & Cramer, L.P. Actin-based cell motility and cell locomotion. *Cell* **84**, 371–379 (1996).
21. Lauffenburger, D.A. & Horwitz, A.F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
22. Olson, M.F., Paterson, H.F. & Marshall, C.J. Signals from Ras and Rho GTPases interact to regulate expression of $p21^{Waf1/Cip1}$. *Nature* **394**, 295–299 (1998).
23. Karlsson, R., Pedersen, E.D., Wang, Z. & Brakebusch, C. Rho GTPase function in tumorigenesis. *Biochim. Biophys. Acta* **1796**, 91–98 (2009).
24. Gidekel Friedlander, S.Y. *et al.* Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. *Cancer Cell* **16**, 379–389 (2009).
25. Borrman, R. in *Handbuch der Speziellen Pathologische Anatomie und Histologie* 4th edn. (ed. Henke, F. & Lubarsch, O.) 812–1054 (Verlag von Julius Springer, Berlin, 1926).
26. Saito, A., Shimoda, T., Nakanishi, Y., Ochiai, A. & Toda, G. Histologic heterogeneity and mucin phenotypic expression in early gastric cancer. *Pathol. Int.* **51**, 165–171 (2001).
27. Henson, D.E., Dittus, C., Younes, M., Nguyen, H. & Albores-Saavedra, J. Differential trends in the intestinal and diffuse types of gastric carcinoma in the United States, 1973–2000: increase in the signet ring cell type. *Arch. Pathol. Lab. Med.* **128**, 765–770 (2004).
28. Hayashi, T. *et al.* The superiority of the seventh edition of the TNM classification depends on the overall survival of the patient cohort: comparative analysis of the sixth and seventh TNM editions in patients with gastric cancer from Japan and the United Kingdom. *Cancer* **119**, 1330–1337 (2013).
29. Kaneko, S. & Yoshimura, T. Time trend analysis of gastric cancer incidence in Japan by histological types, 1975–1989. *Br. J. Cancer* **84**, 400–405 (2001).
30. Shang, X. *et al.* Rational design of small molecule inhibitors targeting RhoA subfamily Rho GTPases. *Chem. Biol.* **19**, 699–710 (2012).
31. Shang, X. *et al.* Small-molecule inhibitors targeting G-protein-coupled Rho guanine nucleotide exchange factors. *Proc. Natl. Acad. Sci. USA* **110**, 3155–3160 (2013).

© 2014 Nature America, Inc. All rights reserved.

ONLINE METHODS

Informed consent and sample preparation. Frozen tissue samples of gastric cancer and paired normal gastric tissue samples were obtained from individuals who underwent gastrectomy at the University of Tokyo Hospital. Informed consent was obtained from all subjects, and this study was approved by institutional review boards at the University of Tokyo. We selected gastric tumor cases that pathologists histologically classified as DGC according to Lauren's classification. Depending on the size of the tissue sample, 15–60 sequential sections were cut at 20-μm thickness using a cryostat (Leica) at −20 °C. Total DNA was extracted from sections using the QIAamp DNA Mini kit (Qiagen).

Exome capture, library construction and sequencing. One microgram of DNA per sample was sheared with a Covaris SS Ultrasonicator. We used a Sciclone NGS workstation (Caliper Life Sciences) for automated library construction. Exome capture was performed with Agilent SureSelect Human All Exon Kit v4 (Agilent Technologies). Each sample was sequenced on an Illumina HiSeq 2000 instrument using a read length of 2 × 100 bp. Image analysis and base calling were performed using the Illumina pipeline with default settings. Summary statistics and data quality metrics for whole-exome sequencing are shown in **Supplementary Table 5**.

Exome sequence processing. Burrows-Wheeler Aligner (BWA)[32] and Novoalign software were used to align sequence reads to the human reference genome GRCh37/hg19. After removal of PCR duplicates, we used SRMA[33] to improve variant discovery through local realignments of short-read next-generation sequencing data. To identify somatic mutations in the low-tumor-content samples, we implemented an original variant caller and a method to estimate tumor cellularity as described below (karkinos genotyper; see URLs). For each sample, tumor cellularity was estimated from allelic imbalance in the matched tumor and normal samples with a program examining the allelic fractions of heterozygous SNPs in regions of loss of heterozygosity (LOH) using an algorism similar to that described in a previous report[34,35]. In some cases where LOH regions were not detected, tumor content ratios were estimated from the distribution of mutant allele frequencies. When both calculations failed to estimate tumor cellularity, we presumed it to be 0.2 for the correction of mutant allele frequencies. Consequently, somatic mutant allele frequencies, adjusted by estimated tumor content ratios, that were ≥15% were retained. Artifacts originating from errors in the sequence and mapping were also filtered out by checking SNV positions and base quality scores for supporting reads. Fisher's exact tests were then used, and SNV candidates with a P value of >0.2 were removed. To eliminate germline variations in this study, we carried out comparative analyses using paired tumor and non-tumor stomach mucosa tissues from the same cases for all of the samples analyzed and we extracted the somatic events detected only in tumor tissues. Annotation of SNVs was performed with ANNOVAR[36].

Gene selection for targeted resequencing. To accurately evaluate the frequency and distribution of somatic mutations, especially in stroma-rich DGCs, targeted resequencing was performed with deeper read coverage. To characterize *RHOA* gene and pathway aberrations, 27 genes were selected that included *RHOA*, RHO modulator genes (RhoGEF, RhoGAP and RhoGDI genes) and effector molecules (BioGRID and the Human Protein Reference Database (HPRD); see URLs). In addition, 19 genes were selected from recurrently mutated genes identified in previous gastric cancer exome studies[9,10] or in our discovery screen, including chromatin-remodeling, cell adhesion and WNT pathway genes. We selected 46 genes in total (listed in **Supplementary Table 2**).

Mutation confirmation using targeted resequencing. To validate somatic mutations in selected genes, we designed 2 panels consisting of 1,504 and 1,525 amplicons (average of 175 bp in length) for 46 genes in total (**Supplementary Table 2**) using DesignStudio for TruSeq Custom Amplicon (TSCA) sequencing (Illumina). Coverage was 98.9%, and 93.5% of the targeted regions were covered by the designed amplicons.

Library preparation for 87 DGC tumors and 51 IGC tumors with matched normal tissues was performed according to the TSCA protocol for Illumina paired-end sequencing. We performed 96-plex deep sequencing for each panel on one flow cell (two lanes) of a HiSeq 2500 instrument in rapid-run mode, using 150-bp paired-end reads. Low-quality parts of the reads, 5 bp from both ends of raw reads, were trimmed off, and the remaining sequences were mapped to the reference hg19 genome using BWA. Initial detection of SNVs and indels was carried out using the Genome Analysis Toolkit (GATK) UnifiedGenotyper with a minimal coverage of 20-fold. Any variants detected in pooled normal data were filtered out as possible germline variants, such as SNPs or errors in PCR amplification and sequencing. Finally, we inspected the mapped reads and called mutations on the Integrative Genomics Viewer (IGV) to confirm the variations. We could not obtain any sequence reads from a part of exon 2 of *RHOA* in several samples, possibly owing to primer synthesis failure; therefore, we performed PCR amplicon sequencing targeting exon 2 of *RHOA* using the Nextera XT DNA sample prep kit (Illumina) according to the manufacturer's protocol and generated 200-bp paired-end reads on a MiSeq system (Illumina). Summary statistics and data quality metrics for TSCA sequencing are shown in **Supplementary Table 5**.

Digital PCR. Digital PCR was carried out using the BioMark nanofluidics system (Fluidigm) with the 12.765 Digital Array. Assay methods were essentially as described[37]. Briefly, reaction mixture was prepared for each assay, containing 1× TaqMan gene expression master mix, 1× *RHOA*wt-FAM and *RHOA*mut-VIC TaqMan assay probes, 1× sample-loading reagent (Fluidigm) and 2 μg of genomic DNA. The *RHOA* DNA fragment was amplified, and, at the end of each PCR cycle, FAM and VIC signals for all chambers were recorded. For each panel, both FAM-positive chambers (wild type) and VIC-positive chambers (mutant) were counted, and the mutant/wild-type copy number ratio was calculated as described[38]. Primer sequences are shown in **Supplementary Table 6**.

Immunohistochemistry for HER2 protein. Paraffin-embedded blocks of gastric tumors were cut into 4-μm sections just before immunohistochemical analyses of HER2 protein. Immunohistochemical staining was performed using the Ventana Benchmark XT autostainer (Ventana Medical Systems) with the labeled streptavidin-biotin peroxidase method, and signals were visualized with 3,3′-diaminobenzidine. The primary antibody used was an antibody to HER2 (4B5, Ventana Medical Systems). Antigen epitopes were retrieved by heating at 100 °C for 60 min with EDTA, pH 8.5 (Ventana Benchmark CC1 standard program). The method by which HER2 immunostaining was scored was based on a previous report[36].

Microscope-based dissection. Sections (10 μm) were sliced from paraffin blocks and mounted on glass slides. Another 4-μm section was sliced for staining with hematoxylin and eosin. After areas with intramucosal tubular components and poorly differentiated cohesive components were examined by microscope, these components were dissected from the corresponding 10-μm sections.

Cell culture. The human SW948, BT474, AGS (obtained from the American Type Culture Collection, ATCC), OE19 (obtained from the European Collection of Cell Cultures, ECACC) and MKN74 (obtained from the Japanese Collection of Research Bioresources, JCRB) cancer cell lines were cultured in Leibovitz's L-15 (Gibco), Hybri-Care Medium (American Type Culture Collection), RPMI-1640 (Sigma), F12K (Gibco) and RPMI-1640 (Sigma), respectively, supplemented with 10% FBS (PAA Laboratories) at 37 °C with 5% CO_2, except for SW948 cells, which were cultured without CO_2. Cells were passaged once every 2–3 d.

siRNA transfection, cell growth inhibition assays and gain-of-function rescue assays. siRNA duplex oligonucleotides against human *RHOA* (*RHOA* siRNAs 2 and 3) and *KIF11* and a non-targeting negative control siRNA (Silencer Select Negative Control No.1 siRNA) were synthesized by Life Technologies (**Supplementary Table 6**). Cells were seeded in 96-well ultra-low-attachment plates (Corning) at a density of 1.0×10^4 cells per well in triplicate wells. At the same time, mixtures of siRNA and Lipofectamine RNAiMAX reagent (Invitrogen) were added to each well as 0.2 or 1 nM siRNA solutions. Under these culture conditions, SW948, BT474, OE19, AGS and MKN74 cells showed spheroid-like growth patterns. ATP molecules were

© 2014 Nature America, Inc. All rights reserved.

doi:10.1038/ng.2984

quantified 5–7 d after transfection of cells with RNA interference (RNAi) duplexes using the CellTiter-Glo Luminescent Cell Viability Assay (Promega), a luminescence-based method to determine the number of viable cells.

For rescue of the *RHOA* knockdown phenotype, SW948 cells expressing siRNA-resistant wild-type and mutant *RHOA* were prepared. *RHOA* siRNAs 2 and 3 recognize the 3′ UTR of *RHOA*, so we introduced the coding sequence of *RHOA* as for siRNA-resistant wild-type and mutant *RHOA* constructs. The coding sequence for *RHOA* was amplified by RT-PCR from AGS cells and inserted into the pEBMulti-Neo vector (Wako). Expression plasmids for Gly17Glu and Tyr42Cys mutant RHOA were generated using site-directed-mutagenesis PCR and the In-Fusion HD Cloning system (Clontech) with plasmid encoding wild-type RHOA. Each plasmid was transfected into SW948 cells by electroporation with the Nucleofector system (Lonza). Cell growth inhibition assays were performed with transfected cells transiently expressing siRNA-resistant wild-type and mutant RHOA.

Quantitative RT-PCR and protein blot analyses. Cells were seeded in 6-well ultra-low-attachment plates (Corning) at a density of 2.5×10^5 cells per well. At the same time, mixtures of siRNA and Lipofectamine RNAiMAX reagent were added to each well as 0.2 or 1 nM siRNA solutions. Two days after transfection, total RNA was extracted using the RNeasy Mini kit (Qiagen). Quantitative RT-PCR was performed with Power SYBR Green PCR Master Mix (Applied Biosystems), using primers for *RHOA* or *RPS18* (primer sequences are listed in **Supplementary Table 6**). Values obtained in quantitative RT-PCR were normalized to those for *RPS18*.

Two days after transfection, cells were also lysed in RIPA buffer (188-02453, Wako) supplemented with a protease inhibitor cocktail (4693132, Roche) and the phosphatase inhibitor PhosSTOP (4906845, Roche), and concentrations for the extracts ware estimated with the DC protein assay (Bio-Rad). Total cell extract (5 μg of protein per lane) was subjected to protein blot analyses. Blotted membranes were probed with rabbit monoclonal antibody to RHOA (2117, Cell Signaling Technology) and with mouse monoclonal antibody to β-actin (A1978, Sigma-Aldrich). Antibodies were diluted by 1:1,000 (RHOA) and 1:4,000 (β-actin).

Structural analysis. The structural model was generated from PDB 1X86, using Discovery Studio (Accelrys) and PyMol (Schrodinger). The Tyr42 and Gly17 residues in RHOA are shown by stick model (magenta, carbon; red, oxygen; blue, nitrogen). GDP is shown by stick model (yellow, carbon; red, oxygen; blue, nitrogen; orange, phosphorus).

32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Homer, N. & Nelson, S.F. Improved variant discovery through local re-alignment of short-read next-generation sequencing data using SRMA. *Genome Biol.* **11**, R99 (2010).
34. Song, S. *et al.* qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* **7**, e45835 (2012).
35. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
36. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
37. Qin, J., Jones, R.C. & Ramakrishnan, R. Studying copy number variations using a nanofluidic platform. *Nucleic Acids Res.* **36**, e116 (2008).
38. Dube, S., Qin, J. & Ramakrishnan, R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS ONE* **3**, e2876 (2008).

© 2014 Nature America, Inc. All rights reserved.

nature COMMUNICATIONS

ARTICLE

Received 30 May 2014 | Accepted 17 Dec 2015 | Published 30 Jan 2015

DOI: 10.1038/ncomms7120

Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity

Akihiro Fujimoto[1,2,*], Mayuko Furuta[1,*], Yuichi Shiraishi[3], Kunihito Gotoh[4], Yoshiiku Kawakami[5], Koji Arihiro[6], Toru Nakamura[7], Masaki Ueno[8], Shun-ichi Ariizumi[9], Ha Hai Nguyen[1,10], Daichi Shigemizu[2], Tetsuo Abe[2], Keith A. Boroevich[2], Kaoru Nakano[1], Aya Sasaki[1], Rina Kitada[1], Kazihiro Maejima[1], Yujiro Yamamoto[1], Hiroko Tanaka[11], Tetsuo Shibuya[11], Tatsuhiro Shibata[12], Hidenori Ojima[13], Kazuaki Shimada[14], Shinya Hayami[8], Yoshinobu Shigekawa[8], Hiroshi Aikata[5], Hideki Ohdan[15], Shigeru Marubashi[4], Terumasa Yamada[4], Michiaki Kubo[16], Satoshi Hirano[7], Osamu Ishikawa[4], Masakazu Yamamoto[9], Hiroki Yamaue[8], Kazuaki Chayama[5,17], Satoru Miyano[3,11], Tatsuhiko Tsunoda[2] & Hidewaki Nakagawa[1]

Intrahepatic cholangiocarcinoma and combined hepatocellular cholangiocarcinoma show varying degrees of biliary epithelial differentiation, which can be defined as liver cancer displaying biliary phenotype (LCB). LCB is second in the incidence for liver cancers with and without chronic hepatitis background and more aggressive than hepatocellular carcinoma (HCC). To gain insight into its molecular alterations, we performed whole-genome sequencing analysis on 30 LCBs. Here we show, the genome-wide substitution patterns of LCBs developed in chronic hepatitis livers overlapped with those of 60 HCCs, whereas those of hepatitis-negative LCBs diverged. The subsequent validation study on 68 LCBs identified recurrent mutations in *TERT* promoter, chromatin regulators (*BAP1*, *PBRM1* and *ARID2*), a synapse organization gene (*PCLO*), *IDH* genes and *KRAS*. The frequencies of *KRAS* and *IDHs* mutations, which are associated with poor disease-free survival, were significantly higher in hepatitis-negative LCBs. This study reveals the strong impact of chronic hepatitis on the mutational landscape in liver cancer and the genetic diversity among LCBs.

[1] Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo 108-8639, Japan. [2] Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. [3] Laboratory of DNA Information Analysis, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. [4] Department of Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka 537-8511, Japan. [5] Department of Medicine & Molecular Science, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. [6] Department of Anatomical Pathology, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. [7] Department of Gastroenterological Surgery II, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan. [8] Second Department of Surgery, Wakayama Medical University, Wakayama 641-8510, Japan. [9] Department of Gastroenterological Surgery, Tokyo Women's Medical University, Tokyo 162-8666, Japan. [10] Genome Analysis Laboratory, Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi Vietnam. [11] Laboratory of Sequence Analysis, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. [12] Division of Cancer Genomics, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. [13] Division of Molecular Pathology, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. [14] Hepatobiliary and Pancreatic Surgery Division, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. [15] Department of Gastroenterological Surgery, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. [16] Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. [17] Laboratory for Digestive Diseases, RIKEN Center for Integrative Medical Sciences, Hiroshima 734-8551, Japan. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.N. (email: hidewaki@ims.u-tokyo.ac.jp) or to T.T. (email: tsunoda@src.riken.jp).

© 2015 Macmillan Publishers Limited. All rights reserved.

Primary liver cancer is the fifth most common cancer and the third leading cause of cancer death worldwide. Virus infection is the most common and strongest aetiological factor for liver cancer development. Pathologically, primary liver cancer can be classified into ~90% hepatocellular carcinoma (HCC), and 5~10% intrahepatic cholangiocarcinoma (ICC), and the combined hepatocellular cholangiocarcinoma (cHCC/CC), representing only a small portion[1–3]. Clinically, ICC and cHCC/CC show much more aggressive behaviour with poorer prognosis than HCC, and no standard treatment currently exists, other than surgical resection[1].

One of the major risk factors for the development of ICC is chronic inflammation of the bile ducts, including chronic infections caused by biliary flukes, primary sclerosing cholangitis and hepatolithiasis[2,4]. Furthermore, recent epidemiological studies recognized that chronic hepatitis associated with viral infection (hepatitis B virus (HBV) and hepatitis C virus (HCV)) is also an important aetiologic factor of ICC, as well as HCC, in Asia[2], and indicated that hepatitis-associated ICC and HCC share a common disease process for carcinogenesis[5]. HCC and ICC have been reported to develop simultaneously in both human and mouse models[5–7] and a combined or mixed phenotype (cHCC/CC) displays intimately mixed components of both hepatocellular and biliary epithelial differentiation[8], as shown in Fig. 1a. The presence of these phenotypes indicates the possibility that some of liver cancers can arise from liver progenitor or liver stem cell, although exact cell origins of ICC and cHCC/CC are still controversial and remain to be elucidated[6,9]. We here define ICC and cHCC/CC, both of which contain varying degrees of biliary epithelial differentiated cells, as liver cancer displaying biliary phenotype (LCB), distinguishing from HCC phenotype (Fig. 1a). Although several genome analyses of HCC and exome studies of ICC have been recently reported[10–20], the whole-genomic aberration signature of LCB and its comparison with HCC has yet to be comprehensively explored. In addition, the influence of aetiological factors, such as chronic hepatitis and virus infection type, on the mutational landscape of primary liver cancer remains unknown and has just begun to be analyzed[21].

To elucidate the molecular features of these liver cancer phenotypes, we compared whole-genome sequencing (WGS) data of 30 LCBs and 60 HCCs, and RNA sequencing data for 25 LCBs and 60 HCCs. We examine the WGS to elucidate the substitution pattern and identify driver genes. Finally, we investigate the tumour heterogeneity to find the genes with clonal mutations by target deep-sequencing analysis. This study demonstrates the first genome-wide comparison of LCBs with and without chronic hepatitis and characterizes their molecular features.

Results

Samples and WGS. In this study, frozen tumour and matched normal tissues were collected from 30 patients with LCB (Table 1) and 60 patients with HCC (Supplementary Table 1). All samples were used for WGS. RNA from 25 LCB and 44 HCC samples, from which high-quality RNA was obtained, was also sequenced. Of the LCB samples, 21 were pathologically classified as typical ICC, 2 as rare type of ICC, cholangiocellular carcinoma (CoCC) and 7 as cHCC/CC. Epidemiologically, 9 were from a HCV-related hepatitis background, 7 from HBV-related hepatitis and 14 were not infected with HBV nor HCV. Among them, 10 cases had normal livers with no pathological feature of chronic hepatitis and fibrosis. RK204 (ICC) and RK209 (HCC) developed metachronously in a single individual as multicentric tumours. Genomic DNA was extracted from the tumour and matched lymphocyte or non-tumour liver samples, and WGS was performed at 40.0x average coverage for tumour samples and 33.0x for matched normal tissue, after removing polymerase chain reaction (PCR) duplicates (Supplementary Table 2). For comparison, we used the somatic mutation data of the 60 HCCs that have been whole-genome sequenced by RIKEN and deposited to the ICGC dataset version 8 released on 2012 March (https://dcc.icgc.org/).

Whole-genome mutational landscape of LCBs. We identified point mutations, short indels, copy number alternations, HBV integration sites and somatic rearrangements using custom algorithms (see Methods). We detected between 345 and 180,117 point mutations per tumour. RK308 had an exceptionally large number of somatic mutations (180,117 point mutations), exhibiting a DNA mismatch-repair deficiency with a homozygous deletion in the *MLH1* gene and a missense mutation (C199R) in the *MSH2* gene, which was previously found in Lynch syndrome patient with DNA mismatch-repair deficiency and proved to disrupt its function[22]. Excluding RK308, the average number of nonsynonymous mutations in the 29 LCBs was 26.2, larger than previously reported for leukaemia, but lower than for HCC, lung cancer and melanoma[23] (Supplementary Fig. 1, and Supplementary Table 3). The number of detected somatic rearrangements varied greatly among samples (0–260) (Supplementary Fig. 2 and Supplementary Dataset 1). Chromothripsis was observed in RK142 and RK316 (Supplementary Fig. 2j and 2ac). HBV integration sites were identified in three LCBs (RK069:cHCC/CC, RK166:cHCC/CC, RK208:ICC) using read-pair information[12,14,15] (Supplementary Table 4 and Supplementary Methods), indicating that HBV-infection and its genomic integration can be involved with the carcinogenesis of LCBs[24,25].

Gene expression patterns of LCBs and HCCs. To gain molecular insight into the differences between LCBs and HCCs, we examined their gene expression profiles (Fig. 1b). RNA-seq analysis on 25 LCBs and 44 HCCs, whose high-quality RNAs were available among the 30 LCBs and 60 HCCs, clustered most LCBs into one group, along with some poorly differentiated HCCs, suggesting that LCBs may be similar to poorly differentiated HCCs (Fig. 1b) and that LCBs may have some progenitor feature similar to poorly differentiated HCCs[7]. Some cHCC/CCs clustered in the LCB group, whereas others were in the main HCC group, which is consistent with their histological combined features. Interestingly, three ICCs with HCC metachronous multicentric tumours were classified within the HCC group (RK204, RK073 and RK137). One CoCC was clustered in the HCC group, which suggests that this particular liver tumour may have an origin similar to that of cHCC/CC[26].

Somatic substitution pattern of LCB and HCC. Next we examined somatic substitution patterns of LCBs and compared with those of 60 HCCs (Fig. 2a–d). The distribution of genome-wide somatic substitution patterns is significantly different from random expectation (χ^2-test; P-value $< 10^{-16}$). In the LCB genomes, the most predominant substitution was C:G to T:A (odds ratio = 2.2, comparison from the assumption of the uniform mutation rate), followed by T:A to C:G (odds ratio = 1.9) and C:G to A:T (odds ratio = 1.3) (Fig. 2b). The distribution of the substitution pattern for LCBs and HCCs was similar (Fig. 2a,b), but the proportion of C:G to T:A was significantly higher in LCBs (Supplementary Fig. 3). To examine the differences between each sample, principal component analysis (PCA) was applied to the somatic substitution patterns. Although most LCB somatic substitution patterns overlapped the 60 HCC cluster, those of the eight LCBs, all of which developed in livers with no evidence of

© 2015 Macmillan Publishers Limited. All rights reserved.

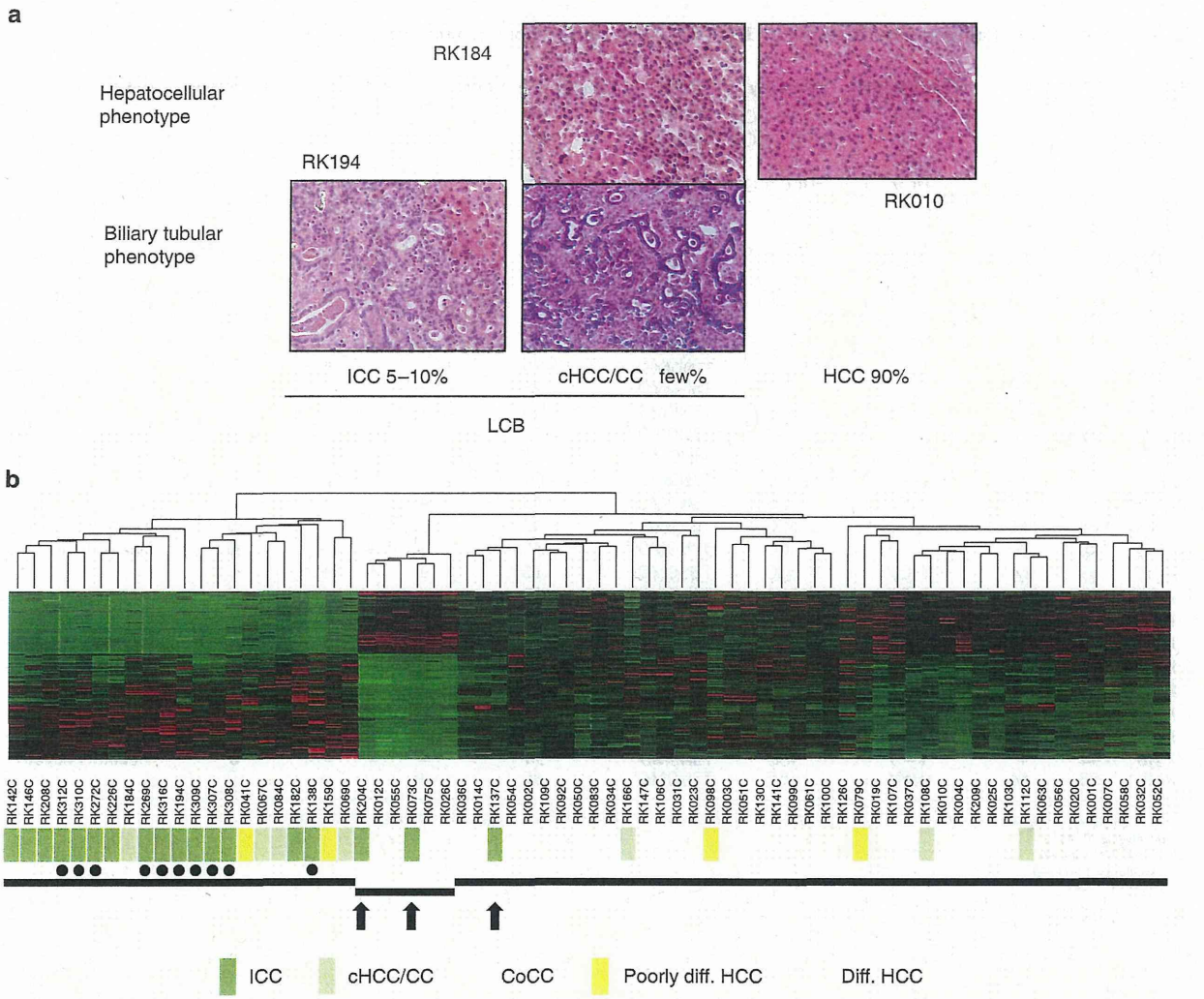

Figure 1 | LCB phenotype and analysis of the transcriptome pattern on the 30 LCBs and 60 HCCs. (**a**) Representative pathological images of ICC, cHCC/CC and HCC. RK194 is intrahepatic well-differentiated cholangiocarcinoma. RK184 is cHCC/CC where some sections show moderately or poorly differentiated cholangiocarcinoma (lower) and some show moderately differentiated HCC (upper). RK010 is moderately differentiated HCC. We define ICC and cHCC/CC, both of which contain varying degrees of biliary tubular-differentiated cells, as liver cancer displaying biliary phenotype (LCB), distinguishing from HCC phenotype. (**b**) Clustering by the transcriptome of 25 LCBs and 44 HCCs. Hepatitis-negative LCBs are indicated by black dots. ICC, cHCC/CC, CoCC (cholangiolocellular carcinoma), poorly differentiated HCC and differentiated HCC are indicated by coloured rectangles. Three ICCs with metachronous MCTs of HCC are indicated by arrows.

chronic hepatitis, diverged (Fig. 2e and Supplementary Fig. 4). To compare the difference between the HCCs, the hepatitis-positive LCBs and the hepatitis-negative LCBs in the PCA, we performed a permutation test. The difference between the hepatitis-positive and -negative LCBs, and between the hepatitis-negative LCBs and the HCCs were significantly larger than those from randomly selected samples after the Bonferroni correction (hepatitis-positive and negative LCBs; P-value = 0.00116, and the hepatitis-negative LCBs and the HCCs; P-value < 0.00001).

To compare the impact of chronic hepatitis and inflammation on the somatic substitution pattern, we performed PCA with several types of cancer[21]. In the PCA plot, cancers strongly influenced by specific mutagens, such as melanoma (UV-exposure) and lung cancer (smoking), were tightly clustered, suggesting that a strong impact of these mutagen exposures on substitution patterns causes a reduction in the divergence among the samples (Fig. 2f). The HCCs, most of which were associated with chronic hepatitis, and the hepatitis-positive LCBs tightly clustered together, whereas the hepatitis-negative LCBs were more spread out (Fig. 2f and Supplementary Fig. 5). This result indicates that chronic inflammation involved with hepatitis strongly influences the somatic substitution pattern (Fig. 2e, f). In addition, the substitution pattern of the hepatitis-negative LCBs was more similar to the recently reported ICCs[18].

To identify the mutational signatures of the hepatitis-positive or -negative LCBs, we used EMu software[27] for the 30 LCBs, and five mutational signatures were detected (Supplementary Fig. 6 and 7). In these signatures, the influence of signature E, which consists of C>T mutations in CpG sites, differed significantly between the hepatitis-positive and -negative LCBs, indicating a potential role for methylated cytosines in carcinogenesis related with chronic hepatitis because C>T transitions preferentially occur in methylated CpG sites (Supplementary Fig. 8).

These findings suggest that the pattern of expressed genes is mainly influenced by cancer type and reflects histological or

© 2015 Macmillan Publishers Limited. All rights reserved.