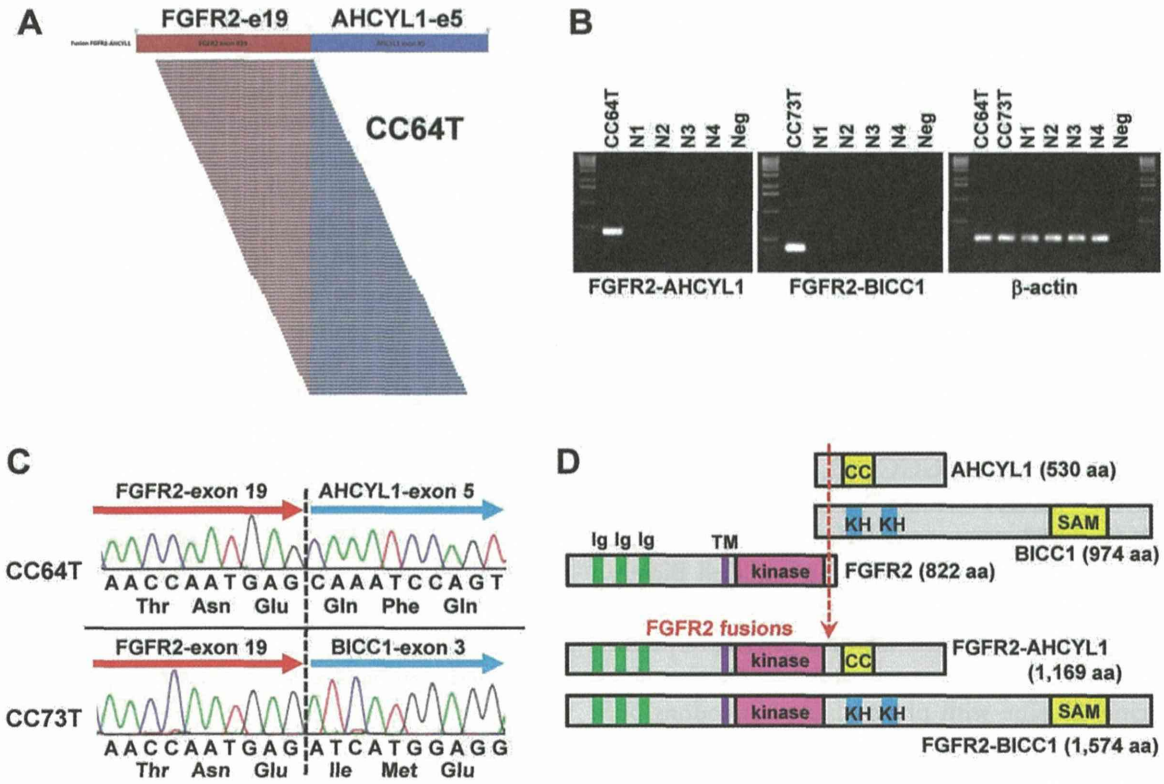

Fig. 1. *FGFR2* fusion genes in cholangiocarcinoma. (A) Junction reads representing *FGFR2-AHCYL1* fusion transcripts in CC64T samples. (B) Confirmation of tumor specific fusion transcripts by RT-PCR. Fusion transcripts were detected only in tumor tissues (CC64T and CC73T), but not in normal liver tissues (N1-N4). Neg: no template. β-Actin expression was used as a control. (C) Sanger sequencing of the RT-PCR product validated in-frame fusion transcripts. (D) Schematic representation of FGFR2-AHCYL1 and FGFR2-BICC1 fusion proteins. Ig: immunoglobulin-like domain, TM: transmembrane domain, kinase: protein tyrosine kinase domain, CC: coiled-coil domain, KH: K homology RNA binding domain, SAM: sterile alpha motif. The dotted vertical line indicates break points.

and *FGFR2-BICC1* encoding 1,169 and 1,574 amino acids, respectively. The chimeric genes consisted of the in-frame fusion of the *FGFR2* amino terminus (exons 1-19) and the *AHCYL1* carboxyl terminus (exons 5-21) or the *BICC1* carboxyl terminus (exons 3-21) (Fig. 1C,D; GenBank/DDBJ accession numbers AB821309 and AB821310). *FGFR2-AHCYL1* is a novel *FGFR2* fusion. *AHCYL1* encodes an S-adenosyl-L-homocysteine hydrolase and inositol 1,4,5-trisphosphate binding protein, and contains a coiled-coil motif in the central domain.[14] *BICC1* encodes an RNA binding protein with a sterile alpha motif (SAM) protein-interaction and dimerization module at the carboxyl terminus.[15] The FGFR2-AHCYL1 and FGFR2-BICC1 fusion proteins are likely to form homodimers through the coiled-coil motif of AHCYL1 and the SAM motif[16] of BICC1, respectively. *FGFR2*, *AHCYL1*, and *BICC1* mapped to chromosome 10q26.1, 1p13.2, and 10q21.1, respectively (Fig. 2A). *FGFR2* and *BICC1* are located on the long arm of chromosome 10 in opposite directions, suggesting that the *FGFR2-BICC1* fusion is generated by intrachromosomal inversion (Supporting Fig. 1B). Gross rearrangement of the *FGFR2* gene locus was verified by FISH with break-apart probes, which showed a split in the signals of the probes flanking the *FGFR2* breakpoint in CC64 and CC73 tumors (Fig. 2B).

Prevalence of FGFR2 Fusions. RT-PCR and Sanger sequencing analysis of 102 cholangiocarcinoma specimens (66 ICCs and 36 ECCs) from Japanese individuals, including eight who had been subjected to whole transcriptome sequencing, identified seven *FGFR2-AHCYL1*-positive and two *FGFR2-BICC1*-positive cases (Table 1; Supporting Table 4). The nine *FGFR2*-fusion-positive cases were ICC type tumors (9/66, 13.6%). *KRAS* mutations were detected in 19 cases (19/102, 17.8%) and *BRAF* mutations in one (1/102, 1%); these mutations were mutually exclusive with the *FGFR2* fusions (Fig. 3A; Supporting Table 4). Although two cases of *FIG-ROS1* fusion (2/23, 8.7%) have been reported by other researchers in cholangiocarcinoma,[10] we did not detect such fusion in this cohort. No significant differences in age, gender, tumor differentiation, clinical stage, and prognosis were detected between fusion-positive and -negative cases. (Table 2, Fig. 3B). Overall survival of ICC cases

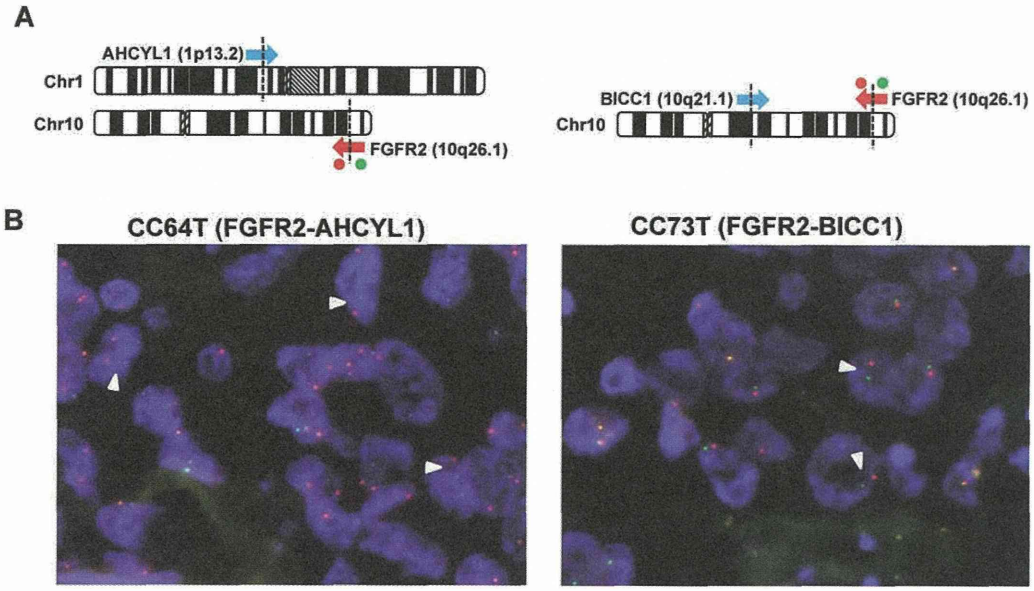

Fig. 2. Detection of *FGFR2* rearrangements. (A) Schematic representation of *FGFR2* gene rearrangements: *FGFR2-AHCYL1* (left) and *FGFR2-BICC1* (right). Arrows indicate the position and direction of the fused genes. Green and red spots indicate the genomic location of 5′ and 3′ FISH probes for the *FGFR2* gene. (B) Representative FISH pattern of *FGFR2* probes in *FGFR2-AHCYL1* and *FGFR2-BICC1*-positive cases. Arrows indicate a split of 5′ green and 3′ red signals.

also showed no great distinction between the two groups. However, fusion positive cases had a propensity for hepatitis virus infection (Table 2). Expression of *FGFR2* mRNA was significantly higher in fusion-positive cases than in fusion-negative ones (Supporting Fig. 2). Especially, *KRAS/BRAF* mutant cases showed reduced *FGFR2* expression. This might afford collateral evidence of mutually exclusive relationships between *FGFR2* fusion and *KRAS/BRAF* mutation. Immunohistological analysis revealed prominent FGFR2 protein expression at both cytoplasm and plasma membrane in fusion-positive cases (Supporting Fig. 3). We further screened 212 gastric cancers, 149 colorectal cancers, and 96 hepatocellular carcinomas by RT-PCR for the presence of these FGFR2 fusion transcripts. The *FGFR2-BICC1* fusion gene was detected in one colorectal cancer (0.7%) and one hepatocellular carcinoma (1.0%). These fusion-positive non-ICC cases were also hepatitis virus-positive (Table 1).

***FGFR2 Fusions Transform NIH3T3 Cells Both* In Vitro *and* In Vivo.** To assess the oncogenic activity of the FGFR2 fusion proteins, stable NIH3T3 clones expressing the retrovirally transfected wild-type fusion proteins or their kinase activity-deficient mutants (KD mutant) were established. As shown in Fig. 4A, wild-type FGFR2-AHCYL1 or FGFR2-BICC1-expressing cells showed anchorage-independent colony formation in soft agar, which was severely suppressed in KD mutant expressing cells. Subcutaneous transplantation of these clones into immunodeficient mice resulted in the formation of tumors from FGFR2-AHCYL1 and FGFR2-BICC1 expressing clones, whereas those expressing KD mutants did not form tumors (Fig. 4B).

To investigate the mechanisms by which the FGFR2 fusion drives oncogenesis, downstream FGFR signaling was analyzed *in vitro* (Fig. 5A; Supporting Fig. 4). The wild-type fusion expressing cells showed constitutive tyrosine phosphorylation in the activation loop of the FGFR kinase domain. FGFR2 signaling activates multiple downstream pathways, including RAS/MAPK and PI3K/AKT.[17] Immunoblot analysis revealed that activation of MAPK, but not AKT or STAT3, was induced in clones expressing FGFR2-AHCYL1 and FGFR2-BHCC1. These results indicate that FGFR2 fusion proteins activate

Table 1. Clinical Features of FGFR2 Fusion Positive Cases

FGFR2 fusion	Gender	Age	Virus status	Pathlogy	Differentiation
FGFR2-AHCYL	F	72	HCV	ICC	mod
FGFR2-AHCYL	F	59		ICC	well
FGFR2-AHCYL	M	62	HCV	ICC	mod
FGFR2-AHCYL	M	73		ICC	well
FGFR2-AHCYL	F	52		ICC	mod
FGFR2-AHCYL	M	59		ICC	well
FGFR2-AHCYL	F	49		ICC	mod
FGFR2-BICC1	M	65	HBV	ICC	mod
FGFR2-BICC1	F	68		ICC	well
FGFR2-BICC1	F	66	HCV	CRC	mod
FGFR2-BICC1	F	46	HBV	HCC	por

ICC: Intrahepatic cholangiocarcinoma
CRC: colorectal cancer
HCC: hepatocellular carcinoma

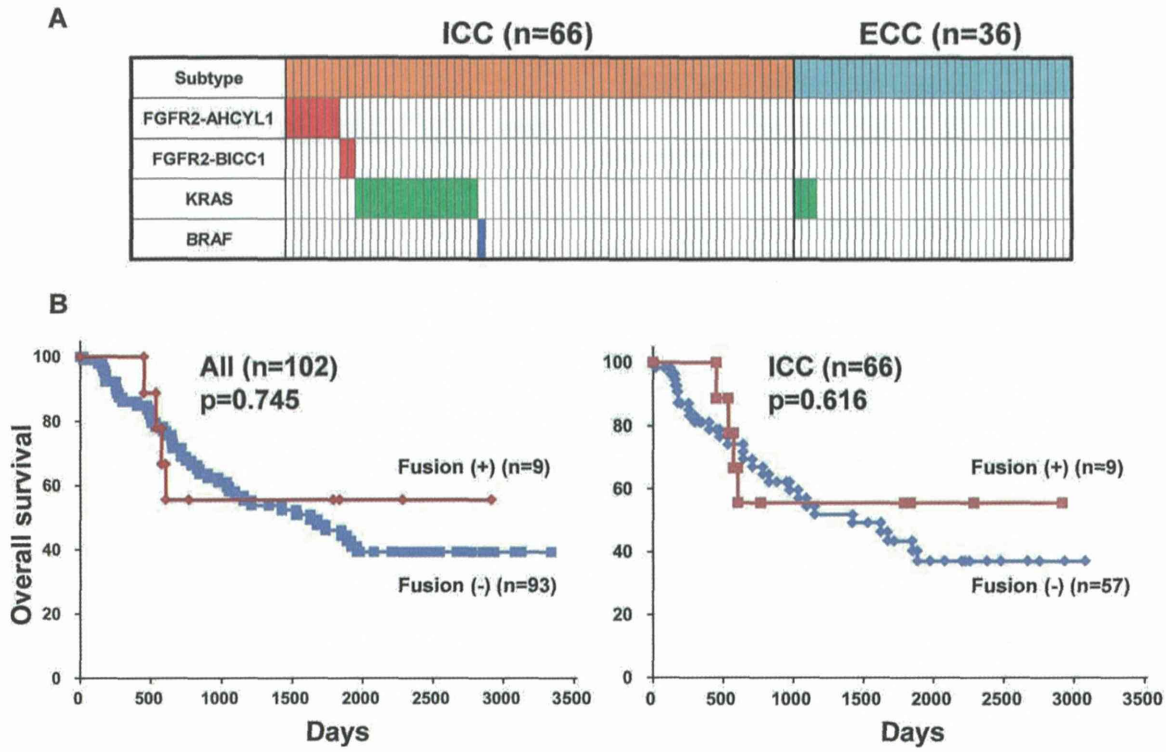

Fig. 3. Clinical subtypes in cholangiocarcinoma. (A) Distribution of genomic alterations. *FGFR2* fusion, *KRAS* mutation, and *BRAF* mutation among ICC and ECC cases are indicated by red, green, and blue, respectively. (B) Overall survival curve stratified by *FGFR2* fusions in all cholangiocarcinoma cases and ICC cases (Kaplan-Meier method). The outcome was not significantly different between *FGFR2* fusion-positive and -negative cases (log-rank test).

canonical FGFR signaling and confer anchorage-independent growth and *in vivo* tumorigenesis, both of which are hallmarks of cellular transformation.

FGFR2 Fusions Are Potential Therapeutic Targets in Cholangiocarcinoma. Next, we examined the sensitivity of *FGFR2* fusion-driven tumor cells to two specific FGFR inhibitors, BGJ398 and PD173074, which selectively inhibit FGFR tyrosine kinase activity.[18,19] These compounds significantly inhibited the phosphorylation of MAPK and reduced *in vitro* anchorage-independent colony formation to the level observed in KD mutant expressing cells (Fig. 5B).

Discussion

FGFR genes are involved in multiple biological processes, ranging from cell transformation, angiogenesis, and tissue repair, to embryonic development. Activating point mutations and amplification of *FGFR* gene members have been explored as therapeutic targets in a wide range of tumors, including bladder, gastric, and lung cancers[20,21]; however, amplification of *FGFR* genes is uncommon in ICC.[22] Diverse fusions involving the *FGFR* gene family have also been reported in hematological and solid cancers[10,11,23,24] and some have shown sensitivity to FGFR inhibition.

The identification of two recurrent *FGFR2* fusions (*FGFR2-AHCYL1* and *FGFR2-BICC1*) that are mutually exclusive with *KRAS/BRAF* mutations warrants a new molecular classification of cholangiocarcinoma and suggests a novel therapeutic approach in cholangiocarcinomas driven by these fusions. Wu et al.[11] recently detected the *FGFR2-BICC1* fusion gene in two cholangiocarcinoma cases, although its prevalence

Table 2. Association Between Clinical Features and FGFR2 Fusion

Clinical factors		Number of fusion positive case	Number of fusion negative case	*P* Value
Gender	Male	4	61	0.207
	Female	5	32	
Age (average)		62.1	66.1	0.104
Virus status	Hepatitis virus positive	3	9	0.035
	Hepatitis virus negative	6	84	
Differentiation	Well	4	21	0.367
	Mod	5	60	
	Poor	0	5	
Stage	I	1	2	0.463
	II	2	14	
	III	2	30	
	IV	4	23	

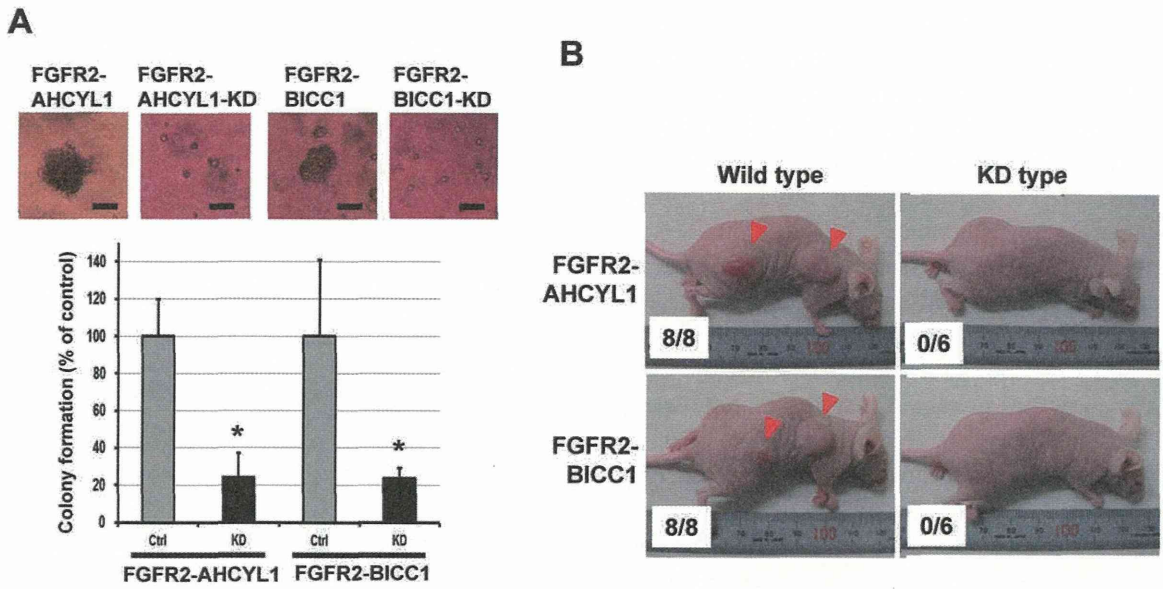

Fig. 4. Oncogenic activity of FGFR2 fusion proteins. (A) Soft agar colony formation in kinase activity-deficient (KD) mutants. The percentage (±SD) of colonies with FGFR2 fusions and their KD mutant transfectants are plotted. $*P < 0.05$. A representative image of colonies expressing wild-type and KD FGFR2 fusions is shown (scale bar = 100 μm). (B) Representative images of mice subcutaneously transplanted with NIH3T3 cells expressing wild-type and KD FGFR2 fusions. The number of tumors per injection in each transfectant is shown.

in cholangiocarcinoma has been lacking. The present study showed a high prevalence of *FGFR2* fusion genes in the intrahepatic subtype of cholangiocarcinoma. Although two cases of another kinase fusion, *FIG-ROS1* (2/23, 8.7%), have been reported by other researchers in CC,[10] we did not detect such fusion in this study. As cholangiocarcinoma is a heterogeneous disease, some epidemiological or clinical specificity may be ascribable to the *FIG-ROS1* fusion. However, no detailed pathological information of the patients was stated in that study. Further investigation is needed to clarify the whole picture of driver fusion genes in CC. Association between *FGFR2* fusion positivity and hepatitis virus infection may suggest an involvement of the virus in the chromosomal rearrangements in CC. However, rare observation of *FGFR2* fusion in hepatocellular carcinoma argues for further analysis of genetic rearrangements.

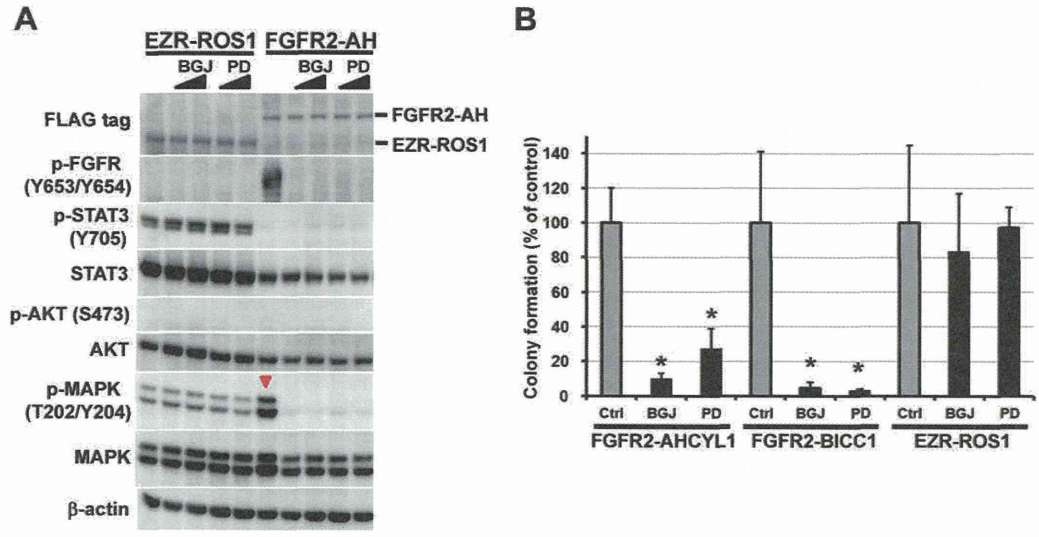

Fig. 5. FGFR inhibitors block signaling in FGFR2-fusion-expressing cells. (A) Activation of FGFR2 and MAPK by FGFR2-AHCYL1 and its suppression by FGFR inhibitors. Lysates from NIH3T3 cells expressing FGFR2-AHCYL1 or EZR-ROS1 (control) treated with vehicle (DMSO), 0.2 and 1 μM BGJ398, and 0.2 and 1 μM PD173074 were immunoblotted with the relevant antibodies. β-Actin was used as a loading control. (B) Anchorage-independent growth of NIH3T3 cells expressing FGFR2 fusions and its suppression by FGFR inhibitors (BGJ: BGJ398 and PD: PD173074). The percentage (±SD) of colonies formed in the presence of FGFR2 inhibitors (0.2 μM) with respect to those formed by DMSO-treated cells are plotted. The NIH3T3 clone expressing EZR-ROS1 was used as a negative control for FGFR inhibitors. $*P < 0.05$.

Overexpression of the FGFR2 fusion protein hyperactivate one of the canonical signaling events downstream of FGFR. This contrast with other FGFR fusion proteins, FGFR1-TACC1 and FGFR3-TACC3 in glioblastoma,[24] which fail to activate canonical downstream MAPK signaling, but induce aneuploidy and oncogenic transformation.[9]

Based on the specific relevant genomic alterations, TKIs have been developed into effective therapies.[7,8] We showed that small molecule FGFR inhibitors, BGJ398 and PD173074, efficiently blocked the downstream signaling and oncogenic activity of ICC-specific *FGFR2* fusions. By the high-throughput cell line profiling assay, amplifications or mutations of *FGFR* genes in cancer cell lines have been reported to predict sensitivity to the selective pan-FGFR inhibitor BGJ398.[25] This drug is currently in a phase I study in patients of advanced solid tumors with FGFR1/2 amplification or FGFR3 mutation (Novartis, Basel, Switzerland; ClinicalTrials.gov identifier: NCT01004224). Clinical investigations, akin to those conducted in other solid tumors with oncogenic fusion kinases, such as *EML4-ALK*,[26] are warranted to examine the efficacy of FGFR inhibitors for the treatment of defined subset of cholangiocarcinoma harboring *FGFR2* fusions.

Acknowledgment: We thank S. Wakai, H. Shimizu, S. Ohashi, W. Mukai, T. Urushidate, and N. Okada of the National Cancer Center for excellent technical assistance.

Author Contributions: Sequencing and data analysis: Y.T., N.H., H.N., F.H.; Molecular biological analysis: Y.A., F.H.; Clinical and pathological analysis: T.Shirota., H.O., K.F., K.S., T.O., T.K.; Article writing: Y.A., Y.T., F.H., T.S.; Study design: Y.A., Y.T., T. Shibata.

References

1. Razumilava N, Gores GJ. Classification, diagnosis, and management of cholangiocarcinoma. Clin Gastroenterol Hepatol 2013;11:13-21.
2. Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD. Cholangiocarcinoma. Lancet 2005;366:1303-1314.
3. Ong, CK, Subimerb C, Pairojkul C, Wongkham S, Cutcutache I, et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. Nat Genet 2012;44:690-693.
4. Voss JS, Holtegaard LM, Kerr SE, Barr Fritcher EG, Roberts LR, Gores GJ, et al. Molecular profiling of cholangiocarcinoma shows potential for targeted therapy treatment decisions. Hum Pathol 2013; 44:1216-1222.
5. Hezel AF, Zhu AX. Systemic therapy for biliary tract cancers. Oncologist 2008;13:415-423.
6. Aljiffry M, Abdulelah A, Walsh M, Peltekian K, Alwayn I, Molinari M. Evidence-based approach to cholangiocarcinoma: a systematic review of the current literature. J Am Coll Surg 2009;208:134-147.
7. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007;7:233-245.
8. Ablain J, Nasr R, Bazarbachi A, de The H. The drug-induced degradation of oncoproteins: An unexpected Achilles' heel of cancer cells? Cancer Discov 2011;1:117-127.
9. Gerber DE, Minna JD. ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. Cancer Cell 2010;18:548-551.
10. Gu TL, Deng X, Huang F, Tucker M, Crosby K, Rimkunas V, et al. Survey of tyrosine kinase signaling reveals ROS kinase fusions in human cholangiocarcinoma. PLoS One 2011;6:e15640.
11. Wu YM, Su F, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao X, et al. Identification of targetable FGFR gene fusions in diverse cancers. Cancer Discov 2013; 3:636-647.
12. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, et al. KIF5B-RET fusions in lung adenocarcinoma. Nat Med 2012;18: 375-377.
13. Arai Y, Totoki Y, Takahashi H, Nakamura H, Hama N, Kohno T, et al. Mouse model for ROS1-rearranged lung cancer. PLoS One 2013; 8:e56010.
14. Ando H, Mizutani A, Matsu-ura T, Mikoshiba K. IRBIT, a novel inositol 1,4,5-trisphosphate (IP3) receptor-binding protein, is released from the IP3 receptor upon IP3 binding to the receptor. J Biol Chem 2003; 278:10602-10612.
15. Wessely O, Tran U, Zakin L, De Robertis EM. Identification and expression of the mammalian homologue of Bicaudal-C. Mech Dev 2001;101:267-270.
16. Kim CA, Bowie JU. SAM domains: uniform structure, diversity of function. Trends Biochem Sci 2003;28:625-628.
17. Brooks AN, Kilgour E, Smith PD. Molecular pathways: fibroblast growth factor signaling: a new therapeutic opportunity in cancer. Clin Cancer Res 2012;18:1855-1862.
18. Skaper SD, Kee WJ, Facci L, Macdonald G, Doherty P, Walsh FS. The FGFR1 inhibitor PD 173074 selectively and potently antagonizes FGF-2 neurotrophic and neurotropic effects. J Neurochem 2000;75:1520-1527.
19. Guagnano V, Furet P, Spanka C, Bordas V, Le Douget M, Stamm C, et al. Discovery of 3-(2,6-dichloro-3,5-dimethoxy-phenyl)−1-{6-[4-(4-ethyl-piperazin-1-yl)-phenylamino]-pyrimidin-4-yl}-1-methyl-urea (NVP-BGJ398), a potent and selective inhibitor of the fibroblast growth factor receptor family of receptor tyrosine kinase. J Med Chem 2011;54:7066-7083.
20. Turner N, Grose R. Fibroblast growth factor signaling: from development to cancer. Nat Rev Cancer 2010;10:116-129.
21. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2010;2:62ra93.
22. Sia D, Hoshida Y, Villanueva A, Roayaie S, Ferrer J, Tabak B, et al. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. Gastroenterology 2013; 144:829-840.
23. Chase A, Grand FH, Cross NC. Activity of TKI258 against primary cells and cell lines with FGFR1 fusion genes associated with the 8p11 myeloproliferative syndrome. Blood 2007;110:3729-3734.
24. Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. Science 2012;337:1231-1235.
25. Guagnano V, Kauffmann A, Wöhrle S, Stamm C, Ito M, Barys L, et al. FGFR genetic alterations predict for sensitivity to NVP-BGJ398, a selective pan-FGFR inhibitor. Cancer Discov 2012;2:1118-1133.
26. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 2010;363:1693-1703.

ARTICLES

nature genetics

Trans-ancestry mutational landscape of hepatocellular carcinoma genomes

Yasushi Totoki[1,14], Kenji Tatsuno[2,14], Kyle R Covington[3,14], Hiroki Ueda[2], Chad J Creighton[3,4], Mamoru Kato[1], Shingo Tsuji[2], Lawrence A Donehower[5], Betty L Slagle[5], Hiromi Nakamura[1], Shogo Yamamoto[2], Eve Shinbrot[3], Natsuko Hama[1], Megan Lehmkuhl[3], Fumie Hosoda[1], Yasuhito Arai[1], Kim Walker[3], Mahmoud Dahdouli[3], Kengo Gotoh[2], Genta Nagae[2], Marie-Claude Gingras[3], Donna M Muzny[3], Hidenori Ojima[6], Kazuaki Shimada[7], Yutaka Midorikawa[8], John A Goss[9], Ronald Cotton[9], Akimasa Hayashi[2,10], Junji Shibahara[10], Shumpei Ishikawa[10], Jacfranz Guiteau[9], Mariko Tanaka[10], Tomoko Urushidate[1], Shoko Ohashi[1], Naoko Okada[1], Harsha Doddapaneni[3], Min Wang[3], Yiming Zhu[3], Huyen Dinh[3], Takuji Okusaka[11], Norihiro Kokudo[12], Tomoo Kosuge[7], Tadatoshi Takayama[8], Masashi Fukayama[10], Richard A Gibbs[3], David A Wheeler[3], Hiroyuki Aburatani[2] & Tatsuhiro Shibata[1,13]

Diverse epidemiological factors are associated with hepatocellular carcinoma (HCC) prevalence in different populations. However, the global landscape of the genetic changes in HCC genomes underpinning different epidemiological and ancestral backgrounds still remains uncharted. Here a collection of data from 503 liver cancer genomes from different populations uncovered 30 candidate driver genes and 11 core pathway modules. Furthermore, a collaboration of two large-scale cancer genome projects comparatively analyzed the trans-ancestry substitution signatures in 608 liver cancer cases and identified unique mutational signatures that predominantly contribute to Asian cases. This work elucidates previously unexplored ancestry-associated mutational processes in HCC development. A combination of hotspot *TERT* promoter mutation, *TERT* focal amplification and viral genome integration occurs in more than 68% of cases, implicating *TERT* as a central and ancestry-independent node of hepatocarcinogenesis. Newly identified alterations in genes encoding metabolic enzymes, chromatin remodelers and a high proportion of mTOR pathway activations offer potential therapeutic and diagnostic opportunities.

HCC is the third leading cause of cancer deaths worldwide[1,2]. Epidemiologically, the incidence of HCC shows marked variance across geographical regions and ancestry groups and between the sexes[3]. HCC incidence predominates in East Asia and Africa, and rapid increases in prevalence have occurred in Western countries[2]. Multiple etiological cofactors are associated with liver cancer, and their contributions might additionally differ according to ancestry. Hepatitis B virus (HBV) infection is dominant in East Asia and Africa, whereas hepatitis C virus (HCV) infection among HCC cases is frequent in Japan. Aflatoxin B1 exposure is a strong risk factor of HCC in China and Africa, whereas alcohol intake is a major etiological factor for HCC in Western countries[3–5]. The average male/female ratio for HCC incidence is greater than two, which could be owing to different environmental exposures or hormone levels[6]. Overlapping but partially distinctive epidemiological backgrounds, such as liver fluke infection, were associated with intrahepatic cholangiocarcinoma (IHCC), another type of liver cancer[5]. Here we conducted the first trans-ancestry HCC genome sequencing research under the umbrella of the International Cancer Genome Consortium (ICGC)[7] and The Cancer Genome Atlas (TCGA)[8]. Thus far, this study represents the largest genomic profiling of liver cancers (608 cases) and compares ancestry groups (Japanese, Asian and European) with distinctive etiological cofactors. This genome data set also uncovers an extensive landscape of driver genetic alterations in HCC.

RESULTS

Whole-exome and oncovirome sequencing of liver cancers

As an ICGC liver cancer project, we collected 503 pairs (413 cases in the Japanese cohort and 90 cases in the US cohort) of liver cancers (488 HCC and 15 IHCC) and matched non-cancerous liver tissues

[1]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan. [2]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [4]Department of Medicine, Baylor College of Medicine, Houston, Texas, USA. [5]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA. [6]Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo, Japan. [7]Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Tokyo, Japan. [8]Department of Digestive Surgery, Nihon University School of Medicine, Tokyo, Japan. [9]Department of Surgery, Baylor College of Medicine, Houston, Texas, USA. [10]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [11]Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital, Tokyo, Japan. [12]Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [13]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [14]These authors contributed equally to this work. Correspondence should be addressed to D.A.W. (wheeler@bcm.edu), H.A. (haburata-tky@umin.ac.jp) or T.S. (tashibat@ncc.go.jp).

Received 31 December 2013; accepted 3 October 2014; published online 2 November 2014; doi:10.1038/ng.3126

ARTICLES

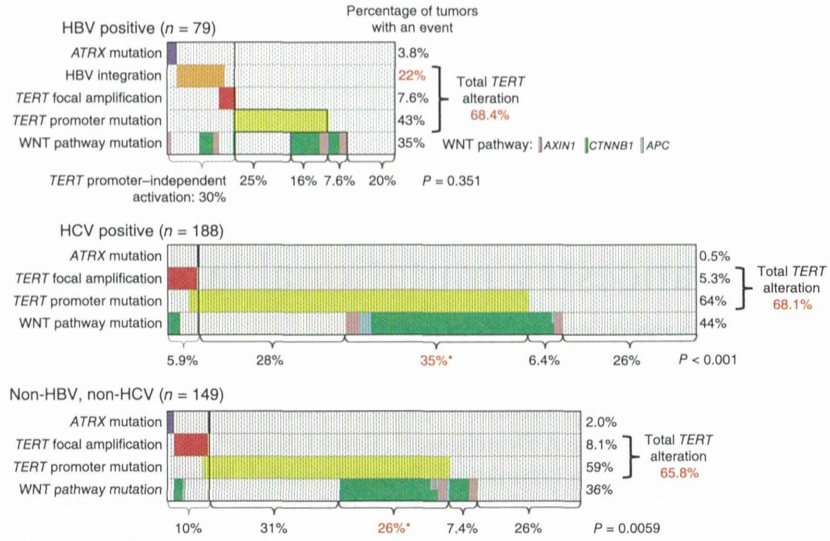

Figure 1 Multiple types of *TERT* alterations in HCC. Mutual exclusivity of HBV genome integration at the *TERT* locus, *TERT* focal amplification and *TERT* promoter mutation in HBV-positive (top), HCV-positive (middle) and non-HBV, non-HCV (bottom) cases. *AXIN1*, *CTNNB1* and *APC* mutations were included as WNT pathway mutations. *TERT* promoter mutation significantly co-occurred with WNT pathway mutation in HBV-negative cases (*$P < 0.001$, χ^2 test). HBV-positive cases without virus capture analysis (41 samples) were excluded (**Supplementary Table 28**).

or blood. This cohort contained 212 HCV-positive, 117 HBV-positive and 150 non-virus cases. The US cohort contained European-ancestry (55%), Asian (defined as US-Asian hereafter; 16%) and African-American (12%) cases. The clinical backgrounds for this cohort are shown in **Supplementary Table 1**.

The exons and surrounding noncoding genomic regions of protein-coding genes were captured in 452 pairs of tumor and non-cancerous liver tissues. Oncoviral genomes, including for HBV, human papillomavirus (HPV-16 and HPV-18) and human T-lymphotrophic virus 1 (HTLV1) (91 kb in total; **Supplementary Table 2**), were also captured in 198 cases. Whole-genome sequencing was conducted in 22 HCC pairs, including 9 exome-sequenced cases, and targeted resequencing of liver cancer genes was carried out for 38 cases. To minimize multicenter study bias due to differences in exome sequencing platform or data analysis pipeline, we optimized the somatic mutation detection algorithms and filtering conditions for three centers using Japanese cohort samples. High concordance (>87%) with a validation rate of >97% in somatic mutation detection was achieved, and substitution patterns among the three centers were consistent (**Supplementary Figs. 1** and **2**). We also confirmed that similar mutation spectra were observed in the same cases in whole-genome sequence and whole-exome sequence (**Supplementary Fig. 3**).

The average mutation rate was 2.8 mutations per megabase, and T>C and C>T substitutions were dominant in this cohort (**Supplementary Fig. 4**). Eight (1.7%) outlier tumors harboring more than 4.3 mutations per megabase showed substitution patterns distinctive from those of other cases and had somatic nonsense or missense mutations in mismatch repair (*MSH3*, *MSH4*, *MSH5* and *MSH6*), DNA polymerase (*POLA1*, *POLK*, *POLE* and *POLL*) or nucleotide excision repair (*ERCC1* and *ERCC2*) genes (**Supplementary Fig. 5**).

Panoramic view of ploidy, copy number and virus integration

We evaluated copy number alteration (CNA) by comparing the sequence depth for paired samples and allelic imbalance in the captured area (**Supplementary Fig. 6**). This digital assessment of CNA and allelic imbalance was consistent with SNP array data in cases analyzed by both methods (**Supplementary Fig. 7**). We also imputed deviation in the allele frequency of heterozygous single-nucleotide variation to predict the tumor purity and ploidy for each sample (H.U., S.Y., K.T. and H.A., unpublished data). A large fraction of cases (28.9%) represented whole-genome duplication with gross chromosomal loss (average ploidy was 3.87, and the average number of CNAs was 11.58) (**Supplementary Fig. 8**), whereas the remainder showed more stable copy number status (average ploidy was 2.08, and the average number of CNAs was 7.56). Tetraploidy was more frequently observed in higher-grade tumors ($P = 0.039$, Fisher's exact test; **Supplementary Fig. 9**).

We observed recurrent arm-level gains (1q, 5p, 6p and 8q) and losses (1p, 4q, 6q, 8p and 17p), as previously described for HCC[9] (**Supplementary Fig. 10**). Recurrent focal amplifications were detected in 25% of cases, including for *TERT* and *CCND1-FGF19*. Homozygous deletions were less frequent events (detected in 17.4% of cases). Recurrent homozygous deletion was observed for 28 genes, including *CDKN2A-CDKN2B*, *MAP2K3* and *PTEN* (**Supplementary Figs. 11** and **12**).

Using paired-end reads mapped to the HBV viral and human genomes, respectively, we detected 628 HBV virus integrations in 68 HBV-positive cases from which viral genomes were captured (9.2 integrations per case) (**Supplementary Table 3**), reflecting a detection rate that was 2–4 times more sensitive than in previous whole-genome sequencing studies[10,11]. Genes close to (less than 10 kb away from) the recurrent HBV integrations included *TERT* ($n = 17$ cases), *KMT2B* (*MLL4*; $n = 6$ cases), and *ALOX5*, *ZFPM2*, *SENP5*, *MYO19* and *RGS22* ($n = 2$ cases each). Recurrent non-genic HBV integrations were observed near the centromere, especially on chromosomes 1p, 8p and 10q. A significant fraction of HBV integrations were colocalized with (less than 500 kb away from) DNA copy number breakpoints (10.7%; $P < 1 \times 10^{-5}$, randomization test) (**Supplementary Figs. 13** and **14**). Despite intimate association between HBV genome integration and CNA breakpoints, the frequency of CNA was not different among the viral subtypes ($P = 0.29$, ANOVA test; **Supplementary Fig. 15** and **Supplementary Table 4**).

Multiple types of *TERT* genetic alteration in HCC

Somatic mutations in the transcriptional regulatory region of the *TERT* gene have been reported in a range of cancers, including HCC[12,13]. By combining captured noncoding sequence data with capillary sequencing validation, we detected *TERT* promoter mutations in 254 cases of the 469 cases analyzed (54% in total). The frequency of these mutations was highest in HCV-positive cases (121/188; 64%), with lower frequencies in non-viral cases (88/149; 59%) and HBV-positive cases (44/120; 37%) (**Supplementary Table 5**). As reported[13], the mutation located 124 bp upstream of the ATG start site (c.–124C>T, on the opposite strand; 93%) was more frequent than the c.–146C>T (4.3%) and c.–57A>C (1.6%) mutations (**Supplementary Table 6**).

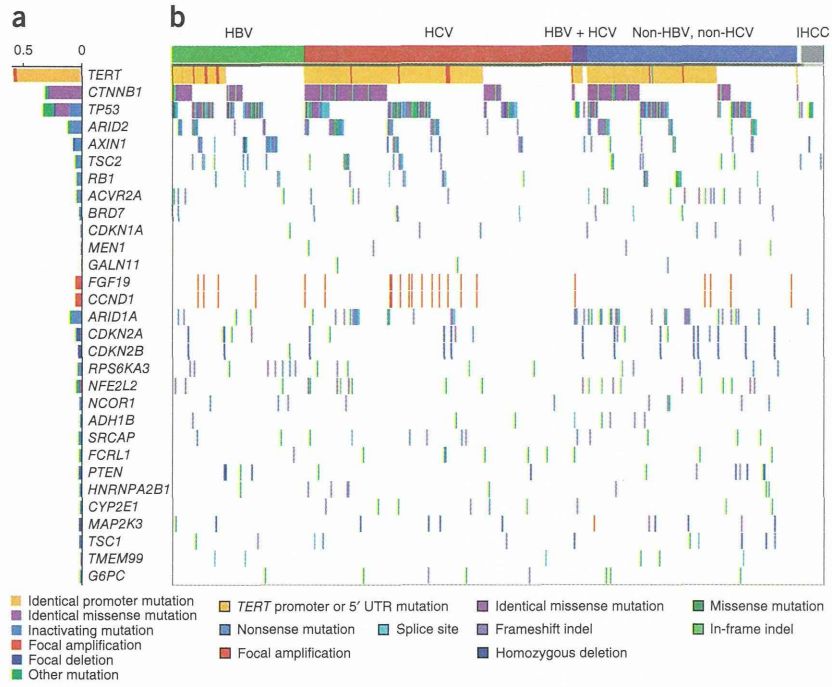

Figure 2 Significant cancer driver genes in HCC. An overview of significant driver genes in HCC. Shown are genes with statistically significant mutations or focal CNAs (**a**) and their alterations in each sample classified by the status of hepatitis virus infection (**b**). Genes were sorted by significant *q* value (**Supplementary Note**).

Additionally, *TERT* focal amplification was detected in 6.7% of the cases in total, and integration of the HBV genome in the *TERT* locus was observed in 22% of HBV-positive samples for which integration was analyzed. *TERT* promoter mutations were mutually exclusive with HBV genome integration in the *TERT* locus in integration-analyzed HBV-positive samples and were almost mutually exclusive with *TERT* focal amplifications, both of which were considered to cause higher *TERT* expression[14] (**Fig. 1**). Alterations of *ATRX* have also been reported to induce telomerase-independent telomere maintenance[15], Altogether, more than 68% of the HCC cases had alterations in either *TERT* or *ATRX*, representing the most frequent molecular event reported (**Supplementary Table 5**). In contrast, no *TERT* promoter mutations were detected in 13 IHCC cases (**Fig. 2**). *TERT* promoter mutations significantly co-occurred with WNT pathway gene alterations, such as *CTNNB1*, *AXIN1* or *APC*, in HCV-positive and non-virus cases, suggesting a cooperative oncogenic activity between *TERT* promoter mutation and the WNT pathway[16] in these subgroups (**Fig. 1**).

Significantly altered genes in HCC

To identify significantly altered genes in HCC, we used a combination of MutSigCV[17], an aggregated somatic alteration method that aggregates somatic substitutions, short indels, homozygous deletions and focal amplifications, and an inactivation bias method that calculates inactivating mutation bias (**Supplementary Fig. 16**, **Supplementary Tables 7–10** and **Supplementary Note**). Furthermore, we eliminated mutated genes that exhibited sequencing center bias and subclone bias as sources of possible false discovery (**Supplementary Tables 11** and **12**). These steps led to a final list of 30 candidate driver genes (**Fig. 2**, **Supplementary Fig. 17** and **Supplementary Tables 13–15**), including 13 that were not recurrently mutated in previous cohorts[18–20] (**Supplementary Table 16**). These 13 genes included *BRD7*, a component of the SWI/SNF nucleosome-remodeling machinery, and *MEN1*, a putative tumor suppressor somatically mutated in neuroendocrine tumors—neither of which has been reported in HCC. Mutations in *TSC2*, *SRCAP* and *NCOR1* have been reported as singletons in other

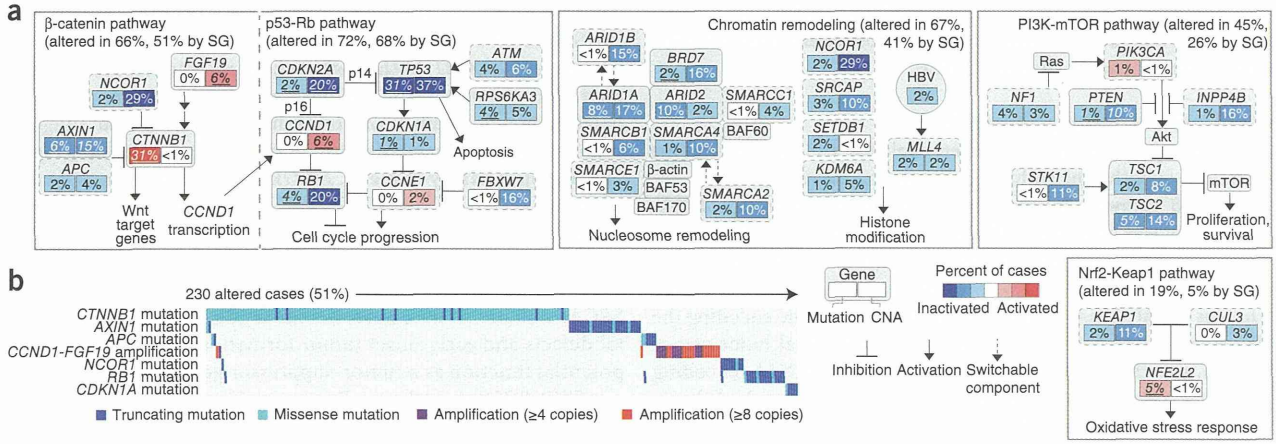

Figure 3 Oncogenic network in HCC. (**a**) Major signaling pathways involving genetic alterations in HCC. Key genes in each pathway are indicated by rectangles, with the percentages of somatic mutations and CNAs shown in the left and right portions of each rectangle, respectively. Significantly altered genes (SG; MutSigCV, $P < 0.05$ or GISTIC, q value < 0.1; percentages are underlined for alterations meeting either criterion) are bounded by solid lines, whereas other key genes in each pathway are bounded by dashed lines. (**b**) Mutual exclusivity plot of genes relevant to the WNT signaling pathway. The plot indicates that somatic mutations in WNT-related genes might contribute to the activation of WNT signaling in over half of all HCCs.

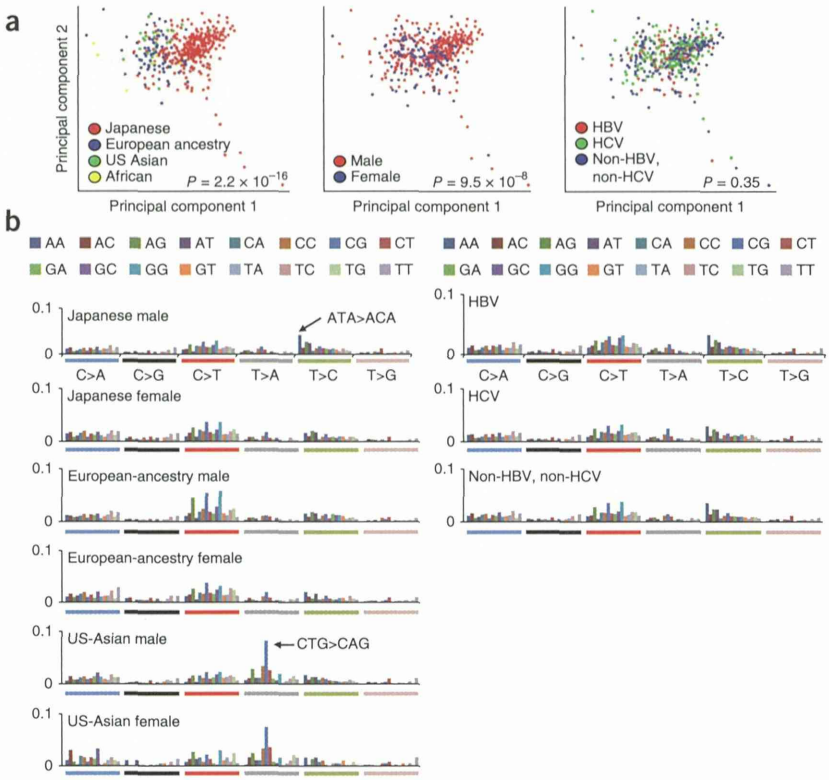

Figure 4 Somatic substitution patterns were associated with ancestry. (**a**) Principal-component analysis of the 96 substitution patterns in the HCC genome by ancestry group (left), sex (middle) and hepatitis virus group (right). (**b**) Average frequency of the 96 substitution patterns in each sample group (ancestry group, sex and virus group). The top legend shows the bases immediately 5′ and 3′ to each substitution. The *y* axis indicates the frequency of the 96 substitution patterns.

studies, but these genes were shown here to be significantly mutated. Some of the difference in results might be attributed to the greatly increased statistical power with our 503-case population, but some of the difference might also reflect contribution from the ancestry composition of the cohorts in this study. Several genes demonstrated differences in mutational frequency among virus subtypes (**Fig. 2b** and **Supplementary Table 17**). *AXIN1* was more frequently mutated in HBV-positive cases in comparison with HCV-positive and non-virus HCC (P = 0.0055, Fisher's exact test), indicating that different viral etiologies might activate WNT signaling in distinct ways. *ARID1A* was more frequently altered in non-virus cases (P = 0.009).

Alterations of drug target kinases were rarely found in HCC; low-level recurrent mutations of *FGFR2* (mutated in 1.8% of cases), *KIT* (1.3%), *FGFR3* (0.9%), *FGFR1* (0.9%), *JAK1* (0.9%) and *EGFR* (0.4%) and focal amplification of *MET* (0.5%) were detected. The specific mutations in these receptor tyrosine kinases were not generally observed in other cancers, with the exception of two *JAK1* mutations (encoding p.Ser703Ile and p.Leu910Pro substitutions), which were previously observed in a liver cancer sequencing study[20]. The liver has a central role in many metabolic processes. Our study identified recurrent mutations of metabolic enzyme genes in HCC (**Fig. 2b** and **Supplementary Tables 7** and **13**). These included *CYP2E1* (2.0%); *ADH1B* (1.8%), encoding alcohol dehydrogenase 1B; and *G6PC* (1.8%), encoding a glucose-6-phophatase catalytic subunit, whose aberrations could be linked to metabolomic changes in HCC.

Significant oncogenic pathways in HCC

Oncogenic pathways were further explored by aggregating the alterations of each gene within a particular pathway (**Fig. 3a**).

TP53-RB pathway. Inactivation of the tumor-suppressor TP53-RB pathway was a consistent theme in HCC. *TP53* mutations were observed in 31% of tumors, and two genes encoding p53-activating kinases, *ATM* and *RPS6KA3*, were also recurrently mutated. The *RB1* gene was mutated in 4.4% of cases. The *CDKN2A* gene encoding the RB regulator $p16^{INK4A}$ was subject to frequent focal homozygous deletion, and the p53 target and RB regulator *CDKN1A* (encoding $p21^{CIP1}$) was significantly mutated. Overall, 72% of cases had alterations in component genes of one or both of these pathways.

WNT pathway. In addition to activating *CTNNB1* mutations, inactivating mutations were frequently observed in WNT regulators, including *AXIN1* and *APC*. *CCND1* is a key downstream target of WNT signaling[21], and *FGF19* has been shown to activate *CTNNB1* transcriptional functions[22]. Mutual exclusivity of *CTNNB1*, *AXIN1* and *APC* mutations and *CCND1-FGF19* amplification supports the functional role of these genes in altering WNT signaling (**Fig. 3b**). Overall, 66% of HCCs showed WNT pathway–related alterations.

Chromatin and transcription modulators. A large proportion of the genes on the list of significantly mutated genes encoded chromatin modulators or transcriptional regulators. Frequent alterations in *NFE2L2*, encoding a transcriptional regulator that activates antioxidant and cytoprotective target genes[23], and its negative regulators *KEAP1* and *CUL3* (ref. 24) were noted. Also mutated were the nucleosome remodelers *ARID1A*, *ARID2* and *BRD7*, with CNAs and mutations in six additional members of the SWI/SNF complex (**Fig. 3a**), *SRCAP* and the transcriptional corepressor *NCOR1*, both of which have roles in steroid receptor–mediated transcription. These genes displayed primarily inactivating frameshift and nonsense mutations that suggest a tumor-suppressor gene function in HCC (**Supplementary Fig. 18** and **Supplementary Table 9**). *NCOR1* has been shown to directly suppress *CTNNB1* function[25] and exhibits mutual exclusivity for mutations with other WNT pathway genes (**Fig. 3b**). *SRCAP* encodes an Snf2-related CREBBP activator in several pathways, including NOTCH[26] and steroid receptors[27]. Truncating *SRCAP* mutations cause a rare hereditary disease with developmental defects and early-onset tumor formation[28,29], highlighting its potential function as a tumor-suppressor gene.

mTOR-PIK3CA pathway. Recurrent inactivating mutations in *TSC1-TSC2* and activating mutations and copy gain in *PIK3CA* were observed (**Fig. 3a**). Other modulators involved with this pathway, such as *NF1*, *PTEN*, *INPP4B* and *STK11*, were also affected, and, in total, 45% of cases had alterations in the mTOR-PIK3CA pathway. Somatic *TSC1* mutation was reported as a potential predictive biomarker of an mTOR inhibitor[30], and *TSC1*-mutated HCC cell lines showed

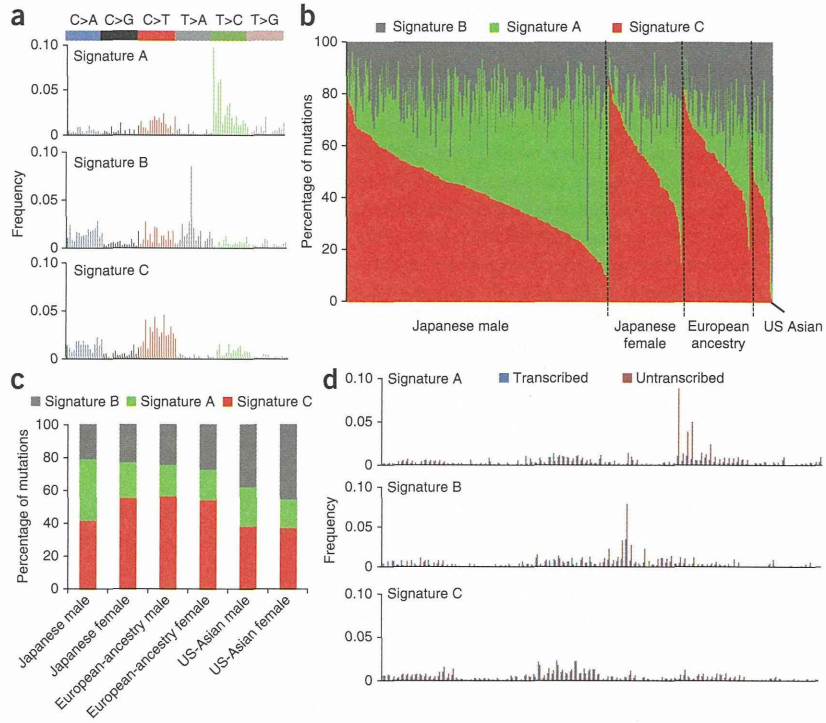

Figure 5 Ancestry-specific mutational signatures with transcriptional strand bias in the HCC genome. (**a**) The 3 mutational signatures in the HCC genome are shown according to the frequencies of 96 substitution types. The *y* axis indicates the frequency of each of the 96 substitution patterns. (**b**) Contribution of the three mutational signatures to each tumor. The *y* axis indicates the percentage of mutations comprised in each signature. The *x* axis indicates tumors classified in each ancestry group and by sex. (**c**) Contribution of the three mutational signatures to tumors from each ancestry group and sex. The *y* axis indicates the percentage of mutations comprised in each signature. (**d**) Transcriptional strand bias in mutational signatures. Each signature is displayed with 192 mutation patterns based on the 96 substitution types with transcriptional strand information. The mutation types are shown on the *x* axis, and the *y* axis indicates the frequency of each of the 192 mutation types contributing to each signature.

higher sensitivity to an mTOR kinase inhibitor (BEZ235) in comparison to cell lines with wild-type *TSC1* (**Supplementary Fig. 19**).

To identify networking among the oncogenic pathways in HCC, we developed a pathway compression algorithm and applied it to the significantly altered genes. We identified 11 core oncogenic network modules in HCC (**Supplementary Table 18**). To visualize these modules in the context of a biological network, we constructed a schematic view of the modules and the additional nodes that can connect them (**Supplementary Fig. 20**). The nodes were typically classified into two types; one type was closely connected to neighboring nodes (with higher value for centrality; **Supplementary Table 19**) and the other type had long-range edges that reached distant nodes, which can be used to measure the effect of each module alteration on the total network. Further comparison of the association between these module alterations and background clinical factors showed that the mTOR module was significantly different ($P < 0.05$, Cochran-Mantel-Haenszel test) in Asian and European-ancestry populations with respect to mutational frequencies (**Supplementary Fig. 21**).

Ancestry-dependent diversity in HCC mutation signatures

Somatic mutation patterns in human cancer are closely associated with epidemiological factors[31–34]; however, their association with ancestry remains unexplored. We integrated genomic data from an additional 105 HCC cases sequenced by TCGA along with the 503 cases sequenced by us (**Supplementary Table 1**) and compared somatic substitution patterns according to epidemiological data and ancestry group. Because mutation patterns in hypermutated cases and IHCC were distinctive (**Supplementary Figs. 4** and **22**), these two groups were excluded from further mutation pattern analysis.

Principal-component analysis of the 96 possible nucleotide triplets, dependent on the bases immediately 5′ and 3′ to each substitution, showed that the constitution of substitution patterns with these triplets was significantly different by ancestry group (Japanese, US Asian and European ancestry; $P = 2.2 \times 10^{-16}$, Wilks' test) and by sex ($P = 9.5 \times 10^{-8}$) (**Fig. 4a**). Notably, substitution patterns were not significantly associated with viral status (HBV, HCV and non-viral, $P = 0.35$; **Fig. 4a** and **Supplementary Fig. 23**). T>C substitutions, particularly in an ATA context, were specifically increased in Japanese male samples, and T>A substitutions (most frequently in a CTG context) were specifically increased in US-Asian male and female samples. The distributions of the frequencies for the 96 substitution types were similar among Japanese female samples and European-ancestry male and female samples (**Fig. 4b**).

We applied non-negative matrix factorization (NMF) analysis to the 96-substitution pattern[33] and identified 3 mutation signatures (HCC signatures A–C; **Fig. 5a** and **Supplementary Fig. 24**). Each signature was composed of context-specific substitutions: HCC signature A was characterized by dominant T>C mutations, especially in an AT(A/G/T) context, whereas HCC signature B contained dominant T>A mutations, with a sharp increase in frequency for a CTG context. HCC signature C contained dominant C>T mutations, especially in an (A/C/G)CG context. The distribution of these signatures was associated with ancestry and sex but not with the virus status (**Supplementary Table 20**). Among the different ancestry groups, HCC signatures A and B more frequently contributed to Japanese male (odds ratio (OR) = 2.2; $P = 0.0025$, Fisher's exact test) and US-Asian (OR = 2.5; $P = 0.00036$) cases, respectively, whereas HCC signature C was common across all ancestry groups and in both sexes (**Fig. 5b,c** and **Supplementary Fig. 25**). Remarkable differences in mutation prevalence between the transcribed and untranscribed strands were observed for T>C substitutions, especially in an AT(A/G/T) context ($P = 7.4 \times 10^{-152}$, χ^2 test), in HCC signature A and for T>A substitutions, especially in a CTG context ($P = 3.3 \times 10^{-8}$), in HCC signature B (**Fig. 5d**). These significant strand biases imply the involvement of transcription-coupled repair, which is tightly associated with known carcinogens in other tumor types[31–34]. There was no significant association between the signature distribution and the *ALDH2* SNP rs671, which is associated with alcohol metabolism and is a more frequent genotype in the Asian population[35] (**Supplementary Table 21**).