

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

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 *IDH*s 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.

<sup>1</sup>Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo 108-8639, Japan. <sup>2</sup>Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. <sup>3</sup>Laboratory of DNA Information Analysis, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>4</sup>Department of Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka 537-8511, Japan. <sup>5</sup>Department of Medicine & Molecular Science, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. <sup>6</sup>Department of Anatomical Pathology, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. <sup>7</sup>Department of Gastroenterological Surgery II, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan. <sup>8</sup>Second Department of Surgery, Wakayama Medical University, Wakayama 641-8510, Japan. <sup>9</sup>Department of Gastroenterological Surgery, Tokyo Women's Medical University, Tokyo 162-8666, Japan. <sup>10</sup>Genome Analysis Laboratory, Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi Vietnam. <sup>11</sup>Laboratory of Sequence Analysis, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. <sup>12</sup>Division of Cancer Genomics, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. <sup>13</sup>Division of Molecular Pathology, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. <sup>14</sup>Hepatobiliary and Pancreatic Surgery Division, National Cancer Center, Chuo-ku, Tokyo 104-0045, Japan. <sup>15</sup>Department of Gastroenterological Surgery, Hiroshima University School of Medicine, Hiroshima 734-8551, Japan. <sup>16</sup>Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. <sup>17</sup>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).

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<sup>1–3</sup>. 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<sup>1</sup>.

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<sup>2,4</sup>. 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 aetiological factor of ICC, as well as HCC, in Asia<sup>2</sup>, and indicated that hepatitis-associated ICC and HCC share a common disease process for carcinogenesis<sup>5</sup>. HCC and ICC have been reported to develop simultaneously in both human and mouse models<sup>5–7</sup> and a combined or mixed phenotype (cHCC/CC) displays intimately mixed components of both hepatocellular and biliary epithelial differentiation<sup>8</sup>, 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<sup>6,9</sup>. 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<sup>10–20</sup>, 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<sup>21</sup>.

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

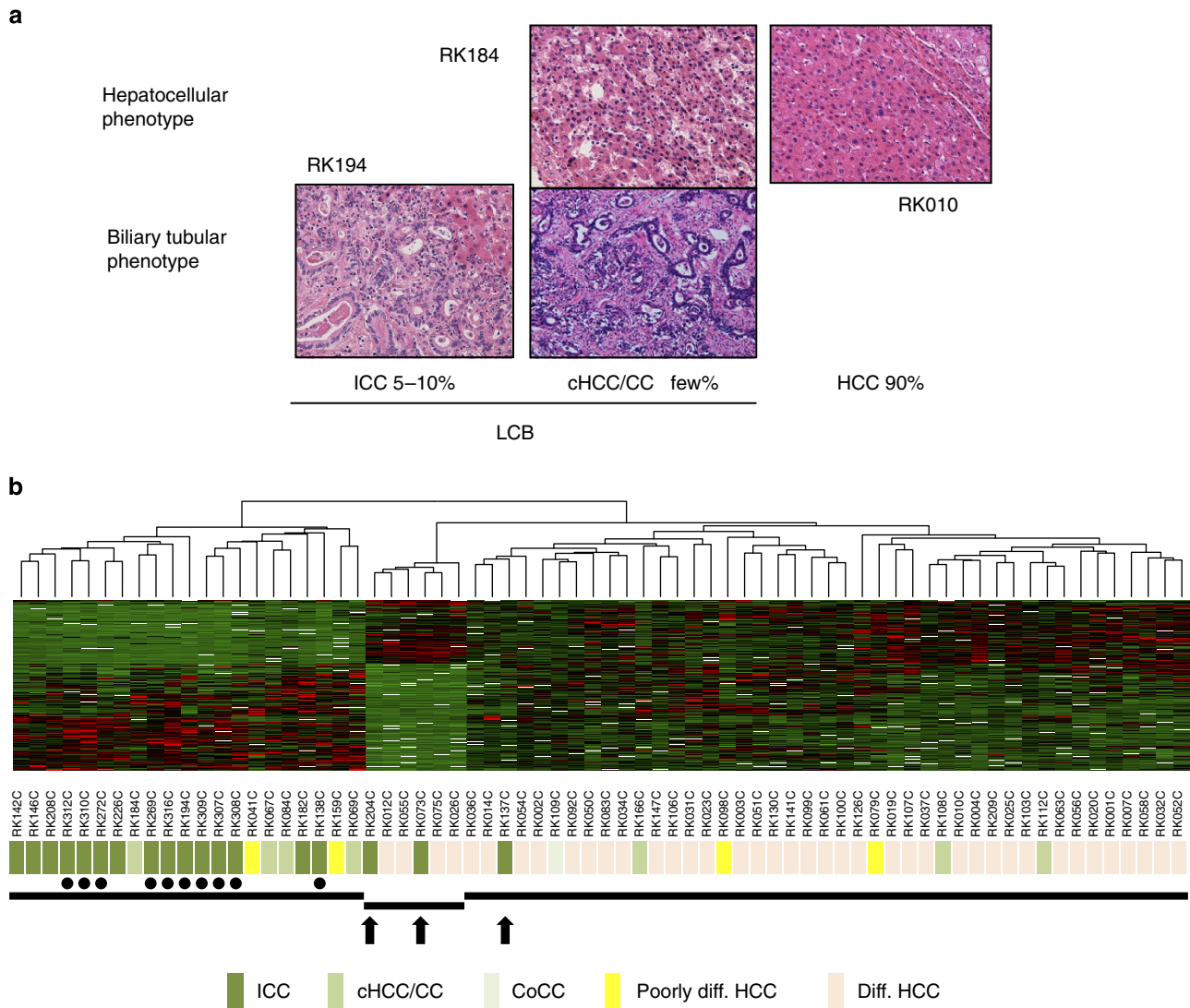
**Samples and WGS.** In this study, frozen tumour and matched normal tissues were collected from 30 patients with LCB (Table 1) and 60 patients with HCC (Supplementary Table 1). All samples were used for WGS. RNA from 25 LCB and 44 HCC samples, from which high-quality RNA was obtained, was also sequenced. 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<sup>22</sup>. 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<sup>23</sup> (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<sup>12,14,15</sup> (Supplementary Table 4 and Supplementary Methods), indicating that HBV-infection and its genomic integration can be involved with the carcinogenesis of LCBs<sup>24,25</sup>.

**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<sup>7</sup>. 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<sup>26</sup>.

**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 ( $\chi^2$ -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



**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<sup>21</sup>. 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<sup>18</sup>.

To identify the mutational signatures of the hepatitis-positive or -negative LCBs, we used EMu software<sup>27</sup> 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

**Table 1 | Clinical and pathological features of 30 LCBs analyzed by whole-genome sequencing.**

ID	Age	Gender	Viral infection	Histology*	TNM†	Tumour size (mm)	vp‡	vv§	b	Liver fibrosis†	Note
RK067	89	M	HCV	cHCC/CC	T2N0M0	25	–	–	–	3	
RK069	85	F	(–)	cHCC/CC	T3N0M0	80	+	+	–	3	
RK073	62	M	HBV	ICC	T1N0M0	10	–	–	–	4	HCC MCT + #
RK084	67	F	HCV	cHCC/CC	T3N0M0	30	+	–	–	4	
RK108	74	M	HCV	cHCC/CC	T1N0M0	12	–	–	–	2	
RK109	84	M	HCV	CoCC	T2N0M0	16	–	+	–	1	
RK112	83	M	HBV	cHCC/CC	T3N0M0	25	–	–	–	1	
RK137	74	F	HCV	ICC	T3N0M0	24	–	–	+	3	HCC MCT +
RK138	75	F	(–)	ICC	T3N0M0	120	+	+	+	0	Hepatitis-negative**
RK142	57	M	(–)	ICC	T1N0M0	15	–	–	+	3	
RK146	57	F	HBV	ICC	T2N0M0	23	+	–	–	3	
RK166	67	M	HBV	cHCC/CC	T3N0M0	25	–	–	+	4	
RK182	65	M	HCV	ICC	T3N0M0	28	+	–	–	3	
RK184	64	M	(–)	cHCC/CC	T3N0M0	110	+	–	–	1	
RK194	67	M	(–)	ICC	T3N0M0	35	+	–	+	0	Hepatitis-negative
RK204	83	M	HCV	ICC	T1N0M0	12	–	–	–	4	HCC MCT + (RK209)
RK208	60	M	(–)	ICC	T3N0M0	60	–	–	+	2	
RK226	59	M	HBV	ICC	T3N0M0	45	+	+	–	2	
RK269	74	M	(–)	ICC	T1N0M0	12	–	–	–	0	Hepatitis-negative
RK272	78	F	(–)	ICC	T3N0M0	45	+	+	–	0	Hepatitis-negative
RK279	69	M	HCV	ICC	T3N0M0	35	+	+	+	3	
RK298	68	M	HCV	ICC	T3N0M0	40	+	–	–	1	
RK303	76	M	(–)	CoCC	T2N0M0	20	+	–	–	2	
RK307	61	F	(–)	ICC	T3N1M0	75	+	–	+	0	Hepatitis-negative
RK308	70	F	(–)	ICC	T1N0M0	14	+	+	+	0	Hepatitis-negative
RK309	56	M	(–)	ICC	T2N1M0	36	–	–	+	0	Hepatitis-negative
RK310	62	F	(–)	ICC	T3N0M0	90	+	+	+	0	Hepatitis-negative
RK312	66	M	HBV	ICC	T3N1M0	48	+	+	–	0	Hepatitis-negative
RK316	54	F	(–)	ICC	T3N3M0	54	+	+	–	0	Hepatitis-negative
RK317	73	M	HBV	ICC	T3N0M0	45	+	+	–	2	

\*ICC (intrahepatic cholangiocellular carcinoma), CoCC (cholangiocellular carcinoma), cHCC/CC (combined hepatocellular cholangiocellular carcinoma).

†TNM staging in UICC.

‡vp; portal vein invasion.

§vv; hepatic vein invasion.

||b; bile duct invasion.

\*Fibrosis in non-cancerous liver tissue is determined according to the New Inuyama Classification.

#MCT; multicentric tumour.

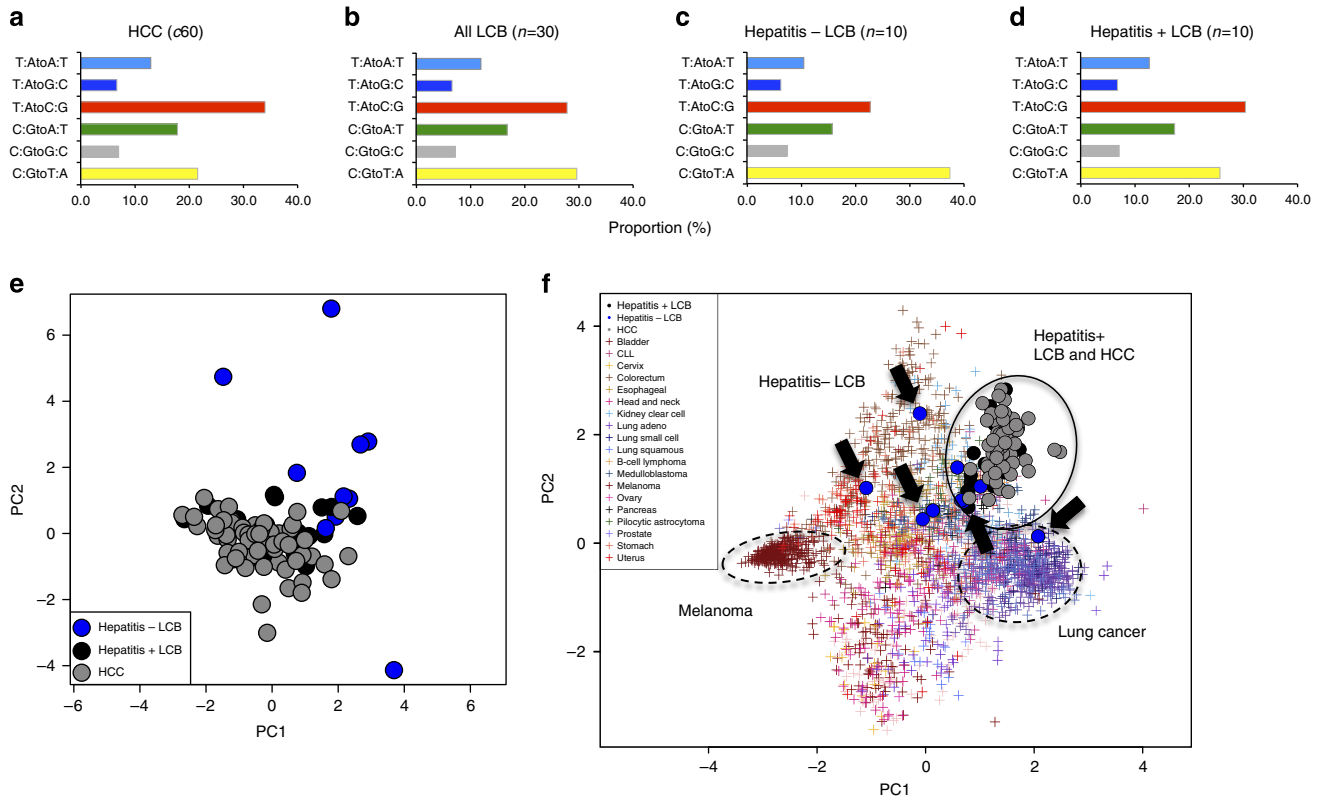
\*\*Liver fibrosis 0 indicates hepatitis-negative LCB.

morphological classification, whereas the somatic substitution pattern in the LCBs is strongly determined by the aetiological background of chronic hepatitis.

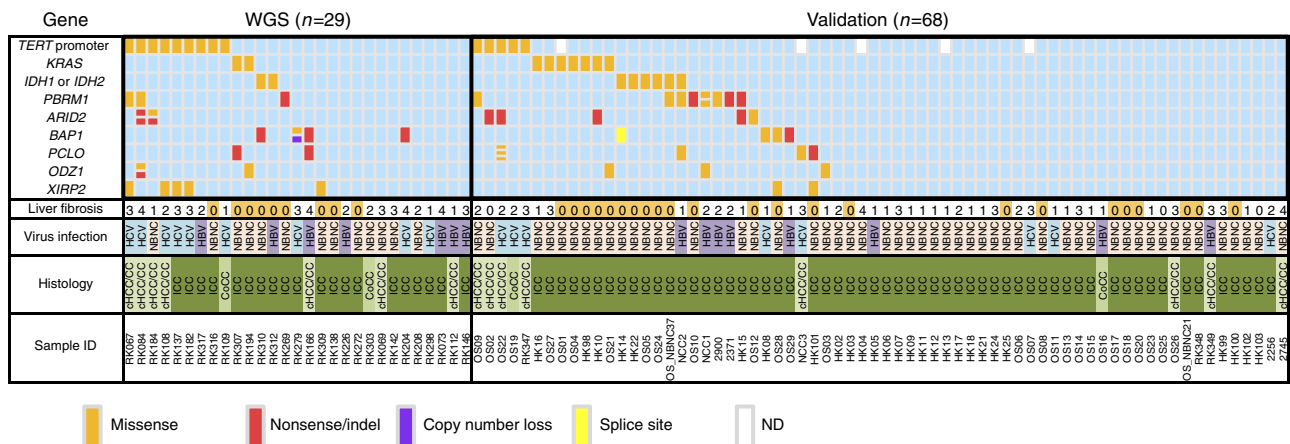
**Recurrently mutated genes and pathway analysis.** We examined recurrently mutated genes in our LCB samples. RK308 had an exceptionally large number of mutations and was excluded from the subsequent analyses. Across the 29 LCB genomes, we detected 892 protein-altering mutations, including 760 nonsynonymous, 108 short coding indels and 24 splice-site mutations (Supplementary Table 3). Thirty-two genes were recurrently mutated (Fig. 3 and Supplementary Table 5): cytoskeleton genes (*XIRP2*, *KIF2B* and *MYO10*), a cell adhesion molecule (*CDH2*), known tumour suppressors (*TP53*, *PTEN* and *BAP1*), known oncogenes (*KRAS* and *PIK3CA*), chromatin regulators (*PBRM1*, *ARID1A* and *ARID2*), which are highly mutated in HCC and other cancers<sup>10,12</sup>, neuron growth genes (*ODZ1*, *EPHA2* and *PLCO*) and tyrosine kinase receptors (*ERBB4* and *EPHA2*). To validate the frequency of the mutations, the protein-coding exonic regions of recurrently mutated genes (*BAP1*, *CDH2*, *EPHA2*, *KIF2B*, *MGAT4C*, *ODZ1*, *PBRM1*, *PCLO*, *SYT1*, *ARID2* and *XIRP2*) were amplified in an additional 68 LCB samples (Supplementary Table 6) and sequenced by the Illumina HiSeq2000 sequencer. In addition, as *KRAS*, *IDH1*, *IDH2* and *TERT* promoter mutations were frequently observed in ICCs<sup>28</sup>, fluke-related ICCs<sup>18</sup> and HCCs<sup>29</sup>, we sequenced exons 2 and 3 of

*KRAS*, exon 4 of *IDH1* (codon 132), exon 4 of *IDH2* (codon 172) and *TERT* promoter hotspots (chr5:1,295,228 and 1,295,250) in the additional 68 samples by Sanger sequencing. *TERT* promoter hotspots were also examined in the 30 WGS samples by Sanger sequencing owing to a low depth of coverage in the WGS. This validation experiment revealed that one *TERT* promoter hotspot (chr5:1,295,228) was mutated in 14 samples (15.2%), *KRAS* and *PBRM1* in 7 samples (10.3%), *ARID2* in 5 (7.4%), *BAP1*, *PCLO* and *IDH1* in 4 (5.9%), *ODZ1* in three (4.4%) and *EPHA2*, *SYT2*, *CDH2*, *XIRP2* and *IDH2* in two samples (2.9%) (Fig. 3 and Supplementary Table 7). In the *ARID2*, *PBRM1* and *BAP1* genes, which encode chromatin regulators, an accumulation of loss-of-function mutations was observed, suggesting that they are likely to function as tumour suppressors in LCBs as well as HCCs<sup>10,12</sup>. As observed in previous HCC studies<sup>10,12,13</sup>, more than half of LCBs had somatic mutations and rearrangements accumulated in chromatin regulators (Supplementary Fig. 8).

We then examined the frequency of gene mutation and its association with clinical information in the WGS and validation samples (Table 2 and Supplementary Table 8). The frequency of the *TERT* promoter hotspot mutations was significantly lower in LCBs than in HCCs (Fisher's exact test  $P$ -value =  $1.2 \times 10^{-9}$ ) (Table 2). Furthermore, the frequency was significantly higher in cHCC/CCs and HCCs than in the ICCs (Fisher's exact test ICCs versus cHCC/CC;  $P$ -value =  $6.5 \times 10^{-5}$  and ICCs versus HCC;  $P$ -value =  $2.1 \times 10^{-11}$ ), but no significant difference was observed between cHCC/CCs and HCCs (Table 2). The frequency



**Figure 2 | Genome-wide substitution pattern on the 30 LCBs and 60 HCCs.** Average proportion of somatic nucleotide substitutions for (a) the 60 HCCs, (b) the 30 LCBs, (c) the 10 hepatitis-negative LCBs and (d) the 20 hepatitis-positive LCBs. (e) Principal component analysis (PCA) of the whole-genome substitution patterns of the 30 LCBs and 60 HCCs. Hepatitis-positive LCBs (black dots) overlap the HCC cluster (gray). LCBs developed in livers without hepatitis (blue dots) diverged from others. (f) PCA of the whole-genome substitution patterns of the 30 LCBs, the 60 HCCs and other types of cancers<sup>21</sup>. Hepatitis-positive liver cancers (HCC and LCBs) are tightly clustered, as are melanomas, indicating that chronic hepatitis or inflammation can strongly impact the somatic mutation signature.



**Figure 3 | Analysis of recurrently mutated genes in LCBs.** Recurrently mutated genes in the 30 LCB WGS set and the 68 validation LCB set. Histological subtype of LCB: ICC; intrahepatic cholangiocarcinoma, cHCC/CC; combined hepatocellular-cholangiocarcinoma and CoCC; cholangiolocellular carcinoma. Liver fibrosis was determined according to the New Inuyama Classification (0~4). Liver fibrosis 0 indicates hepatitis-negative LCB. Virus infection status is classified as HBV, HCV or negative, which was determined by serological study.

of mutations in the *KRAS* and *IDH* gene hotspots and *PBRM1* was higher in the LCBs than in the HCCs (not significant after the Bonferroni correction) and no mutations were observed in the 60 HCCs.

When taking into account the presence of hepatitis, we found significant difference in the frequency of the *TERT* promoter hotspot mutations among the hepatitis-positive LCBs, the hepatitis-negative LCBs and the HCCs. The HCCs and

hepatitis-positive LCBs had a higher frequency of the *TERT* promoter mutations than the hepatitis-negative LCBs (Table 2). In contrast, HCCs had significantly lower frequency of *KRAS* and *IDH* gene hotspot mutations than the hepatitis-negative LCBs, all of which are ICCs (Fisher's exact test *KRAS*; *P*-value = 0.0006, *IDH* genes; *P*-value = 0.0006). Hepatitis-positive LCBs shared more mutated genes with HCCs, which is consistent with the substitution pattern shown in Fig. 2f. However, the frequency of

**Table 2 | Summary of mutations in WGS and validation set of LCBs.**

	LCBs		HCCs	Hepatitis		Histology		Unadjusted <i>P</i> -value for comparison						
	WGS ( <i>n</i> = 29)	Validation ( <i>n</i> = 68)	WGS ( <i>n</i> = 60)	Hepatitis- positive LCBs ( <i>n</i> = 62)	Hepatitis- negative LCBs ( <i>n</i> = 35)	ICC ( <i>n</i> = 82)	cHCC/ CC ( <i>n</i> = 15)	LCB versus HCC	Hepatitis- positive LCB versus hepatitis- negative LCB	Hepatitis- negative LCB versus HCC	Hepatitis- positive LCB versus HCC	ICC versus cHCC/CC	ICC versus HCC	cHCC/ CC versus HCC
<i>TERT</i> Promoter	9 (30%)	5 (10%)	38 (63%)	12 (20%)	2 (6%)	4 (8%)	8 (53%)	$1.2 \times 10^{-9*}$	NS	$1.8 \times 10^{-8*}$	$2.8 \times 10^{-6*}$	$6.5 \times 10^{-5*}$	$2.1 \times 10^{-12*}$	NS
<i>KRAS</i>	2 (7%)	7 (10%)	0 (0%)	2 (3%)	7 (20%)	9 (11%)	0 (0%)	0.013	0.010	0.00061*	NS	NS	0.010	NS
<i>PBRM1</i>	3 (10%)	7 (10%)	0 (0%)	8 (13%)	2 (6%)	7 (9%)	3 (20%)	0.014	NS	NS	0.0062	NS	0.021	0.0067
<i>IDH1/IDH2</i>	2 (7%)	6 (9%)	0 (0%)	1 (2%)	7 (20%)	8 (10%)	0 (0%)	0.024	0.0030*	0.00061*	NS	NS	0.021	NS
<i>ARID2</i>	2 (7%)	5 (7%)	3 (5%)	4 (6%)	3 (9%)	3 (4%)	4 (27%)	NS	NS	NS	NS	NS	0.010	NS
<i>BAP1</i>	2 (7%)	4 (6%)	1 (2%)	3 (5%)	3 (9%)	6 (7%)	0 (0%)	NS	NS	NS	NS	NS	NS	NS
<i>PCLO</i>	2 (7%)	4 (6%)	4 (7%)	4 (6%)	2 (6%)	3 (4%)	3 (20%)	NS	NS	NS	NS	NS	NS	NS
<i>ODZ1</i>	2 (7%)	3 (4%)	0 (0%)	3 (5%)	2 (6%)	4 (5%)	1 (7%)	NS	NS	NS	NS	NS	NS	NS
<i>XIRP2</i>	5 (17%)	1 (1%)	5 (8%)	4 (6%)	2 (6%)	4 (5%)	2 (13%)	NS	NS	NS	NS	NS	NS	NS

\*Significant after the Bonferroni correction.

RK308 had an exceptionally large number of mutations and was excluded from this analysis. Number and frequency (%) of mutations are shown. For *TERT* promoter, *KRAS* and *IDH* genes, hotspot mutations were counted. *TERT* promoter hotspots (chr5:1,295,228 and 1,295,250) were examined by Sanger sequencing method. CoCCs (*n* = 4) were included in ICCs. *P*-values were obtained by the Fisher's exact test.

mutations in the *PBRM1* gene was different between hepatitis-positive LCBs and HCCs, which may be related to cell differentiation in the liver cancer development<sup>30</sup> (marginal significance after the Bonferroni correction).

Mutations in the *KRAS* gene were significantly enriched in patients with lymph node metastasis (Supplementary Table 8), and mutations in the *BAP1* gene were significantly enriched in patients with vascular or bile duct invasion (Supplementary Table 8). Mutations in *IDH* genes were associated with poor disease-free survival after adjustment for age, which is consistent with the previous study<sup>20</sup> (Supplementary Fig. 10).

To identify gene sets and pathways related to the LCB development, we carried out gene set enrichment analysis for all nonsynonymous mutations, short indels and rearrangements<sup>31</sup>. After adjustment for the multiple testing, 36 categories including 'synapse organization' and 'cytoskeleton' were significantly overrepresented (Supplementary Table 9). These results suggest that genes in these categories have an important role in the carcinogenesis or cancer development in the LCBs. As 'axon guidance' genes were significantly mutated in pancreatic cancer<sup>32</sup>, genes related to neuron growth may have an important role in the carcinogenesis and development of aggressive hepatobiliary-pancreatic cancers, including LCBs.

To determine any possible biological activity of these mutated genes in LCBs, we examined four genes (*PCLO*, *EPHA2*, *ODZ1* and *XIRP2*), which are involved in synapse organization and/or cytoskeleton structure. We knocked down the expression of each candidate gene by short interfering RNA in liver cancer cell lines, and examined their proliferation, migration and invasion abilities. These experiments confirmed that silencing of *PCLO* promoted cell invasion in liver cancer cell lines (Supplementary Figs 11 and 12).

**Genetic heterogeneity within liver tumour.** A tumour is a population of heterogeneous cancer cells, and the analysis of this heterogeneity should provide us with deeper insights into tumorigenesis<sup>12,33–36</sup>. To examine intratumour heterogeneity, the clonal proportion of the 1,085 nonsynonymous point mutations and short indels, detected in randomly selected 15 LCBs and 10 HCCs, were sequenced to an average depth of 56,462x by ultra-deep sequencing (Supplementary Methods). Copy number alternations were adjusted for mutant-allele frequencies using allelic imbalance ratio, and proportion of mutated allele (PMA) was obtained. The distribution of PMA in the ICCs and the cHCC/CCs significantly differed (Wilcoxon's test *P*-value = 0.0047), and ICC genomes had a larger number of mutations

with higher PMA (Supplementary Fig. 13). One possible explanation is that the pattern of genetic heterogeneity is different between cHCC/CC and ICC, and cHCC/CC had larger number of shared mutations in the tumour population than ICC, which is consistent with the histological diversity of cHCC/CC, showing mixed components of both hepatocellular and biliary epithelial differentiation. We then examined the clonal proportion for each gene (Supplementary Dataset 2). Genes in 15 categories, including 'replicative senescence' and 'negative regulation of DNA replication' had a higher PMAs after adjustment for multiple testing (Supplementary Table 10). All categories contained the *TP53* gene, indicating that the *TP53* gene conferred clonal advantage to cancer cells. This result is consistent with a breast cancer study<sup>33</sup>. Various genes showed high frequency of mutations in each tumour and would be candidates for tumour initiators (Supplementary Dataset 2).

## Discussion

In the present study, we comprehensively analyzed 30 LCBs by WGS and RNA-seq, and compared their genomic landscapes with those of 60 HCCs. To our knowledge, this is the first study that demonstrates the impact of chronic hepatitis and inflammation on the mutational landscape of the cancer genome and the first whole-genome comparison between LCBs and HCCs. In our analysis, gene expression patterns are consistent with the histological classifications; HCCs and LCBs were differentially clustered and hepatitis-positive and -negative LCBs were clustered together. In contrast, hepatitis-positive and -negative LCBs differentiated in their genome-wide somatic substitution pattern. The hepatitis-positive LCBs clustered more tightly to hepatitis-related HCCs, whereas hepatitis-negative LCBs were more spread out. These results suggest that gene expression depends on the histological phenotype, whereas the somatic substitution pattern is strongly influenced by aetiological background like the occurrence of chronic hepatitis. Previous studies suggested that the expression pattern is consistent with pathological phenotype, but does not reflect tumour origin<sup>37–39</sup>. A mouse study on pancreatic ductal adenocarcinoma suggested that inflammation can promote neoplasia by altering cell differentiation<sup>38</sup>, and a comparison between virus-associated ICCs and HCCs suggested that they can arise from the hepatic progenitor cells<sup>5</sup>. Considering these studies, the similarities between the hepatitis-positive LCBs and the HCCs in the somatic substitution pattern may indicate their same cellular origin, such as liver progenitor cell. In contrast, hepatitis-negative LCB may arise from different origins, such as cholangiocytes<sup>40</sup>.

The frequency of some driver mutations, such as hotspot mutations in *KRAS*, *IDH1/2* and *TERT* promoter, differed among cancer types, hepatitis-positive and -negative LCBs. Mutations of *KRAS* and *IDH* genes were more frequent in the hepatitis-negative LCBs, and the *TERT* promoter mutation was more frequent in the cHCC/CCs and HCCs. As almost all cHCC/CC and HCCs were hepatitis-positive, it is difficult to differentiate the impact of hepatitis from that of the histology. In general, HCC and cHCC/CC, which mainly developed under a hepatitis background, had a larger frequency of *TERT* promoter mutations and a lower frequency of *KRAS* and *IDH1/2* mutations.

In the current study, we found that the occurrence of chronic hepatitis impacted the mutational landscape, discovered new driver gene and examined intratumour heterogeneity in the LCBs. Our analysis indicates that the WGS can reveal the impact of aetiological background on the genome-wide substitution pattern, and suggest that the WGS can contribute to molecular classification based on their aetiology. However, we did not find mutations in the driver gene candidates in about a half of the samples, suggesting that LCB is a highly heterogeneous cancer. Analysis of larger number of samples would be necessary for deeper understanding of LCB.

## Methods

**Clinical samples.** The clinical and pathological features of 30 LCBs that were used in WGS analysis are in Table 1. Our pathologists evaluated hematoxylin and eosin-stained slides and diagnosed HCC, ICC and cHCC/CC according to the 2010 WHO Classification of Tumors of the Digestive System<sup>41</sup>. We defined ICC and cHCC/CC, both of which contain varying degrees of epithelial tubular-differentiated cells (Fig. 1a), as liver cancer displaying biliary phenotype (LCB), distinguishing them from the hepatocellular phenotype (HCC). Viral infection was defined by the presence of HB surface antigen in patient's serum, or by the presence of antibody to HCV in patient's serum. Hepatitis-negative LCB was defined as a tumour showing no sign of chronic inflammation and liver fibrosis, which was determined according to the New Inuyama Classification. All subjects had undergone partial hepatectomy, and pathologists estimated the ratio of viable tumour cells in each sample. High molecular weight genomic DNA was extracted from fresh-frozen tumour specimens and blood. Non-cancerous liver tissues were used as the normal tissue for RK182, RK307, RK308, RK309 and RK310. All subjects agreed with informed consent to participate in the study following ICGC guidelines<sup>42</sup>. IRBs at RIKEN and all groups participating in this study approved this work.

**Whole-genome sequencing.** DNA was extracted from tumours and non-cancer frozen tissues, and 500 bp insert libraries were prepared according to the protocol provided by Illumina. The libraries were sequenced on HiSeq2000 platforms with paired reads of 101 bp. The mutation data for the 60 HCCs have been generated in the same way by RIKEN and deposited to the ICGC dataset version 8 released at 2012 March (<http://icgc.org/>).

**Somatic mutation and short indel calling.** Point mutations and somatic indels were identified using our in-house methods<sup>12</sup>. In brief, read pairs were mapped by BWA<sup>43</sup>, and the result files were converted to pileup file by samtools<sup>44</sup>. After PCR duplications were removed and comparing between cancer genome sequences and non-cancer genome sequences, somatic point mutations and indels were identified by our in-house mutation caller<sup>12</sup>. False-negative and false-positive rates of our analysis pipeline were described previously<sup>12</sup>. Information for all point mutations and indels in the 30 LCBs and the 60 HCCs was deposited to the ICGC web site (<http://www.icgc.org/>).

**Identification of rearrangements.** Inconsistent read pairs which occurred within 500 bp of each other were considered to support the same rearrangement. We identified candidate rearrangements in both tumour (support read pairs  $\geq 4$ ) and normal tissue (support read pairs  $\geq 1$ ) samples, and tumour-specific rearrangement candidates were identified. To exclude mapping errors, we performed a blast search of read pairs that support rearrangements against the reference genome. If a read pair mapped with correct orientation and distance ( $\leq 500$  bp) with an  $E$ -value  $< 10^{-7}$ , we excluded that read pair. Reads mapped with more than two mismatches were also discarded. After filtering, candidates supported by  $\geq 4$  read pairs and at least one perfect match pair were considered as somatic rearrangements. The candidates that the same rearrangement was found in other normal samples were filtered out. False discovery rate of this method was estimated to be 2.3% (4/176).

**Statistical analysis.** The random distribution was calculated by multiplying (proportion of nucleotide in the reference genome sequence) and (total number of mutations) as done in the previous study<sup>12</sup>. Tests for significantly mutated genes and PCA of the substitution pattern were carried out as described previously<sup>12</sup>.

Survival analyses were done using the 'survival' package for the R programming environment (<http://www.r-project.org>). A Cox proportional hazards model was used to test association between disease-free survival and mutations in the genes (*TERT* promoter, *KRAS*, *XIRP2*, *ARID2*, *BAP1*, *PBRM1*, *PCLO*, *ODZ1* and *IDH* genes) and clinical factors (age, gender, virus type and liver fibrosis). Model selection was done by the stepAIC function, and the model with age and mutations in *IDH* genes was selected.

Estimation of PMA was described in the Supplementary Methods.

To test the difference of the clonal proportion of mutations among ICCs, cHCC/CCs and HCCs, we calculated PMA for each mutation, which was standardized by the maximum PMA in each sample. Then we compared the median of the distribution of PMA between ICCs, cHCC/CCs and HCCs by Wilcoxon's test.

To identify gene sets with high clonal proportion, we used 'biological process' terms with depth = 5 in the Gene Ontology (GO) database (<http://www.geneontology.org>). The clonal proportions of the genes within and outside the gene category were compared by Wilcoxon's test as a previous study<sup>33</sup>. Note that we used unadjusted clonal proportions (not PMAs) for this analysis to consider the influence of copy number changes.

**Sanger sequencing and ultra-deep amplicon sequencing.** Sanger sequencing of PCR products was performed on ABI 3770x. For ultra-deep sequencing of mutations, each of the 100 bp target regions was amplified and the amplicons were directly ligated with Illumina TruSeq adaptors and sequenced on HiSeq2000 platform. Mapping was done by BWA to the target region, and uniquely mapped read pairs with proper distance and orientation were selected. More than 98% of the exonic target regions were covered with a depth  $\geq 100$ . We filtered out reads with a mapping quality  $< 10$  and base calls with base quality  $< 10$ . Base calls with a depth  $\geq 100$  were used for the analysis. We identified variants with frequency  $\geq 0.05$ . Variants found in more than one individual in the 1000 Genome database<sup>45</sup> were discarded. We performed Sanger sequencing verification for the predicted candidates in the both cancer and matched normal tissues.

**RNA sequencing.** RNA-seq was carried out for 25 LCBs and 44 HCCs for which high-quality RNA was available among the 30 LCBs and the 60 HCCs. Total RNA was extracted by Trizol from the frozen liver cancer tissues and the corresponding non-cancerous liver tissues and quality and quantity were evaluated by Bioanalyzer (Agilent). The high-quality RNA was subjected to polyA+ selection and chemical fragmentation, and 100–200 base RNA fraction was used to construct complementary DNA libraries according to Illumina's protocol. RNA-seq was performed on HiSeq2000 using the standard paired-end 101 bp protocol.

**Analysis of RNA sequencing data.** First, all sequencing reads were aligned to the known transcript sequences of UCSC known gene database (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/knownGene.txt.gz>) using Bowtie<sup>46</sup>, with -a --best --strata -m 20 -v 3 options, and the coordinates of the aligned reads were converted to the human reference sequence (hg19). Then, reads unaligned in the above step were aligned to the human reference sequence (hg19) and as well as HBV sequence (AP011098) using Blat<sup>47</sup>, with -stepSize = 5 -repMatch = 2253, and aligned reads by Bowtie or Blat were combined together. For each short read, the alignment positions with the maximum number of matched bases were adopted, and mapping quality for each read was assigned to as follows: for a location  $a$ , let  $B(a)$  denote the number of matched bases and let  $a_{\text{best}}$  denote the best location selected arbitrarily from those with the maximum number of matched bases.

$$\min \left( 100 - 10 \times \log_{10} \left( 1 - \frac{1}{\sum_a 0.02^{B(a_{\text{best}}) - B(a)}} \right) \right)$$

Finally, sorting and PCR duplicate removal of short reads were performed by using Picard (<http://picard.sourceforge.net/>). For quantification of expression values, we used a slightly modified version of RPKM (reads per kilobase of exon per million mapped reads) measures<sup>48</sup>. After removing improperly aligned or low-quality (mapping quality  $< 60$ ) sequencing reads, the number of bases on each exonic region for each refSeq genes (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz>) were counted. Then, the numbers of bases were normalized as per kilobase of exon and per 100 million of aligned bases. Finally, the expression value of each gene was determined by choosing the maximum of multiple refSeq genes, if any, corresponding to the gene symbol.

## References

- Blechacz, B. & Gores, G. J. Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment. *Hepatology* **48**, 308–321 (2008).
- Shin, H. R. *et al.* Epidemiology of cholangiocarcinoma: an update focusing on risk factors. *Cancer Sci.* **101**, 579–585 (2010).

3. Sia, D., Tovar, V., Moeini, A. & Llovet, J. M. Intrahepatic cholangiocarcinoma: pathogenesis and rationale for molecular therapies. *Oncogene* **32**, 4861–4870 (2013).
4. Sriya, B. *et al.* Liver fluke induces cholangiocarcinoma. *PLoS Med.* **4**, e201 (2007).
5. Lee, C. H. *et al.* Viral hepatitis-associated intrahepatic cholangiocarcinoma shares common disease processes with hepatocellular carcinoma. *Br. J. Cancer* **100**, 1765–1770 (2009).
6. Sekiya, S. & Suzuki, A. Intrahepatic cholangiocarcinoma can arise from Notch-mediated conversion of hepatocytes. *J. Clin. Invest.* **122**, 3914–3918 (2012).
7. Xu, J. *et al.* Intrahepatic cholangiocarcinoma arising in chronic advanced liver disease and the cholangiocarcinomatous component of hepatocellular cholangiocarcinoma share common phenotypes and cholangiocarcinogenesis. *Histopathology* **59**, 1090–1099 (2011).
8. Coulouarn, C. *et al.* Combined hepatocellular-cholangiocarcinomas exhibit progenitor features and activation of Wnt and TGFβ signaling pathways. *Carcinogenesis* **33**, 1791–1796 (2012).
9. Guest, R. V. *et al.* Cell lineage tracing reveals a biliary origin of intrahepatic cholangiocarcinoma. *Cancer Res.* **74**, 1005–1010 (2013).
10. Li, M. *et al.* Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat. Genet.* **43**, 828–829 (2011).
11. Totoki, Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nat. Genet.* **43**, 464–469 (2011).
12. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
13. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
14. Jiang, Z. *et al.* The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res.* **22**, 593–601 (2012).
15. Sung, W. K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
16. Huang, J. *et al.* Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. *Nat. Genet.* **44**, 1117–1121 (2012).
17. Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).
18. Ong, C. K. *et al.* Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nat. Genet.* **44**, 690–693 (2012).
19. Chan-On, W. *et al.* Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat. Genet.* **45**, 1474–1478 (2013).
20. Jiao, Y. *et al.* Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* **45**, 1470–1473 (2013).
21. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
22. Gammie, A. E. *et al.* Functional characterization of pathogenic human MSH2 missense mutations in *Saccharomyces cerevisiae*. *Genetics* **177**, 707–721 (2007).
23. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
24. Raimondo, G., Pollicino, T., Cacciola, I. & Squadrito, G. Occult hepatitis B virus infection. *J. Hepatol.* **46**, 160–170 (2007).
25. Tsai, W. L. & Chung, R. T. Viral hepatocarcinogenesis. *Oncogene* **29**, 2309–2324 (2010).
26. Komuta, M. *et al.* Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. *Hepatology* **47**, 1544–1566 (2008).
27. Fischer, A., Illingworth, C. J., Campbell, P. J. & Mustonen, V. EMU: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biol.* **14**, R39 (2013).
28. Borger, D. R. *et al.* Frequent mutation of isocitrate dehydrogenase (IDH)1 and IDH2 in cholangiocarcinoma identified through broad-based tumor genotyping. *Oncologist* **17**, 72–79 (2011).
29. Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
30. Romero, O. A. & Sanchez-Cespedes, M. The SWI/SNF genetic blockade: effects in cell differentiation, cancer and developmental diseases. *Oncogene* **33**, 2681–2689 (2014).
31. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
32. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405 (2012).
33. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
34. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
35. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
36. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
37. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
38. Gidekel Friedlander, S. Y. *et al.* Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. *Cancer Cell* **16**, 379–389 (2009).
39. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell of origin for human prostate cancer. *Science* **329**, 568–571 (2010).
40. Komuta, M. *et al.* Histological diversity in cholangiocellular carcinoma reflects the different cholangiocyte phenotypes. *Hepatology* **55**, 1876–1888 (2012).
41. Bosman, F. T. *et al.* WHO Classification of Tumors of the Digestive System 4th edn, 225–227 (IARC, 2010).
42. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
46. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
47. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
48. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

## Acknowledgements

The super-computing resource 'SHIROKANE' was provided by Human Genome Center, The University of Tokyo (<http://sc.hgc.jp/shirokane.html>). This work was supported partially by RIKEN President's Fund 2011, the Princess Takamatsu Cancer Research Fund and Takeda Science Foundation.

## Author contributions

A.F., Y. Shiraiishi, M.F., D.S., T.A., K.A.B., T.T. and H.N. performed data analyses. M.F., K.N., A.S., R.K. and H.N. performed WGS and validation sequencing study. M.F., A.F. and H.H.N. performed functional experiments. H.T., T. Shibuya and S. Miyano operated the super-computer system. Y.K., M.U., K.G., S.A., T.N., T. Shibata, K.A., H.O., K.S., Y. Shigekawa, S. Maruhashi, T.Y., O.I., H.A., H.O., S.H., M.Y., H.Y. and K.C. collected clinical samples and cell lines. A.F., M.F., Y. Shiraiishi, K.A.B. and H.N. wrote the manuscript. H.N. conceived the study and led the design of the experiments. A.F., M.K. and H.N. contributed to the funding for this study.

## Additional information

**Accession codes:** WGS data have been deposited in the EGA under the accession codes: EGAN00001187542, EGAN00001187543, EGAN00001187546, EGAN00001187547, EGAN00001187552, EGAN00001187553, EGAN00001187568, EGAN00001187569, EGAN00001187612, EGAN00001187613, EGAN00001187614, EGAN00001187615, EGAN00001187618, EGAN00001187619, EGAN00001187650, EGAN00001187651, EGAN00001187652, EGAN00001187653, EGAN00001187658, EGAN00001187659, EGAN00001187664, EGAN00001187665, EGAN00001187690, EGAN00001187691, EGAN00001187707, EGAN00001187708, EGAN00001187709, EGAN00001187710, EGAN00001187712, EGAN00001187713, EGAN00001187714, EGAN00001187715, EGAN00001187716, EGAN00001187717, EGAN00001187719, EGAN00001187720, EGAN00001187721, EGAN00001187722, EGAN00001187723, EGAN00001187724, EGAN00001187725, EGAN00001187726, EGAN00001187727, EGAN00001187728, EGAN00001187729, EGAN00001187730, EGAN00001187731, EGAN00001187732, EGAN00001187733, EGAN00001187734, EGAN00001187735, EGAN00001187736, EGAN00001187737, EGAN00001187738, EGAN00001187739, EGAN00001187740, EGAN00001187741, EGAN00001187742, EGAN00001187744 and EGAN00001187743 (summarized in Supplementary Table 2).

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interest.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Fujimoto, A. *et al.* Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. *Nat. Commun.* 6:6120 doi: 10.1038/ncomms7120 (2015).