Dietary cholesterol promotes steatohepatitis related hepatocellular carcinoma through dysregulated metabolism and calcium signaling

Jessie Qiaoyi Liang¹, Narcissus Teoh², Lixia Xu¹, Sharon Pok², Xiangchun Li¹, Eagle S.H. Chu¹, Jonathan Chiu¹, Ling Dong³, Evi Arfianti², W. Geoffrey Haigh⁴, Matthew M. Yeh⁵, George N. Ioannou⁴, Joseph J.Y. Sung¹, Geoffrey Farrell² & Jun Yu¹

The underling mechanisms of dietary cholesterol and nonalcoholic steatohepatitis (NASH) in contributing to hepatocellular carcinoma (HCC) remain undefined. Here we demonstrated that high-fat-non-cholesterol-fed mice developed simple steatosis, whilst high-fat-high-cholesterol-fed mice developed NASH. Moreover, dietary cholesterol induced larger and more numerous NASH-HCCs than non-cholesterol-induced steatosis-HCCs in diethylnitrosamine-treated mice. NASH-HCCs displayed significantly more aberrant gene expression-enriched signaling pathways and more non-synonymous somatic mutations than steatosis-HCCs (335 ± 84/sample vs 43 ± 13/sample). Integrated genetic and expresional alterations in NASH-HCCs affected distinct genes pertinent to five pathways: calcium, insulin, cell adhesion, axon guidance and metabolism. Some of the novel aberrant gene expression, mutations and core oncogenic pathways identified in cholesterol-associated NASH-HCCs in mice were confirmed in human NASH-HCCs, which included metabolism-related genes (ALDH18A1, CAD, CHKA, POLD4, PSPh and SQLE) and recurrently mutated genes (RYR1, MTOR, SDK1, CACNA1H and RYR2). These findings add insights into the link of cholesterol to NASH and NASH-HCC and provide potential therapeutic targets.
Hepatocellular carcinoma (HCC) is the second leading cause of cancer death in men and sixth in women worldwide. China alone accounts for about 50% of the total case burden and deaths worldwide. While hepatitis B and C viruses (HBV and HCV) remain the most important risk factors, virus-related HCC is expected to decrease in the near future due to widespread adoption of HBV vaccination and curative HCV treatments. However, the prevalence of obesity and type 2 diabetes is increasing. This is associated with an increase in non-alcoholic fatty liver disease (NAFLD), including its histologically progressive form, non-alcoholic steatohepatitis (NASH). Strong evidence has emerged that obesity and NASH are major risk factors for HCC.

Further, there is growing evidence from case series that HCC can occur in patients with non-cirrhotic NASH. Dietary cholesterol has been shown to play a role in the development of steatohepatitis in both animal models and human, and in epidemiological studies cholesterol intake is an independent risk factor for HCC. However, the mechanism by which cholesterol promotes NASH and leads to HCC development is unclear, and it has also not been established whether the resultant HCCs harbour a different set of genetic mutations and other molecular changes that differs from HCCs in other etiopathogenic settings.

In the present study, we explored the role of dietary cholesterol in contributing to NASH and NASH-HCC development in mice fed high-fat diets with or without high cholesterol. We found that animals fed high-fat with high cholesterol (HFHC) showed NASH development, whereas animals fed high-fat without cholesterol (HF diet) developed simple steatosis. HCCs were more numerous and larger in HFHC-fed than in HF-fed mice exposed to diethylnitrosamine (DEN). Paired tumor and adjacent non-tumorous liver from mice fed HFHC vs. HF diets were analyzed by expression profiling and whole-exome sequencing to seek aberrant gene expression, genomic mutations and signaling networks that could be enriched in the development of dietary cholesterol-induced liver tumors. Finally, to establish the clinical relevance of the discoveries in murine models, we searched for similar, disease-specific gene expression changes and somatic mutations in human NASH-HCCs.

**Results**

**Dietary cholesterol causes NASH in mice fed high-fat diet.** To explore the effects of dietary cholesterol on fatty liver disease, we fed mice either normal chow, HF, or HFHC diets (Fig. 1a). HF- and HFHC-fed mice gained more weight than controls fed normal chow (both \( P < 0.001 \)) (Fig. 1b). Likewise, both HF- and HFHC-fed mice were hyperglycemic (both \( P < 0.001 \)) and displayed impaired glucose tolerance (both \( P < 0.05 \)) compared to control mice (Fig. 1c). Although HF- and HFHC-fed mice were similarly obese, HFHC-fed mice showed significantly higher liver weight and liver/body weight ratios (both \( P < 0.001 \)) (Fig. 1d). Further, HFHC-fed mice showed significantly increased hepatic triglyceride, free cholesterol and cholesterol ester contents, as well as serum cholesterol compared to HF-fed mice (Fig. 1e). Histologically, only steatosis was found in the non-tumorous liver of HF-fed mice, while those from HFHC-fed mice showed steatosis, hepatocyte ballooning and inflammatory cell infiltration sufficient to be deemed NASH (Fig. 1f). Accordingly, steatosis and lobular inflammation scores were significantly higher in HFHC vs. HF livers, contributing to significantly higher NAFLD activity scores in HFHC livers (all \( P < 0.05 \); Fig. 1g).

**Dietary cholesterol augments hepatocarcinogenesis.** To explore the role of dietary cholesterol in HCC development, we compared DEN-induced hepatocarcinogenesis in mice fed normal chow, HF, or HFHC diets. HCC incidence was numerically higher in HF-fed mice (90%) and HFHC-fed mice (100%) compared with normal chow-fed controls (67%). Impressively, HCC multiplicity and size were markedly increased in HFHC-fed versus HF-fed mice (\( P < 0.0001 \); Fig. 1h). Furthermore, 40% of HCCs in HFHC-fed mice but none in HF-fed mice showed lung metastases (Supplementary Fig. 1). These findings demonstrate that high dietary cholesterol augments high-fat diets in promoting hepatocarcinogenesis, in association with inducing NASH liver pathology.

**Dietary cholesterol upregulated inflammatory genes.** In light of the effect of cholesterol on NASH development, we compared the gene expression profiles of NASH livers from HFHC-fed mice and steatosis livers from HF-fed mice. We identified 634 upregulated genes and 248 downregulated genes (\( \geq 2 \) or \( \leq -2 \) fold, \( P < 0.05 \)) (Fig. 2a). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that, in HFHC-fed mouse liver, aberrant gene expression was mainly enriched in inflammation, metabolism and cancer-related pathways (Fig. 2b). The affected inflammatory pathways ranged from macrophage infiltration (upregulation of macrophage markers), early inflammatory response (signaling of cytokines, chemokines and interleukins), late-phase fibrotic response to NASH, and transforming growth factor (TGF)-\( \beta \) and Wnt-related signaling (Fig. 2c, d).

Dysregulation of metabolism and cancer-related pathways (metabolic pathways, pathways in cancer, calcium signaling, insulin signaling, cell adhesion, axon guidance, etc) could possibly function in promoting malignant transformation of HFHC-livers (Supplementary Fig. 2). These findings demonstrate the necro-inflammatory nature of gene expression changes in liver in response to HFHC diet that might be associated with NASH and with accelerated development of HCC.

**Aberrant gene expression in murine NASH- and steatosis-HCCs.** To understand the molecular basis for accentuated hepatocarcinogenesis in HFHC-fed vs. HF-fed mice, we compared gene expression profiles of NASH-HCC with steatosis-HCC directly. We noted that 4,660 genes were aberrantly expressed (2-fold or more) in NASH-HCC versus steatosis-HCC. Because many of these genes could include those pertinent to the disease process of NASH versus steatosis, we further analyzed the differential gene expression in HCCs as compared to adjacent non-tumorous livers, and then compared cancer-related expression changes between the two groups of HCCs. Similar numbers of aberrantly expressed genes were identified in NASH-HCCs and steatosis-HCCs, and more genes were upregulated (315 and 364 respectively) than downregulated (40 and 47, respectively) in both groups (Fig. 3a). Cancers that developed between the two dietary groups shared about half (164) of upregulated genes. These were enriched in 15 pathways (Fig. 3b), including pathways important in cancer (ErkB signaling, MAPK signaling, PPAR signaling, etc), necro-inflammation (TGF-\( \beta \) signaling, chemokine signaling, etc) and cellular metabolism. More pathways were specifically dysregulated in NASH-HCCs, including those previously reported in NAFLD (calcium, insulin, hedgehog and adipocytokine signaling), and cancer (axon guidance and cell adhesion). Notably, most of the pathways dysregulated in NASH-HCCs (vs.

---

**ARTICLE**

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06931-6 | www.nature.com/naturecommunications
**Fig. 1** Cholesterol augmented high-fat (HF) diet in accelerating hepatocarcinogenesis in mice.  

**a** Schematic illustration of the treatment of mice. Male C57BL/6J mice were administered a single injection of the carcinogen diethylnitrosamine (DEN) intraperitoneally at age 15 days and fed normal chow \((n = 9)\), HF \((n = 10)\) or high-fat high-cholesterol (HFHC; \(n = 10\)) diets starting from 6 weeks of age till 8 months.  

**b** Body weights of mice fed HF or HFHC diets were similar and both significantly higher than normal chow (NC)-fed mice.  

**c** Fasting blood glucose and intraperitoneal glucose tolerance test (IPGTT) levels in mice fed NC, HF or HFHC diets.  

**d** Mice fed HFHC diet had significantly higher liver weights and liver-body weight ratios than HF-fed mice.  

**e** Levels of hepatic triglyceride, hepatic free cholesterol and cholesterol ester, and serum cholesterol were measured in HF- and HFHC-fed mice.  

**f** Representative H&E staining histological images (f) and NAFLD scores (g) of liver tissues from HF- and HFHC-fed mice.  

**h** HCC incidence, number and size in NC-, HF- and HFHC-fed mice after DEN injection. Tumor number: few < 5; multiple ≥ 5. Tumor size: small, all < 5 mm^3; large, at least one ≥ 5 mm^3. The data are shown as means ± SE for IPGTT and means ± SD for others. Data in b-g between each two groups were compared using ANOVA Tukey’s multiple comparison tests.
Late-phase response

Steatohepatitis

surrounding livers) were already dysregulated in NASH (vs. steatosis) by aberrant gene expression (Figs. 2b, 3b). These include calcium signaling, pathways in cancer, ErbB signaling, MAPK signaling, and chemokine signaling, metabolic pathways etc. Genes commonly upregulated in both NASH- and steatosis-associated HCCs when compared to adjacent non-tumorous liver (such as Cdhl1, Pak1, Sgp, and others) may be important in hepatocarcinogenesis. We also identified genes aberrantly expressed only in NASH-HCCs, such as downregulation of Cfd (complement factor D, encoding adipsin), and upregulation of Ddit3, Itgα6 and others (Fig. 3c). These findings implicate sets of genes and pathways specifically associated with the acceleration of hepatocarcinogenesis occasioned by addition of cholesterol to a HF diet through expressional dysregulation.

Aberrant gene expression verified in human NASH-HCCs. To establish whether the present dietary model of NASH-enhanced hepatocarcinogenesis has relevance to human disease, we determined whether genes aberrantly expressed in mouse NASH-HCCs (Fig. 3c) were also differentially expressed in human HCCs obtained from patients with histologically pedigreed NASH. Among the 17 available human NASH-HCCs, we verified 12 genes to be significantly up- or downregulated in NASH-HCCs compared to their surrounding liver by RNA sequencing (all \( P < 0.05 \) by paired \( t \) test and false discovery rate (FDR) < 0.15; Fig. 4a). The adipisin-encoding gene CFD was significantly downregulated in NASH-HCCs. Upregulated genes included two cell adhesion molecules (ALCAM and ITGA6), 3 MAPK signaling genes (DDIT3, MAP3K6 and PAK1) and 6 other metabolic genes (ALDH1A1, CAD, CHKA, POLD4, PSHF, SQLE). Differential expression of seven of these genes (CFD, ALDH1A1, CHKA, DDIT3, ITGA6, PSHF and SQLE) was replicated in another set of 12 paired NASH-HCCs and adjacent non-tumor livers by RT-qPCR (all \( P < 0.05 \) by paired \( t \) test; Fig. 4b). The generalizability of these particular expressional changes between species provides evidence of their likely etiopathogenic importance in HCC related to NASH.

NASH-HCCs harbor more mutations than steatosis-HCCs in mice. To identify genomic alterations associated with HCC development in HFHC- and HF-fed mice, whole-exome sequencing was performed. The somatic mutation spectra of both experimental groups were similar, with no observed differences in nucleotide base changes. However, NASH-HCCs harbored significantly more somatic mutations than steatosis-HCCs (circos illustration in Fig. 5a). The average mutated gene numbers were \( 452 \pm 119 \) vs. \( 58 \pm 20 \) in total somatic mutations, and \( 335 \pm 84 \) vs. \( 43 \pm 13 \) in non-synonymous somatic mutations (both \( P < 0.05 \); Fig. 5b; Supplementary Data 1). Only 28 mutated genes overlapped between NASH- and steatosis-associated HCCs. These findings demonstrate distinct genetic alterations in HCCs related to NASH produced by HFHC diet vs. HCCs in mice consuming only a HF diet. Differences between the two groups
were evident both in the number of genes mutated and the particular genes affected.

**Identification of mutations related to NASH-HCC in mice.** Several of the mutated genes we identified in murine NASH-HCCs have been reported as drivers in other human cancers, such as Mtor, Sdk1, Braf, Pik3cb, Ctnnd1, Akt3, and Nos2. We identified 82 genes to be recurrently mutated in two or more NASH-HCCs (Fig. 5c; Supplementary Data 2); these included two genes mutated in 4/5 NASH-HCCs (Ryr1 and Sdk1); 8 genes mutated in 3/5 NASH-HCCs (Epha8, Pdcd15, Fat2, Cep152, Tnn, Rsfpl, Aox3l1 and Pkd1l3) and 71 genes mutated in 2/5 NASH-HCCs (notably Mtor, Ryr2, Cacn1h, Col7a1, Fgfbp, Adam29, Gpr98 and Pelo). Only 2 of the 82 genes (Tnn and Obscn) were also mutated in one HF-associated steatosis-HCC each, and only one gene (Kif3c) was recurrently mutated in steatosis-HCCs. These results suggest that mutations of these known or potential cancer driver genes contribute to the development of cholesterol-associated NASH-HCC. The recurrent mutations suggest that Mtor and Sdk1 may also function as cancer drivers in NASH-HCC, while other recurrently mutated genes could represent novel candidates associated with NASH-HCC.

**Pathways enriched by gene mutations in NASH-HCC in mice.** To investigate the signaling networks enriched by mutations in NASH- and steatosis-HCCs, we conducted KEGG pathway analysis. Seven pathways were found to be significantly enriched by gene mutations in three or more NASH-HCCs, whereas no pathway was significantly affected in steatosis-HCCs (Fig. 6a and Supplementary Data 3). Pathways affected in NASH-HCCs include metabolic, insulin and calcium signaling, cell adhesion and tight junction molecules, ABC transporters, and axon guidance. It is of particular interest that gene mutations expected to dysregulate calcium signaling were present in all five dietary cholesterol-associated NASH-
HCCs (Supplementary Table 1) as calcium signaling has been reported to be directly affected by cellular cholesterol content21,22 and plays an important role in cancer development23,24. Twenty-eight calcium signaling genes were mutated in NASH-HCCs, with 6 recurrently mutated (Ryr1, Ryr2, Cacna1d, Cacna1h, P2rx1 and Itpr1) all encoding calcium channel proteins, and Ryr1 was mutated in 4/5 samples (Fig. 6b). In NASH-HCCs, the 16 mutated genes involved in insulin signaling included insulin receptor (Insr) and some other well-known cancer-related genes (Mtor, Hras1, Akt3, etc.), with Mtor and Hras1 recurrently mutated (Fig. 6c). The 75 mutated genes involving in metabolism were expected to dysregulate 15 specific metabolic signaling pathways, including 6 associated with lipid metabolism (Supplementary Fig. 3b and Supplementary Table 2).

Dysregulated pathways by mutations and aberrant expression. Only 4.4% of mutated genes were differentially expressed in NASH-HCCs as compared to adjacent non-tumorous liver (Supplementary Fig. 4). This result indicates that genetic alteration contributed little effect to expressional aberration in NASH-HCC. However, five pathways were commonly and uniquely enriched by mutations and aberrant gene expression in NASH-HCCs; these include calcium, insulin signaling, cell adhesion, axon guidance and metabolic pathways (Fig. 6d). Furthermore, these five pathways were significantly enriched by aberrant gene expression in HFHC-induced NASH as compared to HF-induced steatosis (Fig. 2b). These findings indicate that dysregulation of the five core pathways is involved in the transition of steatosis to NASH (by aberrant gene expression) and also in NASH-related hepatocarcinogenesis (by aberrant gene expression and mutations).

Mutations verified in human NASH-HCCs. Among the 80 recurrently mutated genes identified in mouse NASH-HCCs, 66 have human orthologs (Supplementary Data 2). We investigated the exome-sequencing data to examine for the presence of these mutations in 37 human NASH-HCCs (the 17 in-house samples as mentioned earlier, and 20 from The Cancer Genome Atlas (TCGA) database). We found 38 of the 66 genes were mutated in human NASH-HCCs, including 5 calcium signaling genes (Supplementary Data 4). Of note, 21 genes were recurrently mutated in at least two human NASH-HCCs (Fig. 7a). These included Mtor (insulin signaling, and mutational cancer driver in kidney and lung)13, Sdk1 (cell adhesion molecule and potential driver of asbestos-associated cancers)14, and three calcium signaling genes (Ryr1, Ryr2 and Cacna1H).

Mutations in calcium signaling pathway in human NASH-HCCs. As already noted, calcium signaling genes were most
reproducibly mutated in murine NASH-HCCs, prompting us to further investigate them in human NASH-HCCs. Of note, we identified 19 genes in calcium signaling to be recurrently mutated in at least two of the 37 human NASH-HCCs. Similar to murine NASH-HCCs, most of these genes mutated in human NASH-HCCs (11/19) encode calcium channel proteins (RYR1, RYR2, CACNA1B, CACNA1E, CACNA1H, CACNA1I, GRIN2C, ATP2A2, ATP2B4, SLC8A1, and ITPR3) (Fig. 7b). Eight genes showed significantly higher mutation frequencies in NASH-HCCs (the combined cohort \( n = 37 \)) or TCGA cohort \( n = 20 \)) as compared to other HCCs not complicating NASH/NAFLD (TCGA, \( n = 353 \)), and two of them encode calcium channel proteins (RYR1 and CACNA1H) (Fig. 7c). RYR1 was mutated most frequently in both murine (4/5 = 80%) and human (5/37 = 13.5%) HCCs associated with NASH pathology, and RYR1 was less often mutated in other HCCs (7.37%, \( n = 353 \), TCGA data) (Fig. 7c). The 9 RYR1 mutations identified in mouse and human NASH-HCCs were located across the whole gene, with 3 stop-gain or truncating mutations inferring likely loss-of-function mutations (Fig. 7d). These findings suggest that genetic disruptions of calcium channels may alter calcium homeostasis in a way that contributes to NASH-related carcinogenesis.

**Discussion**

Hepatic free cholesterol has been implicated in hepatic lipotoxicity\(^ {25} \), and may thereby activate inflammatory recruitment to fatty livers of HIFHC-fed mice to induce NASH. In this study, aberrant gene expression affecting inflammatory signaling pathways (macrophage markers, cytokine, chemokine, interleukin, TGF-β and Wnt-related signaling) was identified in HIFHC-associated NASH compared with HF-associated steatosis. This is consistent with a previously report that aberrant gene expression was involved in abnormal inflammation during NAFLD progression\(^ {26} \). Although others have found that genes involved in Wnt signaling may be downregulated in NASH\(^ {27} \), our study showed upregulated expression of Wnt signaling pathways in NASH.
genes in NASH compared to steatosis, including Ctnnb1 and Myc (Fig. 2). As similar upregulated expression of Wnt signaling genes has been reported in HCC, our finding supports the hypothesis that NASH is more prone to malignant transformation than steatosis through inducing inflammatory pathways, including those pertinent to Wnt signaling.

One of our important and novel findings is that the aberrantly expressed genes and pathways accelerated DEN-induced HCC based on NASH. We further confirmed that some of these genes were also aberrantly expressed in human NASH-associated HCCs as compared to their adjacent non-tumorous liver. We demonstrated for the first time the aberrant expression of cancer-related genes (ALCAM, ITGA6, DDIT3, MAP3K6 and PAK1) and metabolism-related genes (ALDH18A1, CAD, CHKA, POLD4, PSPH, SQLE and CFD) in human NASH-HCCs. These findings highlight their particular importance in the development of

---

**Table:**

| Pathway                          | Affected HFHC | Mutated gene | Recurrently mut.gene | Genes in pathway | Mutated genes in HFHC (NASH-HCCs) |
|----------------------------------|---------------|--------------|----------------------|------------------|-----------------------------------|
| 1 Calcium signaling             | 5             | HFHC 28      | 1                    | 2                | Ryr1, Ryr2, Caonath, Caona1d, Itpr1, P2tx1, etc |
| 2 Insulin signaling             | 3             | HFHC 16      | 0                    | 2                | Mtor, Hras1, Akt3, Braf, Insr, Shc1, Pik3cb, etc |
| 3 Cell adhesion                 | 3             | HFHC 12      | 0                    | 2                | Vcan, Jam2, Cntnap2, Itga6, etc |
| 4 Axon guidance                 | 4             | HFHC 17      | 0                    | 2                | Epha8, Hras1, Limk2, Elna3, etc |
| 5 ABC transporters              | 3             | HFHC 10      | 0                    | 2                | Abcb1b, Abcc2, Abcc9, Abcb4, Cflr, etc |
| 6 Tight junction                | 3             | HFHC 17      | 1                    | 2                | Hras1, Jam2, Actn2, Ctnna3, Exoc3, etc |
| 7 Metabolic pathways            | 3             | HFHC 75      | 3                    | 2                | Aox3f1, Cyp2b19, Ext1, Prps113, Tm7ifs2, etc |

**Fig. 6** Pathways dysregulated by genetic alterations. a Seven pathways were significantly dysregulated in three or more NASH-HCCs from HFHC-fed mice (≥ 3 genes involved in each sample, adjusted P < 0.05). b Mutated genes in calcium signaling pathway are indicated. 28 genes were mutated in NASH-HCCs, 6 of which were recurrently mutated. Only one gene in calcium signaling pathway was mutated in one of the steatosis-HCCs. Color-filled icons show mutated genes. Recurrently mutated genes were highlighted with affected NASH-HCC numbers denoted. c Mutated genes in insulin signaling are indicated. 16 gene were mutated in NASH-HCCs, 2 of which were recurrently mutated, while no gene was mutated in steatosis-HCCs. d Five pathways were enriched commonly by mutations and aberrant gene expression in NASH-HCCs.
NASH-related HCC. Tumor-promoting roles of MAP3K6, PAK1, ALCAM and ITGA6 have been reported, thus their upregulation may also function in the development of NASH-related HCC. DDIT3 is a multifunctional transcription factor in endoplasmic reticulum stress and is associated with inflammatory responses, thus upregulated DDIT3 should be involved in the development of NASH-HCC. Uproregulation of the metabolic genes (ALDH1B1A1, CAD, CHKA, POLD4, PSPH, SQLE) may function in accelerating cellular metabolism. Of note, SQLE (squalene epoxidase) encodes a rate-limiting enzyme in cholesterol biosynthesis and exerts oncogenic roles in cancers. Our team has recently revealed that upregulation of SQLE plays an important oncogenic role in driving NASH-HCC. CFD encodes human adipin that functions as an adipokine and regulates insulin secretion. Downregulation of CFD expression may be associated with dysregulation of insulin signaling in NASH-HCC. Collectively, dietary cholesterol induced hepatic aberrant expression of genes and their associated necro-

**Fig. 7** Validation of recurrently mutated genes in human NASH-HCCs. a Recurrently mutated genes identified in mouse NAHS-HCCs were checked in 37 human NASH-HCCs (17 in-house and 20 from TCGA). Recurrent mutations in 21 genes were verified in human NASH-HCCs. b Recurrently mutated calcium signaling-related genes in human NASH-HCCs. c Mutation frequencies of eight genes were higher in HCCs with risk history of NAFLD than others without NAFLD (Chi-square test). *P < 0.05 and **P < 0.001 for all NAFLD-HCCs (in-house and TCGA, n = 37) versus other HCCs; #P < 0.05 for NAFLD-HCCs from TCGA (n = 20) vs. other HCCs; ns non-significant as compared with other HCCs. d Schematic illustration of the mutation sites in Ryr1 identified in mouse and human NASH-HCCs.
inflammatory (TGF-β signaling, chemokine signaling, etc), oncogenic signaling pathways (MAPK signaling, axon guidance and cell adhesion, etc) and metabolic pathways to contribute to the development of NASH-HCC.

In contrast to liver cancers in HFHC-fed mice, significantly fewer genes were mutated during hepatocarcinogenesis in HF-fed mice. It therefore seems likely that alterations of gene expression contribute more importantly to hepatocarcinogenesis than do gene mutations in HF-fed mice. In addition to the anticipated dysregulated expression of cancer-related and metabolic pathways, 13 out of the 16 genes related to cholesterol synthesis were significantly upregulated (all P < 0.05; Supplementary Fig. 5). We interpret this phenomenon as indicating that without an augmented level of cholesterol intake, hepatocytes and/or HCC precursor cells are required to undergo de novo cholesterol synthesis to promote cancer development. This concept supports a previous hypothesis that cancer cells need excess cholesterol and intermediates of the cholesterol biosynthesis pathway to maintain cell proliferation, presumably because a large quantum of cholesterol is required for the synthesis of cell membrane36. High levels of dietary cholesterol could therefore accelerate hepatocarcinogenesis by bypassing the requirement for enhanced endogenous cholesterol synthesis. Recent data have shown that major cholesterol metabolites, such as 27-hydroxycholesterol and 6-oxocholestan-3β,5α-diol, display tumor-promoter effects the involvement of pro-inflammatory (TGF-

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06931-6

In the present study, the calcium signaling gene RYR1 that was mutated most frequently in both murine and human NASH-HCCs appears to be of pathogenic significance. Calcium signaling plays an important physiological role in metabolism, and its dysregulation has previously been implicated in cancer development23,24. Furthermore, calcium signaling is known to be affected by cellular cholesterol content21,22. Intracellular calcium ions are important second messengers that regulate gene transcription, cell proliferation, migration and death. Thus, altered intracellular calcium homeostasis is involved in tumor initiation, angiogenesis, tumor progression and metastasis. Some important calcium channels, transporters and calcium-ATPases are altered in human cancer patients58. It is therefore salient to note that most recurrently mutated calcium signaling genes identified in NASH-HCCs (all 6 in mice and 11/19 in human) encode key calcium channels; genetic disruptions of these channels may alter calcium homeostasis and consequently contribute to carcinogenesis. These findings highlight the potential importance of the dysregulated calcium signaling during hepatocarcinogenesis in both experimental and clinical NASH. Targeting derailed calcium signaling by blockers, inhibitors or regulators for calcium channels/transporters/pumps in cancer therapy has become an emerging research area60. The altered calcium signaling-related genes identified in this study with available compounds for targeted therapy include RYR (encoded by RYR1 and RYR2) and IP3R3 (encoded by ITPR3) channels, for which chemotherapeutic agents include RYR agonists (4-chloro-m-cresol, caffeine and its analogs) and IP3R3 inhibitors (heparin and caffeine). Notably, intake of two or more cups of coffee each day reduces HCC incidence by more than 30% in patients with cirrhosis51. Future study to investigate the efficacy of these compounds targeting RYR and IP3R channels in inhibiting NASH-HCC is warranted.

In summary, this study provides the first systematic overview of altered gene expression, gene mutations and the affected signaling networks in HCC developing in the context of NASH compared to steatosis using appropriate murine models (Fig. 8). Key findings are that dietary cholesterol causes NASH via dysregulating expression of genes involved in inflammation and...
metabolism, and augments the accelerated development of HCC through induction of oncogenic mutations and gene expression. In particular, we identified novel aberrant gene expression, mutations and core oncogenic pathways, most notably those linked to calcium transport channels that contribute to HCC with cholesterol-induced NASH, and demonstrated analogous changes in clinical samples. Further studies should interrogate the functional consequences of such molecular changes and test whether they can be exploited pharmacologically as chemoprevention or therapy of NASH-related HCC.

Methods
Animals and diets. From 6 weeks of age, male C57BL/6J mice were randomly fed one of three diets (Specialty Feeds, Glen Forrest, Western Australia) for 6 weeks: 1) normal chow (n = 9), 2) high saturated-fat (HF) diet, 23% fat, 45% carbohydrate, 20% protein, 0% cholesterol w/w (n = 10), and 3) combined high saturated-fat and high cholesterol (HFHC) diet (23% fat, 45% carbohydrate, 20% protein, 0.19% cholesterol w/w) (n = 10). To accelerate HCC development in mice, a single injection of chemical carcinogen DEN (10 mg/kg, Sigma-Aldrich, St. Louis, MO) intraperitoneally was applied at age 15 days (0.9% saline to controls) to induce HCC. This is the most widely used chemical to induce liver cancer in mice.

Assessment of metabolic phenotypes and liver histology. Blood glucose and lipids were measured using automated techniques for clinical pathology at the Canberra Hospital. To evaluate insulin sensitivity, we performed intraperitoneal glucose tolerance tests (IPGTT) 2 weeks prior to sacrifice. Hepatic triglyceride, free cholesterol, cholesterol ester fractions were quantified using high-performance liquid chromatography and results normalized to liver wet weight (g), as described previously. Liver sections (5 µm) were stained with hematoxylin and eosin and assessed blind by an expert liver pathologist (Matthew M Yeh) according to the NAFLD activity score.

Whole-exome sequencing. DNA samples of paired tumor and adjacent non-tumorous livers of 3 HF-fed mice and 5 HFHC-fed mice were subjected to whole-exome sequencing. Exome capture was performed using the SureSelect Target Enrichment System (Agilent, Santa Clara, CA). Captured libraries were sequenced by HiSeq2000 (Illumina, San Diego, CA). Sequencing mean depths were > 100-fold and coverages of the targeted exome were > 99% for all samples (Supplementary Table 3).

Mouse genome alignment. The whole-exome sequencing reads were mapped to the UCSC mm10 reference genome (http://genome.ucsc.edu/) using the Burrows-Wheeler Aligner. The likely PCR duplications with the same match interval on the genomic sequence duplications were removed by Samtools. Local realignment around small insertions and deletions (Indels) were performed using GATK to improve alignment quality and help reduce the false discovery rate in the following mutation detection analysis.

Single-nucleotide variants (SNVs) calling. Candidate SNVs with mutated allele frequencies > 15% in cancer tissue and no more than 2% in paired normal tissue (Fisher’s exact test, P < 0.05) were filtered by the following thresholds: high-quality (Q > 10) coverage ≥ 10X for each sample; reads supporting the mutated allele not a result of sequencing error (Binomial test, f = 0.1, P > 0.01); mutation scores for mutated alleles not lower than normal alleles (Wilcoxon rank sum test, P > 0.01); mutated alleles not come from repeatedly aligned reads (Fisher’s exact test, P > 0.01). Mutant alleles should not be enriched in 10 bps of 5’ or 3’ ends of reads (Fisher’s exact test, P > 0.01). Annotation was performed by ANNOVAR.

Figs. 1A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z. Schematic summary of this study. NASH development promoted by dietary cholesterol is associated with aberrant gene expression linked with activated inflammatory signaling, dysregulated metabolic and cancer-related pathways. Dietary cholesterol-induced NASH-HCC development in mice is associated with aberrant gene expression, genomic mutations and the associated core pathways. The identified aberrant gene expression and mutations were verified in human NASH-HCCs. NC, normal chow; HF, high-fat diet; HFHC, high-fat high-cholesterol diet.
Indels calling. Candidate somatic indels were firstly predicted by GATK SomaticIndelDetector with default parameters. Predicted indels were then filtered if (1) total coverage at the site < 20; (2) average mapping qualities of consensus indel-supporting reads < 30; (3) average number of mismatches per consensus indel-supporting read ≥ 2; (4) average quality of bases from consensus indel-supporting reads < 20; (5) median of indel offsets from the ends starts of the reads within 10 bp, or (6) percentage of forward- or reverse-aligned indel-supporting reads < 20%. Resulting somatic indels were filtered against dbsNP128. Annotation was performed by ANNOVAR.

Identification of known or potential cancer driver genes. Mutated genes identified in mouse models were cross-checked using the Catalog of Cancer Driver Mutations (version 2016.5: https://www.intogen.org/analysis/home).

Genome-wide expression analysis using microarrays. Gene expression profiles of tumors and adjacent non-tumor liver tissues from 3 HF- and 5 HHFC-fed mice (same samples as for exome sequencing) were analyzed using the SurePrint G3 Mouse GE 8 × 60 K Microarray Kit (Agilent Technologies, Palo Alto, CA), which contained 39,430 Entrez Gene RNAs. In brief, RNA was extracted using Qiazol reagent (Qiagen, Valencia, CA). The CDNA probes were prepared from 5 μg of total RNA labeled with Cy3-dUTP (red) or Cy5-dUTP (green) by reverse transcription (Amersham Biosciences, Piscataway, NJ). Two labeled CDNAs were competitively hybridized to the microarray. Signal intensities were analyzed using a SureScan microarray scanner (Agilent Technologies). Array data were presented as log base 2 ratio of the Cy5/Cy3 signals. Gene expression patterns between HF and HHFC groups were compared using unsupervised hierarchal clustering.

Pathway enrichment analyses. Mutated genes or differentially expressed genes were subjected to KEGG pathway enrichment analysis using Gene Set Analysis Toolkit V2 (https://bioinfo.vanderbilt.edu/webgestalt/)[61,62]. The hypergeometric test statistical method and the BH multiple test adjustment method were used. All mouse genes were used as reference. Pathways with at least 4 aberrantly expressed genes/3 mutated genes and adjusted P < 0.05 were considered as significantly enriched. WikiPathways analysis was also used for pathway identification and data visualization (http://www.wikipathways.org)[63].

Human samples, gene expression and mutation analyses. Recurrently mutated genes identified in tumors of HHFC-fed mice were further investigated in human NASH-HCC samples. Tumors and paired non-tumorous liver tissues from 17 patients diagnosed with HCC complicating NASH were obtained from Prince of Wales Hospital, The Chinese University of Hong Kong and Zhongshan Hospital of Fudan University. All patients gave informed consent for participation in this study. This study was approved by the ethics committee of the Chinese University of Hong Kong and the clinical research ethics committee of Zhongshan Hospital of Fudan University. Tumor and paired non-tumorous liver tissues were subjected to RNA sequencing and whole-exome sequencing. Gene expression was analyzed and compared between tumors and adjacent non-tumorous liver as our previous description[64]. Somatic mutations were called as our previous description[65]. Expression and somatic mutations in the genes of interest were then checked in the human HCC samples associated with NAFLD that had been analyzed by exome sequencing in TCGA project. Mutation data were downloaded from TCGA database at dbioPortal Cancer Genomics (http://www.cbioportal.org/public-portal/index.do).

Statistical analyses. All measurements are shown as means ± SD or means ± SEM where appropriate. ANOVA corrected for multiple comparisons using Tukey’s test was used to compare data of each two groups when multiple groups were involved, and multiplicity adjusted P values were reported for each comparison. Unpaired t tests were used for comparison of data involving only HF- and HHFC-fed groups. Paired t tests were used for comparison of expression levels between paired tumor and adjacent non-tumorous livers from mice and human. FDR was adjusted using the Benjamini–Hochberg method for RNA seq data. Mutation frequencies were compared by chi-square test. All statistical tests, except FDR, were performed using Graphpad Prism 5.0 (Graphpad Software Inc., San Diego, CA), and a two-tailed P value of < 0.05 was considered statistically significant.

Data availability. Exome sequencing data are deposited in NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRP115597 (https://www.ncbi.nlm.nih.gov/sra?view=study&term=PRJNA398450). Microarray data are deposited in EMBL-EBI ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-6002 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6002/).

Received: 28 November 2017 Accepted: 1 October 2018
Published online: 26 October 2018

References
1. Torre, L. A. et al. Global cancer statistics, 2012. CA Cancer J. Clin. 65, 87–108 (2015).
2. Chen, J. C. et al. Asian Consensus on the relationship between obesity and gastrointestinal and liver diseases. J. Gastroenterol. Hepatol. 31, 1405–1413 (2016).
3. Larson, S. C. & Wolk, A. Obesity, weight and risk of liver cancer: a meta-analysis of cohort studies. Br. J. Cancer 97, 1005–1008 (2007).
4. Chen, H. F., Chen, P. & Li, C. Y. Risk of malignant neoplasms of liver and biliary tract in diabetic patients with different age and sex stratifications. Hepatology 52, 155–163 (2010).
5. Williams, C. D. et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 140, 124–131 (2011).
6. Wong, V. W. et al. Prevalence of non-alcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. Gut 61, 409–415 (2012).
7. Starley, B. Q., Calagno, C. J. & Harrison, S. A. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. Hepatology 51, 1820–1832 (2010).
8. Van Rooyen, D. M. et al. Hepatic free cholesterol accumulates in obese, diabetic mice and causes nonalcoholic steatohepatitis. Gastroenterology 141, 1393–1403 (2011).
9. Ioannou, G. N., Morrow, O. B., Connole, M. L. & Lee, S. P. Association between dietary nutrient composition and the incidence of cirrhosis or liver cancer in the United States population. Hepatology 50, 175–184 (2009).
10. Matsuzawa, N. et al. Lipid-induced oxidative stress causes steatohepatitis in mice fed an atherogenic diet. Hepatology 46, 1392–1403 (2007).
11. Swiderska-Syn, M. et al. Hedghog pathway and pediatric nonalcoholic fatty liver disease. Hepatology 57, 1814–1825 (2013).
12. Estepe, J. M. et al. Adipocytokine expression associated with miRNA regulation and diagnosis of NASH in obese patients with NAFLD. Liver Int. 35, 1367–1372 (2015).
13. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43–49 (2013).
14. Maki-Nevala, S. et al. Driver gene and novel mutations in asbestos-exposed lung adenocarcinoma and malignant mesothelioma detected by exome sequencing. Lung 194, 125–135 (2016).
15. Davies, H. et al. Mutations of the Braf gene in human cancer. Nature 417, 949–954 (2002).
16. Tamborero, D. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci. Rep. 3, 2650 (2013).
17. Sebestyen, E., Zawisza, M. & Eyras, E. Detection of recurrent alternative splicing switches in tumour samples reveals novel signatures of cancer. Nucleic Acids Res. 43, 1345–1356 (2015).
18. Fanjul-Fernandez, M. et al. Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumor suppressors frequently mutated in laryngeal carcinomas. Nat. Commun. 4, 2531 (2013).
19. Toujani, S. et al. High resolution genome-wide analysis of chromosomal alterations in Burkitt’s lymphoma. PLoS ONE 4, e7089 (2009).
20. Heinecke, J. L. et al. Tumor microenvironment-based feed-forward regulation of NOS2 in breast cancer progression. Proc. Natl Acad. Sci. USA 111, 6323–6328 (2014).
21. Kannan, K. B., Barlos, D. & Hauser, C. J. Free cholesterol alters lipid raft structure and function regulating neutrophil Ca2+ entry and respiratory burst: correlations with calcium channel raft trafficking. J. Immunol. 178, 5253–5261 (2007).
22. Jie, Z., Kendrick, A., Quenby, S. & Wray, S. Contractility and calcium signaling of human myometrium are profoundly affected by cholesterol manipulation: implications for labor? Reprod. Sci. 14, 456–466 (2007).
23. Monteith, G. R., Davis, F. M. & Roberts-Thomson, S. J. Calcium channels and pumps in cancer: changes and consequences. J. Biol. Chem. 287, 31666–31673 (2012).
24. Chen, Y. F., Chen, Y. T., Chiu, W. T. & Shen, M. R. Remodeling of calcium signaling in tumor progression. J. Biomed. Sci. 20, 23 (2013).
25. Gan, L. T. et al. Hepatocyte free cholesterol lipotoxicity results from JNK1-mediated mitochondrial injury and is HMGB1 and TL1R4-dependent. J. Hepatol. 61, 1376–1384 (2014).
26. Lefebvre, P. et al. Interspecies NASF disease activity whole-genome profiling identifies a fibrogenic role of PPARalpha-regulated dermatopontin. JCI Insight 2, 92264 (2017).

27. Clarke, J. D. et al. Characterization of hepatocellular carcinoma related genes and metabolites in human nonalcoholic fatty liver disease. Dig. Dis. Sci. 59, 365–374 (2014).

28. Devis, L. et al. Activated leukocyte cell adhesion molecule (ALCAM) is a marker of recurrence and promotes cell migration, invasion, and metastasis in early-stage endometrioid endometrial cancer. J. Pathol. 241, 475–487 (2017).

29. Brooks, D. L. P. et al. ITGA6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. Mol. Cancer 15, 26 (2016). https://www.ncbi.nlm.nih.gov/pubmed/27001172

30. Eto, N., Miyagishi, M., Inagai, R., Fujita, T. & Nagakura, M. Mitogen-activated protein 3 kinase 6 mediates angiogenic and tumorigenic effects via vascular endothelial growth factor expression. Am. J. Pathol. 174, 1553–1563 (2009).

31. Ong, C. C. et al. Targeting p21-activated kinase 1 (PAK1) to induce apoptosis of tumor cells. P Natl Acad. Sci. USA 108, 7177–7182 (2011).

32. Liu, D. et al. Squalene epoxidase drives NAFLD-induced hepatocellular carcinoma and is a pharmaceutical target. Sci. Transl. Med. 10, eaap9840 (2018).

33. Brown, D. N. et al. Squalene epoxidase is a bona fide oncogene by amplification with clinical relevance in breast cancer. Sci. Rep. 6, 14935 (2016).

34. Sui, Z. H., Zhou, J. H., Cheng, Z. J. & Lu, P. H. Squalene epoxidase (SQLE) promotes the growth and migration of the hepatocellular carcinoma cells. Tumor Biol. 36, 6173–6179 (2015).

35. Lo, J. C. et al. Adipsin is an adipokine that improves beta cell function in diabetes. Cell 158, 41–53 (2014).

36. Cruz, P. M., Mo, H., McConathy, W. J., Sabnis, N. & Lacko, A. G. The role of cholesterol metabolism and cholesterol transport in carcinogenesis: a review of scientific findings, relevant to future cancer therapeutics. Front. Pharmacol. 4, 119 (2013).

37. Nelson, E. R. The significance of cholesterol and its metabolite, 27-hydroxycholesterol in breast cancer. Mol. Cell. Endocrinol. 466, 73–80 (2018).

38. Voisin, M. et al. Identification of a tumor-promoter cholesterol metabolite in human breast cancers acting through the glucocorticoid receptor. Proc. Natl Acad. Sci. USA 114, E9355–E9366 (2017).

39. Kloudova, A. et al. Expression of oxysterol pathway genes in oestrogen-positive breast carcinomas. Clin. Endocrinol. (Oxf.) 86, 852–862 (2017).

40. Tanaka, S. et al. Increased hepatic oxidative DNA damage in patients with nonalcoholic steatohepatitis who develop hepatocellular carcinoma. J. Gastroenterol. 48, 1249–1258 (2013).

41. Zhu, W. et al. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. Nat. Commun. 8, 4565 (2017).

42. Cancer Genome Atlas Research Network. Electronic address wbc, cancer genome atlas research n. comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 169, 1327–1341 (2017). e1323.

43. Ki Kim, S. et al. TERT promoter mutations and chromosome 8p loss are characteristic of nonalcoholic fatty liver disease-related hepatocellular carcinoma. Int. J. Cancer 139, 2512–2518 (2016).

44. Calderaro, J. et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. J. Hepatol. 67, 727–738 (2017).

45. Boughora, L. K. & DeBerardinis, R. J. Metabolic pathways promoting cancer cell survival and growth. Nat. Cell Biol. 17, 351–359 (2015).

46. Bikanin, A. V. et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399–405 (2012).

47. Liang, Q. et al. Integrative identification of Epstein–Barr virus-associated mutations and epigenetic alterations in gastric cancer. Gastroenterology 147, 1350–1362 (2014). e1354.

48. Okegawa, T., Pong, R. C., Li, Y. & Hsieh, J. T. The role of cell adhesion molecule in cancer progression and its application in cancer therapy. Acta Biochim. Pol. 51, 445–457 (2004).

49. Martin, T. A., Mason, M. D. & Jiang, W. G. Tight junctions in cancer metastasis. Front. Biosci. 16, 898–936 (2011).

50. Cui, C., Merritt, R., Fu, L. & Pan, Z. Targeting calcium signaling in cancer therapy. Acta Pharm. Sin. B 7, 3–17 (2017).

51. Shen, C., Teoh, N. C., Chitturi, S. & Farrell, G. C. Coffee and non-alcoholic fatty liver disease: breeding evidence for hepatoprotection? J. Gastroenterol. Hepatol. 29, 435–441 (2014).

52. Park, E. J. et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. Cell 140, 197–208 (2010).

53. Yu, J. et al. Inhibitory role of peroxisome proliferator-activated receptor gamma in hepatocarcinogenesis in mice and in vitro. Hepatology 51, 2008–2019 (2010).

54. Wang, J. et al. Inhibitory role of Smad7 in hepatocarcinogenesis in mice and in vitro. J. Pathol. 230, 411–452 (2013).

55. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41, 1313–1321 (2005).

56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

57. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

58. McKenna, A. et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

59. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).

60. Rubio-Perez, C. et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell. 27, 382–396 (2015).

61. Zhang, B., Kirov, S. & Snoddy, J. Webgestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 33, W741–W748 (2005).

62. Wang, J., Duncan, D., Shi, Z. & Zhang, R. WEB-based GEnE SeT Analysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 41, W77–W83 (2013).

63. Kutmon, M. et al. WikiPathways: capturing the full diversity of pathway information. Nucleic Acids Res. 44, D488–D494 (2016).

64. Yu, J. et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. Gut 64, 636–645 (2015).

Acknowledgements

The results shown here are in part derived using data generated through the TGCA Research Network: http://cancergenome.nih.gov/. Initial dietary experiments leading up to the conduct of these studies were planned by Dr Claire Latter, and Vanessa Barn who provided excellent technical support in the animal work. This project was supported by the RGC-GRF Hong Kong (14106415, 14111216, 14183171); Collaborative Research Fund (RCF) (Ref. No. C4401–17GF); Vice-Chancellor’s Discretionary Fund CUHK; CUHK Direct grant for research (4052468); Shenzhen Virtual University Park Support Scheme to CUHK Shenzhen Research Institute; Australian National Health and Medical Research Council (NHMRC) project grants 585411, 1048416, 1044288 and 1120898.

Author contribution

I.Q.L conceived and performed the analyses, summarized all the data and figures, and wrote the manuscript. J.Y. and G.F conceived and supervised the study and revised the manuscript. J.Y., N.T. and G.F. secured funding. N.T., S.P., L.X. E.S.H.C., J.C., E.A., W.G. H. and M.M. performed experiments. X.L. helped bioinformatics analysis of mutational calling. L.D. and J.Y. provided human samples. G.N.I., G.F. and J.I.Y.S. provided expertise and feedback.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06931-6.

Competing interests: The authors declare no competing interests.

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

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.