Accepted Manuscript

Genetic Features of Aflatoxin-associated Hepatocellular Carcinomas

Weilong Zhang, Huan He, Mengya Zang, Qifeng Wu, Hong Zhao, Ling-ling Lu, Peiqing Ma, Hongwei Zheng, Nengjin Wang, Ying Zhang, Siyuan He, Xiaoyan Chen, Zhiyuan Wu, Xiaoyue Wang, Jianqiang Cai, Zhihua Liu, Zongtang Sun, Yi-xin Zeng, Chunfeng Qu, Yuchen Jiao

PII: S0016-5085(17)30333-5
DOI: 10.1053/j.gastro.2017.03.024
Reference: YGAST 61056

To appear in: *Gastroenterology*
Accepted Date: 20 March 2017

Please cite this article as: Zhang W, He H, Zang M, Wu Q, Zhao H, Lu L-l, Ma P, Zheng H, Wang N, Zhang Y, He S, Chen X, Wu Z, Wang X, Cai J, Liu Z, Sun Z, Zeng Y-x, Qu C, Jiao Y, Genetic Features of Aflatoxin-associated Hepatocellular Carcinomas, *Gastroenterology* (2017), doi: 10.1053/j.gastro.2017.03.024.

Genetic Features of Aflatoxin-associated Hepatocellular Carcinomas

Weilong Zhang,[1,2,*] **Huan He**,[2,*] **Mengya Zang**,[2,*] **Qifeng Wu**,[1,*] **Hong Zhao**,[3,*] **Ling-ling Lu**,[4,*] Peiqing Ma,[5] Hongwei Zheng,[4] Nengjin Wang,[4] Ying Zhang,[1] Siyuan He,[1] Xiaoyan Chen,[1] Zhiyuan Wu,[2] Xiaoyue Wang,[6] Jianqiang Cai,[3] Zhihua Liu,[2] Zongtang Sun,[2] Yi-xin Zeng,[1,8,+] Chunfeng Qu,[2,7,+] Yuchen Jiao[1,2,9,+]

[1]National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Collaborative Innovation Center for Cancer Medicine, Beijing, China; [2]State Key Lab of Molecular Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [3]Department of Abdominal Surgical Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [4]Qidong People's Hospital and Qidong Liver Cancer Institute, Qidong, Jiangsu Province, China; [5]Department of Pathology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [6]State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China; [7]Immunology Department, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical Colleges, Beijing, China; and [8]Department of Experimental Research, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong Province, China; [9]Laboratory of Cell and Molecular Biology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical Colleges, Beijing, China

*****Names in bold indicate shared first authorship**.**

Funding

This work was supported by State Key Projects Specialized for Infectious Diseases (2012ZX10002008-001,C.Q.), National Key Basic Research Program of China (973 program no.2015CB553900/2015CB553902, Y.J.; 2013CB910303, C.Q.; 2014CBA02001, J.C.), National Natural Science Foundation Fund (81472559,Y.J.; 30973387, C.Q.; 81350007, J.C.), PUMC Fundamental Research Funds (JK2013A21, JK2014B10, Y.J.; 3332015058, H.H), CAMS Innovation Fund for Medical Sciences（CIFMS）(Y.J. and X.W.), CAMS Initiative for Innovative Medicine (2016-I2M-1-001, Y.J. ; 2016ZX310181, Y.J. and X.W.; 2016-I2M-1-007, C.Q.), The National High Technology Research and Development Program of China (863 program 2015AA020408, J.C.).

Abbreviations

AFB1, aflatoxin B1; AF-HCC, aflatoxin-associated HCC; CBS, CTCF (CCCTC-binding factor)/cohesin-binding sites; HCC, hepatocellular carcinoma; IHC, immunohistochemistry; MANAs, potential mutation-associated neoantigens; SBSs, somatic single-base substitutions;

WES, whole-exome sequencing; WGS, whole-genome sequencing.

Corresponding author contact information

+Address correspondence to:

Chunfeng Qu, MD, PhD, Cancer Hospital, Chinese Academy of Medical Sciences, No 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China. e-mail: quchf@cicams.ac.cn; Phone: +86-10-87783103;

Yi-xin Zeng, MD, PhD, Cancer Hospital, Chinese Academy of Medical Sciences, No 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China. e-mail: zengyx@sysucc.org.cn; Phone: +86-10-87788045;

Yuchen Jiao, MD, PhD, Cancer Hospital, Chinese Academy of Medical Sciences, No 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China. e-mail: jiaoyuchen@cicams.ac.cn; Phone: +86-10-87788045.

Conflicts of interest

The authors disclose no conflicts.

Author Contributions

Study concept and design: Yuchen Jiao, Chunfeng Qu, Yi-xin Zeng;

Acquisition of data: Weilong Zhang, Huan He, Mengya Zang, Hong Zhao, Qifeng Wu, Ling-ling Lu, Hongwei Zheng, Nengjin Wang;

Analysis and interpretation of data: Yuchen Jiao, Chunfeng Qu, Yi-xin Zeng, Weilong Zhang, Huan He, Mengya Zang, Hong Zhao, Siyuan He, Qifeng Wu, Ling-ling Lu;

Drafting of the manuscript: Yuchen Jiao, Chunfeng Qu, Weilong Zhang, Huan He;

Statistical analysis: Weilong Zhang, Huan He, Xiaoyue Wang, Siyuan He;

Obtained funding: Yuchen Jiao, Chunfeng Qu; Administrative: Yuchen Jiao, Chunfeng Qu, Yi-xin Zeng, Jianqiang Cai, Zhihua Liu, Zongtang Sun;

Technical support: Peiqing Ma, Ying Zhang, Xiaoyan Chen, Zhiyuan Wu;

Material support: Ling-ling Lu, Hongwei Zheng, Nengjin Wang, Hong Zhao, Jianqiang Cai;

Study supervision: Yuchen Jiao, Chunfeng Qu.

BACKGROUND & AIMS: Dietary exposure to aflatoxin is an important risk factor for hepatocellular carcinoma (HCC). However, little is known about the genomic features and mutations of aflatoxin-associated HCCs compared with HCCs not associated with aflatoxin exposure. We investigated the genetic features of aflatoxin-associated HCC that can be used to differentiate them from HCCs not associated with this carcinogen.

METHODS: We obtained HCC tumor tissues and matched non-tumor liver tissues from 49 patients, collected from 1990 through 2016, at the Qidong Liver Cancer Hospital Institute in China—a high-risk region for aflatoxin exposure (38.2% of food samples test positive for aflatoxin contamination). Somatic variants were identified using GATK Best Practices Pipeline. We validated part of the mutations from whole-genome sequencing and whole-exome sequencing by Sanger sequencing. We also analyzed genomes of 1072 HCCs, obtained from 5 datasets from China, the United States, France, and Japan. Mutations in 49 aflatoxin-associated HCCs and 1072 HCCs from other regions were analyzed using the Wellcome Trust Sanger Institute mutational signatures framework with non-negative matrix factorization. The mutation landscape and mutational signatures from the aflatoxin-associated HCC and HCC samples from general population were compared. We identified genetic features of aflatoxin-associated HCC, and used these to identify aflatoxin-associated HCCs in datasets from other regions. Tumor samples were analyzed by immunohistochemistry to determine microvessel density and levels of CD34 and CD274 (PDL1).

RESULTS: Aflatoxin-associated HCCs frequently contained C>A transversions, the sequence motif GCN, and strand bias. In addition to previously reported mutations in *TP53*, we found frequent mutations in the adhesion G protein-coupled receptor B1 gene (*ADGRB1*), which were associated with increased capillary density of tumor tissue. Aflatoxin-associated HCC tissues contained high-level potential mutation-associated neoantigens, and many infiltrating lymphocytes and tumors cells that expressed PDL1, compared to HCCs not associated with aflatoxin. Of the HCCs from China, 9.8% contained the aflatoxin-associated genetic features, whereas 0.4%–3.5% of HCCs from other regions contained these genetic features.

Conclusions: We identified specific genetic and mutation features of HCCs associated with aflatoxin exposure, including mutations in *ADGRB1*, compared to HCCs from general populations. We associated these mutations with increased vascularization and expression of PDL1 in HCC tissues. These findings might be used to identify patients with HCC due to aflatoxin exposure, and select therapies.

KEY WORDS: liver cancer, Aspergillus, risk factor, pathogenesis

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common and the second most lethal cancer in the world.[1] Dietary aflatoxin exposure is an important environmental risk factor of HCC development.[2] Aflatoxin, among the most potent naturally occurring human hepatocarcinogens, was classified as a "group 1" human carcinogen by the International Agency for Research on Cancer in 1994.[3] Aflatoxin is produced primarily by the fungi *Aspergillus flavus* and *A. parasiticus*. Millions of people worldwide experience uncontrolled exposure to aflatoxin.[4, 5] Aflatoxin contamination is common in improperly stored food such as maize, peanuts, and tree nuts.[6] In some special cases, aflatoxin exposure clusters in high-risk regions where the combination of humid weather and improper food storage conditions expose the majority of the local population to aflatoxin, significantly increasing the incidence of HCC. Identification and removal of these risk factors has greatly decreased its incidence in young adults, indicating the importance of controlling aflatoxin as a preventative measure.[7]

However, most aflatoxin-associated HCC cases arise mainly as a consequence of individual life style and occur sparsely in the general population. In addition, another aflatoxin exposure source with increasing importance is from improperly processed food, such as illegally recycled cooking oil[8] Some fast food has been produced using such oil, too. Aflatoxin is one of the toxins found in such swill-cooked dirty oil. Such contaminations would expose the general population to aflatoxin and are not detectable by direct observation. In such cases, patients cannot always provide a clear history of aflatoxin exposure, and aflatoxin might be undetectable by the time HCC diagnostics come into play. It is difficult to define and identify such aflatoxin-associated HCC (AF-HCC) cases.

Exposure to exogenous mutagens leads to carcinogenesis and the mutagens leave their special mutational signature as a "fingerprint" of the mutagenic process. The mutagenic potency of aflatoxin B1 (AFB1) has been studied in several species of animals, including rodents and nonhuman primates.[9] The most famous driver mutation in human is the *TP53* R249S hotspot mutation.[10,11] However, the mutagenic signature of this carcinogen at the genome-wide level in human remains largely unknown. It is not known whether the aflatoxin-associated subtype harbors mutations in driver genes that are canonical in common HCC or harbors some novel driver mutations with a special tumorigenic mechanism.

In this study, we performed whole-genome sequencing (WGS) and exome sequencing on 49 AF-HCC samples collected from the classic high-risk area of Qidong in China.[12] We identified and refined the mutational signature of aflatoxin, which is characterized by increased i) C>A transversions; ii) the sequence motif of GCN in C>A mutations; and iii) strand bias. Besides the canonical *TP53* hotspot mutation, we found frequent mutations in *ADGRB1*, a potential novel driver gene, and in the noncoding region of CTCF/CBS. We further analyzed the genomics data from 1072 HCC cases in the general population and identified the aflatoxin-associated mutational signature and landscape in a significant fraction of these HCCs without known aflatoxin exposure, indicating the tumorigenic role of aflatoxin

in the general population. Furthermore, the AF-HCCs from the high-risk region and the general population both harbored high levels of potential mutation-associated neoantigens (MANAs) and had PD-L1 expression in infiltrating lymphocytes and tumor cells, indicating sensitivity to anti-checkpoint therapies.

Materials and Methods

Subjects and tissue samples

We obtained the HCC tumor tissues and their matched normal tissues from 49 individuals collected between1990 and 2016 at Qidong Liver Cancer Hospital Institute. The study protocol was approved by the institutional review board. AFB1 levels in the staple food (maize) sampled from Qidong were assayed by thin layer semiquantitative chromatography and 38.2% of food samples tested positive for aflatoxin contamination (Supplementary Table 1). The individual levels of urine AFM1 were evaluated by high-pressure liquid chromatography (HPLC) after the pooled urine samples in eight months were immunoconcentrated (Supplementary Table 2). Supplementary Table 8 summarizes the general information for the HCC patients.

HCC samples from The Cancer Genome Atlas (TCGA)[13] and International Cancer Genome Consortium (ICGC)[14] were included in Supplementary Tables 3 and 9. This study analyzed five datasets of HCCs in the general population. They came from the general population in China (CN-HCC), USA (US-HCC or TCGA-HCC), France (FR-HCC) and Japan (JP1-HCC and JP2-HCC). All 1072 samples from these five datasets were named TCGA/ICGC-HCC. Samples with the aflatoxin-associated signature (AF-model) identified from a general population were named "**-AF". Those not harboring the aflatoxin signature were named "**-others". For example, CN-AF denotes the HCCs with the aflatoxin signature from the general population in China, and CN-others refers to those without the aflatoxin signature (Supplementary Table 3).

Exome sequencing and data analysis

We sequenced the exome of around 21,000 protein-coding genes in 13 tumor DNA and matched normal DNA (Supplementary Table 5). We constructed genomic libraries and capture the coding sequences with the Agilent SureSelect Human All Exon 50Mb Kit and sequenced the captured libraries with the Illumina HiSeq genome analyzer. More details are in Supplementary Materials and Methods section.

Whole-genome sequencing and data analysis

We extracted genomic DNA with DNeasy Blood & Tissue Kit (cat# 69504, QIAGEN, Germany), following the manufacturer's protocol. Genomic DNA was loaded on a 0.8% agarose gel for quality control. The DNA samples were used to construct a genomic library and sequenced by Hiseq X Ten (Illumina) with 150PE (pair-end sequencing). The average coverage of each base in the genome was 63 (range from 47 to 83) for the tumor and 38

(range from 31 to 43) for the normal samples. The average of Q30 (the quality value with 99.9% accuracy) was 86% (range from 84 to 89%) (Supplementary Table 4). More details are in Supplementary Materials and Methods section.

HBV integration analysis

Human genome hg19 and viral genome (HBV, NC_003977.1) were downloaded from NCBI and included in the reference files when reads were mapped by BWA using the VirusFinder software (version 2.0)[18]. HBV integration analysis follows a four-step procedure: (1) read subtraction, (2) virus detection, (3) virus integration site detection, and (4) viral mutation detection. The confidence of HBV integrations was sorted by supporting reads (pair break point reads and softclip reads).

Mutational signature analysis

The mutational signature of 49 AF-HCC and 1072 TCGA/ICGC-HCC were analyzed using the Wellcome Trust Sanger Institute mutational signatures framework with the algorithm known as non-negative matrix factorization (NMF)[19]. We used a four-step procedure for mutational signature analysis: (1) converting mutation data into 96 mutational classes comprising the 6 mutation types (C>A, C>G, C>T, T>A, T>C, and T>G), 5′ context (C, A, G, T), and 3′ context (C, A, G, T) for all samples; (2) identifying the number of processes operative in 49 AF-HCC and 1072 TCGA/ICGC-HCC samples based on the signature stability and Frobenius reconstruction errors obtained for K = 1 to 15 signatures; (3) deciphering the mutational signatures (Signature A, B, and C) of all samples with the number of processes operative from step 2 using NMF algorithm; (4) comparing Signature A, B, C with previously known signatures using cosine similarity through unsupervised hierarchical clustering of three mutational signatures identified in our series (Signature A, B, and C) and 30 mutational signatures previously identified in a pan-cancer study (Sig 1-30) [20].

Classifier development

The whole-genome sequencing results from 36 AF-HCC cases from the high-risk region and 72 samples randomly selected from a cohort of 260 Japanese liver cancer patients were used to derive the classifier. The sample DO45299 in non-AF group was excluded due to an extremely high mutational burden observed in this sample. The mutational signatures were extracted from the 108 samples using the NMF method, and their exposure intensities in each sample were compared. The extracted Signature A, which is the one most similar to Signature 24 in the COSMIC database, showed the highest intensities in the AF group.

For the classifier, the status of AF exposure was used as the outcome variable, and the extracted Signature A was used as the predictor. We chose a logistic regression model to fit the data, using the R package {glmnet}. We assessed the performance of the classifier using cross-validation in the training set with a 30% sample hold-out procedure. Based on 200 iterations, we calculated the area under the curve (AUC), sensitivity, and specificity.

Determination of capillarization in HCC tumorous tissues

Immunohistological staining of CD34 was used to determine the microvessel density within the tumorous tissues.[21] Details are in Supplementary Materials and Methods section.

Determination of PD-L1 expression in HCC tissues

We used 1:100 diluted rabbit monoclonal anti-human PD-L1 (Cat. No.13684, Cell Signaling, MA, USA) and 1:200 diluted mouse monoclonal anti-human CD45 (Cat. ZM-0183, ZSGB-Bio, Beijing, China) to examine the expression of PD-L1 in tumor cells and in immune cells. Details are in Supplementary Materials and Methods section.

Analysis of mutations in the CBS region

The transcription factor binding sites were defined as genomic regions covered by peaks in the ChIP-seq data from the HepG2 cell line in ENCODE data.[22] The total size of the CTCF and RAD21 defined with ENCODE regions was approximately 28 Mb (about 1% of the whole genome) and close to 15 Mb (about 0.6% of the whole genome), respectively. The overlap between CTCF, RAD21 peaks, and ChIP-exo CBSs regions[23] were about 1.2M (approximately 0.04% of the whole genome).

To test the enrichment for mutations on CBS regions compared to the flanking regions, we compared the ratio of the total number of mutations to the total number of nucleotide positions in the CBS regions (−20 to + 20 nt) and in the flanking region (21 to 1,000 nt on either side, respectively) using a Fisher's exact test. We performed this test and fold change for CBS regions of six classes of mutation type (C>A, C>G, C>T, T>A, T>C and T>G) respectively.

Mutant peptide MHC binding prediction

Somatic frameshift insertions, frameshift deletions, and missense mutations were analyzed for potential MHC class I binding focusing on HLA-A, HLA-B, and HLA-C. First, all mutations in exomes were translated to amino acid alterations using the CCDS database from NCBI, and we identified 9-mer and 10-mer epitopes surrounding each mutation. Then we analyzed the predicted MHC Class I binding strength of each mutant amino acid fragments, using Immune Epitope Database (www.iedb.org). A predicted affinity of less than 500 nm were considered to be potential binders. We repeated that same process for the complementary wild-type peptide for each mutant peptide, to further refine the total neoantigen burden. Mutation-associated neoantigen (MANA) was defined with two criteria: (1) it was a strong potential binder (2) its complementary wild-type peptide was a weak potential binder.

Results

Genome-wide study of AF-HCC cases

We sequenced the genomes of 49 AF-HCC subjects collected in Qidong, a classic high-risk region of aflatoxin exposure in China[12]. The AFB1 level of staple food and the

AFM1 level of urine in Qidong tested far above the normal range before the 1980s (Supplementary Tables 1 and 2). We performed WGS on 36 samples and whole-exome sequencing (WES) on 13 samples and identified an average of 28,408 somatic single-base substitutions (SBSs) per genome or 738 per exome (Figure 1A and B). In both the genome scale and the exome scale, the mutation load was significantly higher than that in HCCs from the general population ($P < .0001$, unpaired t test; Figure 1A and B). The average number of somatic SBSs was 9 per Mb, ranging from 4.7 to 19.0 per Mb (Figure 3A). We compared the nonsynonymous mutation load of AF-HCCs with those in four tumor types that were associated with group 1 carcinogens (Figure 1C).[24-26] UV-exposed melanomas (n = 7), smoking-associated lung cancers (n = 10), and *Helicobacter pylori*-associated gastric cancers (n = 8) harbored 336, 192, and 60 nonsynonymous SBSs per exome, respectively. The aflatoxin (242 nonsynonymous SBSs per exome) showed a strong mutagenic potential — at a level similar to UV or smoking and much stronger than *H. pylori*.

Mutational signature of aflatoxin in HCC

At both the genome and exome levels, the SBSs in AF-HCCs exhibited a mutagenic signature, with a dominant mutation pattern of C>A (equal to G>T) transversions (Figure 1D and Supplementary Figure 1A). To facilitate comparison of our WGS and exome data with that of the HCC samples from the public database, we focused our analysis on the SBSs in the coding regions. The C>A mutations showed a strong strand preference, with G>T mutations on nontranscribed strands occurring 2.12 times as frequently as those on transcribed strands (Figure 1E and Supplementary Figure 1B). In addition, we found a strong preference for a G in the base preceding the mutated C residue for both synonymous and nonsynonymous C>A mutations (Figure 1F; Supplementary Figures 1C and 3A). In HCC cases reported in the TCGA database, C>T (equal to G>A) transitions were the predominant change (28% of SBSs) and C>A transversions accounted for 21% of SBSs (Figure 1D). The strand bias of G>T mutations between nontranscribed and transcribed was 1.36 in the TCGA cases, compared with 2.05 in AF-HCCs (Figure 1E). Our analysis did not identify any significant preference in the context of the mutated C>A (Figure 1F).

We further compared the proportion of the 96 mutational classes between the cases that were positive in the urine AFM1 test and those had not had the test. There is a strong correlation between the two groups not only in GCN, but also in almost all other classes (Supplementary Figure 12, R = 0.9780, $P < 0.0001$, Pearson Correlation).

We extracted the mutational signatures from the AF-HCCs with Wellcome Trust Sanger Institute Framework software[19, 27,28]. We identified three signatures and compared them to the Signatures of Mutational Processes in Human Cancer in the COSMIC database[20] (Figure 1G). Signature A, the dominant signature, showed a strong correlation to Signature 24 with a cosine correlation similarity of 0.94 (Figure 1G). Compared with Signature 24, Signature A showed more mutations in GCN>GAN, and fewer mutations in other classes of alterations (Supplementary Figure 8A). The other two signatures correlated with Signature 22 and Signature 5, with cosine similarities of 0.98 and 0.94, respectively (Supplementary Figure 7). We built a classifier based on Signature A, the most prominent mutational signature extracted

from the mutation data of AF-HCC tumors (see Methods). Tested in the training dataset of 36 AF-exposed HCC and 72 randomly selected cases from a cohort of 260 Japanese liver cancer patients, the classifier had a ROC curve AUC of 0.963 (95% CI, 0.962–0.964).

Mutational signature in individual AF-HCC cases

In addition to the global mutational signature in the AF-HCC cohort, we examined the mutation patterns in each individual case. In 42 of the 49 AF-HCC cases, the majority (more than 40%) of mutations were C>A transversions (Figure 2A). The C>A fraction of AF-HCC was far higher than in the TCGA cohort, which had a median fraction of C>A mutation of 21% and a Standard Deviation (SD) of 6% (Figure 2B). These cases had significant strand preferences and an increased GCN pattern in C>A mutations (Figure 2C and D). In four cases, C>A transversions were not dominant (Figure 2A). However, C>T transitions were not dominant in these samples, either. Instead, A:T>T:A transversions predominated, with a preference for the CTG>CAG pattern (Supplementary Figure 2A) and a significant strand bias of 2.3 (Supplementary Figure 2B). This signature was similar to that associated with aristolochic acid (AA) in prior studies.[2,29,30] We identified two cases that harbored both a C>A signature (characteristic of aflatoxin exposure) and a T>A signature (characteristic of AA-exposure; Figure 2A). The mutations in each signature showed the expected context and strand bias (Supplementary Figure 2A and B).

The distributions of Signatures A, B and C showed a similar pattern. In 43 of the 49 AF-HCC cases, Signature A dominated or contributed to at least one third of the mutations (Figure 1E). In the same four cases, Signature B was dominant (Figure 1E). In two cases, Signature A and B were both significant (Figure 1E).

Contribution of C>A mutational signature to the driver mutations

The AF-HCC samples displayed a characteristic pattern of mutated genes. Among them, *TP53*, *AXIN1*, *TERT*, and *ADGRB1* had the highest mutation frequency in AF-HCC (Figure 3A). Previous studies had reported *TP53* mutations on HCCs from the general population. However, the *TP53* mutation frequency was much higher in AF-HCC (81.6%), and the major genotype was the R249S hotspot mutation (Figure 3A and Supplementary Table 6).[31] Hotspot SBS mutations in the TERT promoter region, as well as HBV integration near the TERT promoter region, were frequent in AF-HCCs, which were similar to those in the general populations in previous studies (Supplementary Figure 10)[32]. Several potential driver genes, previously unidentified in HCC, were found to have mutated in the AF-HCC cases. Nine of the 49 samples harbored mutations in the Adhesion G Protein-Coupled Receptor B1 (*ADGRB1*, also known as brain-specific angiogenesis inhibitors 1, *BAI1*) gene.[33,34] *ADGRB1* had a significantly higher mutation frequency in AF-HCC (18.4%) than in HCCs from the general population (TCGA/ICGC-HCC; 1.5%, n = 1072; $P < .0001$, Fisher's exact test) (Figure 3B and C). We also identified four mutations in *ADGRB2* and two in *ADGRB3*, respectively. All together, 30.6% (15/49) AF-HCCs were mutant in the *ADGRB* family genes (Supplementary Table 7).

We found no histological differences between the architectures and the cytology of the

tumors with *ADGRB* mutations and the wild-type *ADGRB*. Interestingly, more CD34-positive cells presented in the *ADGRB*-mutant tumor tissues than in *ADGRB*-wild-type samples (P = .005, unpaired t test; Figure 3D and E).

The C>A mutational signature was the major contributor to some of the driver mutations in the AF-HCC samples. 33 out of the 41 *TP53* mutations were C>A changes (Figure 3A). In the ICGC database, on the other hand, nine of the top ten SBS hotspot mutations in *TP53* were C>T mutations, and none of them were C>A mutation (Supplementary Figure 3B). In addition, seven of the nine *ADGRB1* mutations in AF-HCC displayed this specific C>A signature. Altogether, 84 (57%) of the 148 potential SBS driver mutations were typical changes of the aflatoxin-associated signature, strongly supporting the hypothesis that the exposure to aflatoxin was the causative event in these cancers.

Contribution of the C>A mutational signature to the mutations in noncoding region

The C>A pattern was also a dominant contributor of the mutations in the noncoding regions of transcription binding sites such as CTCF (CCCTC-binding factor)/cohesin-binding sites (CBS) (Figure 4). By analyzing transcription factor binding motifs identified in chromatin immune-precipitation sequencing (ChIP-seq) data from an HCC cell line,[22] we identified a 2.15-fold enrichment in CTCF/CBS mutations in the AF-HCC samples (Figure 4A; $P = 3.29 \times 10^{-7}$, Fisher's exact test). Interestingly, the C>A mutation was the dominant contributor of the CTCF/CBS mutations (Figure 4C–H). CBS is an important mutation target in the noncoding region of multiple cancer types,[23] and aflatoxin-associated exposure provided a mechanism to explain the accumulation of mutations in CBS in AF-HCC cases. In other transcription factors, such as MAFK (MAF BZIP Transcription Factor K), we did not observe any significant increase in the mutation frequency (Figure 4B; P = .55, Fisher's exact test).

HCC with aflatoxin signature in the general population

We further analyzed the genomes of 1072 HCCs from TCGA and ICGC to find those with an aflatoxin fingerprint. Strikingly, 9.8% of the HCCs in China showed a typical mutational signature of aflatoxin (CN-AF; Figure 5A), which differed significantly from the rest of the HCCs (CN-others; Figure 5C–F). This subgroup, defined by its mutational signature, showed a mutational landscape highly consistent with the AF-HCCs. The distribution of *TP53* genotypes, including the R249S mutation, the non-R249 C>A mutation, the non-C>A mutation, and its wild type, was similar to those from the AF-HCCs and significantly different from the rest of the HCCs (Figure 5B and Supplementary Table 7). In the Chinese population, 4 of the 24 samples with a full or partial aflatoxin signature harbored *ADGRB1* mutations, while only 4 out of the rest of the 139 samples had mutations in this gene (Supplementary Figure 5; P = .02, Fisher's exact test).

In the USA, 3.5% of HCCs harbored the aflatoxin signature; the ratio in France was 1.7% (Figure 5G; Supplementary Figure 4; Supplementary Table 3). The lower incidence did not compromise the typicality of the positive cases. Three of the seven AF-HCC cases (43%) in the US harbored the canonical *TP53* R249S hotspot mutation (Figure 5C), while only 5 of the

other 195 HCC cases harbored the R249S mutation (P = .004, Fisher's exact test). Three of the four cases with the aflatoxin signature in France harbored *TP53* R249S mutations, and the other case had a *ADGRB1* mutation. On the other hand, France had only one *TP53* R249S mutation (P < .0001, Fisher's exact test) and one *ADGRB1* mutation (P < .03, Fisher's exact test) in the rest of the 230 cases. Two independent studies in Japan showed consistently low incidence of AF-HCC (Figure 5G), indicating that the difference was mainly due to environmental factors instead of genetic background.

Based on the mutational signature extracted from all of the AF-HCC samples and the 1072 HCCs in these general populations, we set up a threshold to identify AF-HCC cases in the general population without known aflatoxin exposure history (Supplementary Figure 11). The threshold was: the proportion of Signature 24 was greater than 45%, and the total number of SBSs was greater than 70. With these criteria, we found that the percentage of aflatoxin-associated HCCs in the general populations were 8.6, 2.5, 2.1, and 0% in China, the USA, France, and Japan, respectively.

Increased MANAs and PD-L1 expression in AF-HCCs

As AF-HCCs harbored a high mutation load, we evaluated the levels of MANAs in these tumors. The AF-HCCs from the high-risk region and CN-AFs from the general population both harbored significantly more MANAs than their counterparts (Figure 6A and B). The AF-HCC samples showed an increased expression of PD-L1 in both infiltrated immune cells (CD45-positive) and tumor cells (Figure 6C). The percentage of PD-L1-positive cells in CD45-positive cells in AF-HCC was significantly higher than that in HCCs from the general population (GP-HCC; Figure 6D).

Discussion

Exposure to exogenous mutagens leads to carcinogenesis, and the mutagens leave their special mutational signature as their unique identifier of the mutagenic process.[19] Some hotspot mutations in frequently mutated driver genes like *TP53* have been associated with exposure to individual mutagens. However, it remains difficult to determine the association of a cancer case to a mutagen based simply on the genotype of a hotspot or a driver gene. Genome-wide sequencing of tumors identifies many mutations and provides the statistical power to define the detailed characteristics of the mutational signature precisely.[30] The precise signature makes it possible to detect the involvement of a mutagen in sporadic cancers not previously known to be associated with the mutagen.[29,30]

In our study, we sequenced the genomes of 49 well-defined AF-HCC samples from a high-risk region where the aflatoxin exposure had been widely validated in staple foods, blood and urine tests, and public health studies.[6,35] Some of the samples had clear evidence of aflatoxin exposure in the urine test. The rest of the samples did not have available blood or urine results, but their genomes showed highly similar mutational signatures to those with

such evidence (Figures 1G and 2A-E). The similarity was not only in the major mutational classes of the signature, but also in almost all the 96 classes. This genetic evidence supports the former studies that the HCC cases in our cohort are highly consistent in aflatoxin exposure. This history of aflatoxin exposure is very difficult to validate in general population, which is different from other risk factors like drinking or smoking. To our knowledge, this is the first genome-wide study on HCC cases in a cohort with clear evidence of aflatoxin exposure.

The dominant signature is similar to Signature 24, which earlier studies have indicated as being associated with aflatoxin exposure[36]. Taken together, our study clearly identifies Signature A (or Signature 24) as the mutational signature of aflatoxin exposure. Compared with Signature 24, signature A showed more preference in GCN>GAN mutations, and harbored less mutations in other classes of alterations (Supplementary Figure 8). In this case, our data from AF-HCCs in a high-risk region could refine the signature with more AF-HCC cases and more dominance of aflatoxin exposure in the process of carcinogenesis. Studies on more cohorts of AF-HCCs in other regions at high risk for aflatoxin exposure could further validate the signature in highly intensive exposure of aflatoxin. To our knowledge, this study is the first to validate the genome-wide aflatoxin signature for human cancers in cases with a clinical history of aflatoxin exposure.[19, 36]

We obtained the precise signature because of the huge number of mutations from WGS, and we identified the HCCs with the same signature from the general population. The HCCs selected simply by the mutational signature from the general population was harboring the same mutational landscape as the AF-HCCs in high-risk region, but different from the rest HCCs in the general population that were not harboring the aflatoxin signature. Furthermore, we evaluated the incidence of AF-HCCs in the general populations from four different countries all over the world. By direct observation of the characteristics of SBS mutations (percentage of C>A mutations among all mutations, of GCN>GAN mutations among C>A mutations, and the strand bias), we found that the incidence of aflatoxin-associated HCCs in all HCCs in the general populations was 9.8, 3.5, 1.7, and 0.4% in China, the USA, France, and Japan, respectively. Analysis based on the fraction of mutational signatures showed a similar result with 8.6, 2.5, 2.1, and 0% of AF-HCCs in the general population in China, the USA, France, and Japan, respectively.

The dramatic difference between the two Asian cohorts (China and Japan) strongly suggests that the different distribution among countries was mainly a consequence of environmental factors rather than genetic background. Furthermore, five out of the seven AF-HCC cases identified in the USA had Asian or African ancestry. They may have been exposed to aflatoxin in their country of origin.

The distribution of AF-HCC in the general population as calculated by our genomics method was consistent with the figures from epidemiology studies.[35,37] In light of the genome-wide nature of these studies and the consistency to epidemiological studies, there is little doubt that exposure to aflatoxin or a mutagen with similar effects to aflatoxin played a pathogenic role in the development of these cancers.

Besides the aflatoxin signature, we also identified a T>A dominated signature. This Signature B, or Signature 22, had been identified in cancers associated with AA exposure. The AA-rich herb, Caulis aristolochiae manshuriensis, was widely used in Chinese traditional medicine regimens for a variety of symptoms, including those associated with liver diseases — until 2002, when it was largely replaced with Caulis akebiae without AA. These HCC cases might be associated with AA taken before 2002. As the genome-wide study provided a quantitative identification of aflatoxin-associated cases, we can also classify the cases that harbored a partial aflatoxin signature in high-risk or general populations. In the AF-HCCs, we identified two cases harboring both aflatoxin- and AA- associated signatures, and the C>A or T>A mutations showed the expected context and strand bias of each signature. The proportion of Signature A and Signature B were greater than 25% in each case (Figures 1G and 2E). Aflatoxin exposure may have played a partial role in the carcinogenesis of these cases, and the aflatoxin signature was mixed with some other signatures. Genome-wide studies provide a precise way to identify such mixed signatures.

Our study also identified the characteristic mutation landscape of AF-HCC. Frequently mutated driver genes include previously implicated genes like *TP53*, *AXIN1* and *TERT*. There are also several genes that have not been reported in HCCs in previous studies. Among them, *ADGRB1* is the most frequently mutated in the aflatoxin- associated cohort. *ADGRB1* belongs to an orphan G-protein-coupled receptor (GPCR) subfamily comprising three brain-specific angiogenesis inhibitory genes.[38] The other two members of this subfamily, *ADGRB2* and *ADGRB3*, are also recurrently mutated in AF-HCC samples. A systemic analysis of multiple tumor types has identified *ADGRB3* as a driver gene in lung cancer.[33] Although lung cancer and some other tumor types have a few *ADGRB1* and *ADGRB2* mutations, the mutation frequency was low and it was hard to determine whether these mutations were driver mutations or just random.[33] In this study, we identified frequent mutations of *ADGRB1* in AF-HCC. We also identified frequent *ADGRB1* mutations in the HCCs from the general population harboring the aflatoxin-associated signature.

Previous studies have shown that the expression of ADGRB1 occurs especially in brain tissues and that it inhibits experimental angiogenesis and tumor formation.[39,40] The thrombospondin type 1 repeats (TSR) domain of ADGRB1 binds to CD36 receptors of microvascular endothelial cell membranes, which then leads to endothelial cell apoptosis.[41] Interestingly, the GSE1133 data of expression profiling by microarray (Gene Expression Omnibus database) shows that *ADGRB1* has a similarly high level of expression in the liver as in the brain (Supplementary Figure 9)[42]. In our study, CD34-staining for neovascularization showed significantly higher capillarization in the *ADGRB* mutated HCC tumorous tissues than in those with the wild type *ADGRB* genes (Figure 3D and E). These data suggest that ADGRB1 could play a role similar to the one it plays in brain tissues and that *ADGRB1* alterations could contribute to the tumorigenesis of AF-HCC by supporting angiogenesis. Our study identified frequent *ADGRB1* mutations and its association with angiogenesis, indicating that *ADGRB1* could be a novel driver gene. The function and underlining mechanism of *ADGRB* mutations in HCC need to be further studied for potential targets to improve HCC therapy.[43]

In recent studies, the nonsynonymous mutation burden and MANA load were associated with increased PD-L1 expression levels and were strong predictive markers of anti-PD1 therapy in melanoma, lung cancer, and other tumor types.[13,44,45] The mutation load of AF-HCC was comparable to that in UV-associated melanoma and smoking-associated lung cancer (Figure 1C). The AF-HCCs from high-risk and general populations both harbored significantly more MANAs. The PD-L1 expression was higher in the lymphocytes and tumor cells of the AF-HCCs. Taken together; our results indicate that an immune checkpoint inhibitor could potentially offer an efficient therapy for AF-HCCs from the high-risk region and the general population.

Our study adds support to the idea that genomic sequencing can inform epidemiologic studies. Since China has one of the high-incidence populations of HCC, this study points out the importance of managing HCC by controlling the exposure of the general population to environmental carcinogens. In addition, it demonstrates that a genome-wide study of individual HCC cases from the general population can precisely identify the HCCs associated with aflatoxin exposure — that is, those which could benefit from anti-checkpoint therapy.

Acknowledgements

We thank Drs. B. Vogelstein, K.W. Kinzler, N. Papadopoulos, S. Zhou, and M. Hoang for their helpful discussions. We thank Dr. Taoyang Chen in Qidong Liver Cancer Institute for organizing the sample collection.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Poon D, Anderson BO, Chen LT, et al. Management of hepatocellular carcinoma in Asia: consensus statement from the Asian Oncology Summit 2009. Lancet Oncol 2009;10:1111-8.
3. Lopez C, Ramos L, Bulacio L, et al. Aflatoxin B1 content in patients with hepatic diseases. Medicina (B Aires) 2002;62:313-6.
4. Williams JH, Phillips TD, Jolly PE, et al. Human aflatoxicosis in developing countries: a review of toxicology, exposure, potential health consequences, and interventions. Am J Clin Nutr 2004;80:1106-22.
5. Kensler TW, Qian GS, Chen JG, et al. Translational strategies for cancer prevention in liver. Nat Rev Cancer 2003;3:321-9.
6. Kensler TW, Roebuck BD, Wogan GN, et al. Aflatoxin: a 50-year odyssey of mechanistic and translational toxicology. Toxicol Sci 2011;120 Suppl 1:S28-48.
7. Sun Z, Chen T, Thorgeirsson SS, et al. Dramatic reduction of liver cancer incidence in young adults: 28 year follow-up of etiological interventions in an endemic area of China. Carcinogenesis 2013;34:1800-5.
8. Likun Z. Million tons of swill-cooked oil back on table. http://www.chinadaily.com.cn/china/2010-03/18/content_9610919.htm, 2010.
9. Meier B, Cooke SL, Weiss J, et al. C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. Genome Res 2014;24:1624-36.
10. Hsu IC, Metcalf RA, Sun T, et al. Mutational hotspot in the p53 gene in human hepatocellular

carcinomas. Nature 1991;350:427-8.
11. Bressac B, Kew M, Wands J, et al. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature 1991;350:429-31.
12. Ming L, Thorgeirsson SS, Gail MH, et al. Dominant role of hepatitis B virus and cofactor role of aflatoxin in hepatocarcinogenesis in Qidong, China. Hepatology 2002;36:1214-1220.
13. Chin L, Meyerson M, Aldape K, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061-1068.
14. Hudson TJ, Anderson W, Aretz A, et al. International network of cancer genome projects. Nature 2010;464:993-998.
15. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-595.
16. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology 2013;31:213-219.
17. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568-76.
18. Wang Q, Jia P, Zhao Z. VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. PLoS One 2013;8:e64465.
19. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3:246-59.
20. Alexandrov LB. Signatures of mutational processes in human cancer. Nature 2013;500:415-421.
21. Xue Y, Lim S, Brakenhielm E, et al. Adipose angiogenesis: quantitative methods to study microvessel growth, regression and remodeling in vivo. Nat Protoc 2010;5:912-20.
22. Hoffman MM, Ernst J, Wilder SP, et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Research 2013;41:827-841.
23. Katainen R, Dave K, Pitkanen E, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet 2015;47:818-21.
24. Nikolaev SI, Rimoldi D, Iseli C, et al. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. Nat Genet 2012;44:133-9.
25. Govindan R, Ding L, Griffith M, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 2012;150:1121-34.
26. Zang ZJ, Cutcutache I, Poon SL, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet 2012;44:570-4.
27. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463:191-6.
28. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
29. Hoang ML, Chen CH, Sidorenko VS, et al. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. Sci Transl Med 2013;5:197ra102.
30. Poon SL, Pang S-T, McPherson JR, et al. Genome-Wide Mutational Signatures of Aristolochic Acid and Its Application as a Screening Tool. 2013;5:197ra101-197ra101.
31. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of

somatic mutations in human cancer. Nucleic Acids Research 2015;43:D805-D811.
32. Killela PJ. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc. Natl Acad. Sci. USA 2013;110:6021-6026.
33. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 2010;466:869-73.
34. Cork SM, Van Meir EG. Emerging roles for the BAI1 protein family in the regulation of phagocytosis, synaptogenesis, neurovasculature, and tumor development. J Mol Med (Berl) 2011;89:743-52.
35. Liu Y, Chang CC, Marsh GM, et al. Population attributable risk of aflatoxin-related liver cancer: systematic review and meta-analysis. Eur J Cancer 2012;48:2125-36.
36. Schulze K, Imbeaud S, Letouze E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505-11.
37. Liu Y, Wu F. Global burden of aflatoxin-induced hepatocellular carcinoma: a risk assessment. Environ Health Perspect 2010;118:818-24.
38. Stephenson JR, Purcell RH, Hall RA. The BAI subfamily of adhesion GPCRs: synaptic regulation and beyond. Trends Pharmacol Sci 2014;35:208-15.
39. Nishimori H, Shiratsuchi T, Urano T, et al. A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. Oncogene 1997;15:2145-50.
40. Kaur B, Cork SM, Sandberg EM, et al. Vasculostatin inhibits intracranial glioma growth and negatively regulates in vivo angiogenesis through a CD36-dependent mechanism. Cancer Res 2009;69:1212-20.
41. de Fraipont F, Nicholson AC, Feige JJ, et al. Thrombospondins and tumor angiogenesis. Trends in Molecular Medicine 2001;7:401-407.
42. Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004;101:6062-7.
43. Teschke R. Traditional Chinese Medicine Induced Liver Injury. J Clin Transl Hepatol 2014;2:80-94.
44. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124-8.
45. Alsuliman A, Colak D, Al-Harazi O, et al. Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: Significance in claudin-low breast cancer cells. Molecular Cancer 2015;14.

Figure Legends

Figure 1. Mutational signature of aflatoxin in AF-HCC samples. (*A*) Mutation counts in the whole genome of AF-HCCs and HCCs from the general population in Japan (JP2-HCC). (*B*) Mutation counts in the exome of AF-HCCs and HCCs from the general population in TCGA database (TCGA-HCC). (*C*) The nonsynonymous mutation load in four tumor types associated with exposure to Group 1 human carcinogens. (*D*) Percentage of the six possible

mutation classes in the exome of AF-HCC and TCGA-HCC. (*E*) The proportions of the six possible SBSs found on transcribed strand (TS, left side) and nontranscribed strand (NTS, right side) in AF-HCC and TCGA-HCC. (*F*) Sequence contexts of C>A mutations of AF-HCC and TCGA-HCC. The height of each base indicates the probability that the base appears in the position around the mutated cytosine in the middle. *(G)* Patterns of three signatures (Signature A, B, and C) observed in 49 AF-HCC genomes using the Wellcome Trust Sanger Institute mutational signatures framework.

Figure 2. Mutational signature of individual AF-HCC cases. (*A*) Distribution of the six mutation classes in the 49 individual AF-HCC cases. (*B*) Scatter plot of the percentage of C>A mutations in all SBS mutations in the exome of AF-HCC and TCGA-HCC. (*C*) Scatter plot of the percentage of C>A mutations with GCN contexts in all C>A mutations in coding regions. (*D*) Scatter plot of the strand bias of C>A mutations in the coding regions of the AF-HCC and TCGA-HCC samples. This paper expressed the data as mean+SEM in a scatter plot. *(E)* Proportion of signature observed in each AF-HCC samples.

Figure 3. Recurrently mutated genes in AF-HCC. (*A*) The mutational spectrum of AF-HCC. (*B*) Diagrams of the mutations in the *ADGRB1* gene of 10 AF-HCC samples and 3 samples with the *ADGRB1* gene mutations from TCGA/ICGC-HCC samples with aflatoxin-associated mutational signature (black trilateral, missense mutation; red trilateral, stop gain mutation). (*C*) The mutation frequency of *ADGRB1* in AF-HCCs and in HCCs from the general population. (*D*) Representative pictures of CD34 staining in the HCC tissues with (left, *ADGRB*-Mut) or without (right, *ADGRB*-WT) *ADGRB* (*ADGRB1*, *ADGRB2*, or *ADGRB3*) mutation. Scale bar = 500 μm. (*E*) Quantification of vessel density in the microscopic fields (n = 9 for *ADGRB* mutant and n = 21 for *ADGRB* wild type) revealed by CD34, and data represent mean±SEM determinants.

Figure 4. Mutation rate in transcription factor binding sites. (*A–B*) Mutation rate of SBSs in CBS and MAFF transcription factor binding sites and their flanking regions. The relative positions of the binding sites are labeled on the X axis. (*C–H*) Mutation rate of SBSs in the six mutation classes in the CBS region and its flanking region.

Figure 5. Aflatoxin-associated HCC cases in the general population. (*A*) The scatter plots of the individual mutational pattern in HCCs in the Chinese population (CN-HCC). The red spots represent the AF-HCC cases from the high-risk region. The green spots (CN-AF) cluster with AF-HCC cases (red) and keep apart from the rest of the CN-HCC cases (blue). (*B*) The distribution of *TP53* mutations in HCCs with the aflatoxin signature from the high-risk region (AF-HCC) and the general population (CN-AF, US/FR-AF) and those in HCCs from the general population without the aflatoxin signature (CN-Others and US/FR-Others). (*C–E*) The percentage, context, and strand bias of C>A mutations in the 16 CN-AF cases clustered with AF-HCC and the rest of the CN-HCC cases. (*F*) Mutation count in the exome of 16 CN-AF cases clustered with AF-HCC and the rest of the CN-HCC cases. (*G*) Aflatoxin-associated HCC cases identified from 1072 HCCs of the general population without known aflatoxin exposure.

Figure 6. Mutation-associated neoantigens (MANAs) and immunohistochemistry of PD-L1 expression in AF-HCC. (*A*) MANA of the CN-AFs and CN-Others. (*B*) MANA of 49 AF-HCC and 50 TCGA-HCC cases. (*C*) Representative images of PD-L1 (brown dots) and CD45 (red dots) staining of aflatoxin-associated HCCs from Qidong (AF-HCC, ID: HQ27, top) and the HCCs from the general population (GP-HCC, ID: GP363544, bottom). The yellow dashed line separates tumor (T) and stromal (S) tissues. Scale bar, 100 μm. (*D*) Based on the numbers of CD45-positive cells in 3 independent fields under 20× magnification, the average percentage of PD-L1-positive cells in each of the HCC cases. Each dot represents one case.

Fig.1

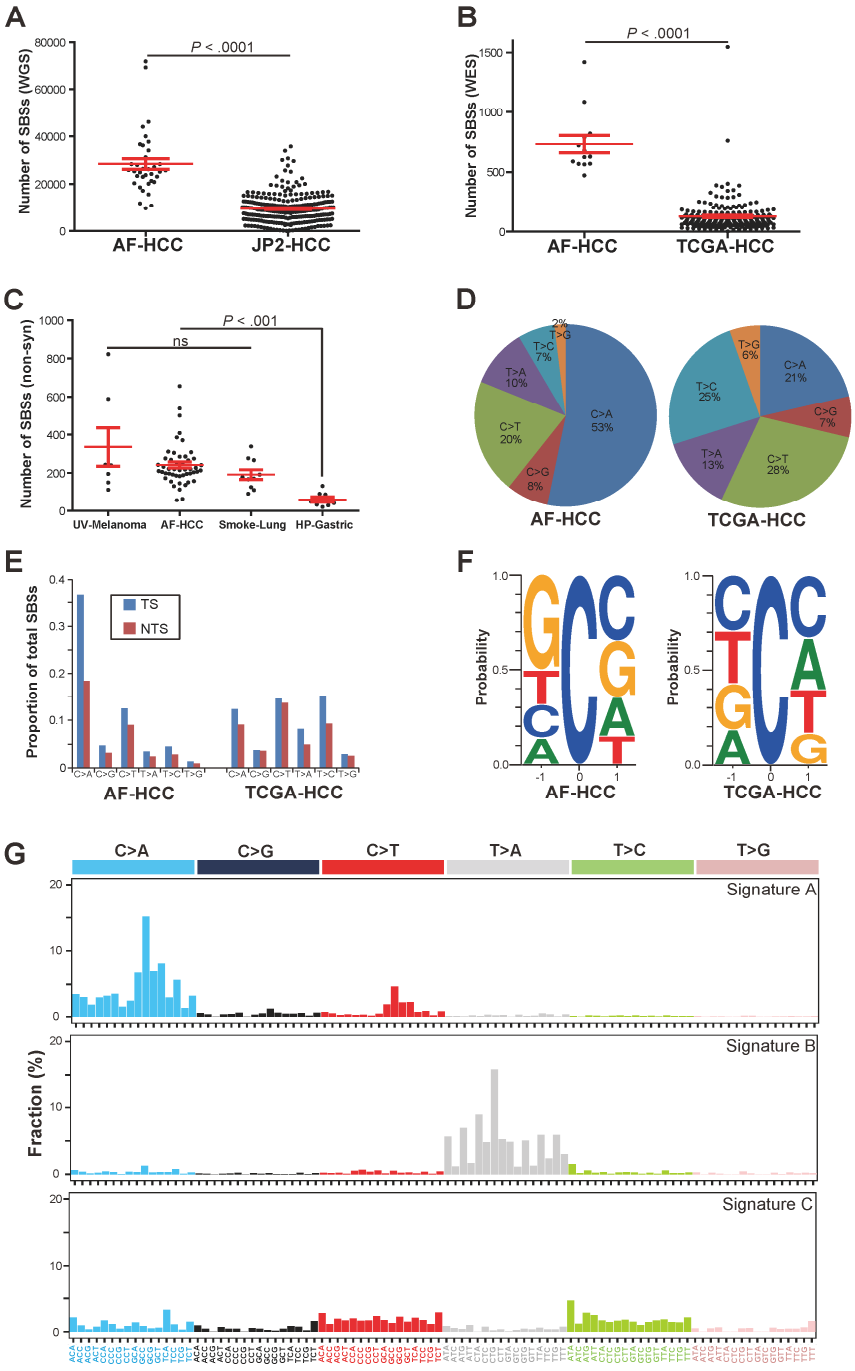

A
Number of SBSs (WGS)
P < .0001
AF-HCC
JP2-HCC
B
Number of SBSs (WES)
P < .0001
AF-HCC
TCGA-HCC
C
Number of SBSs (non-syn)
P < .001
ns
UV-Melanoma
AF-HCC
Smoke-Lung
HP-Gastric
D
C>A 53%
C>G 8%
C>T 20%
T>A 10%
T>C 7%
T>G 2%
C>A 21%
C>G 7%
C>T 28%
T>A 13%
T>C 25%
T>G 6%
AF-HCC
TCGA-HCC
E
Proportion of total SBSs
TS
NTS
AF-HCC
TCGA-HCC
F
Probability
AF-HCC
TCGA-HCC
G
C>A
C>G
C>T
T>A
T>C
T>G
Signature A
Signature B
Signature C
Fraction (%)

Fig.2

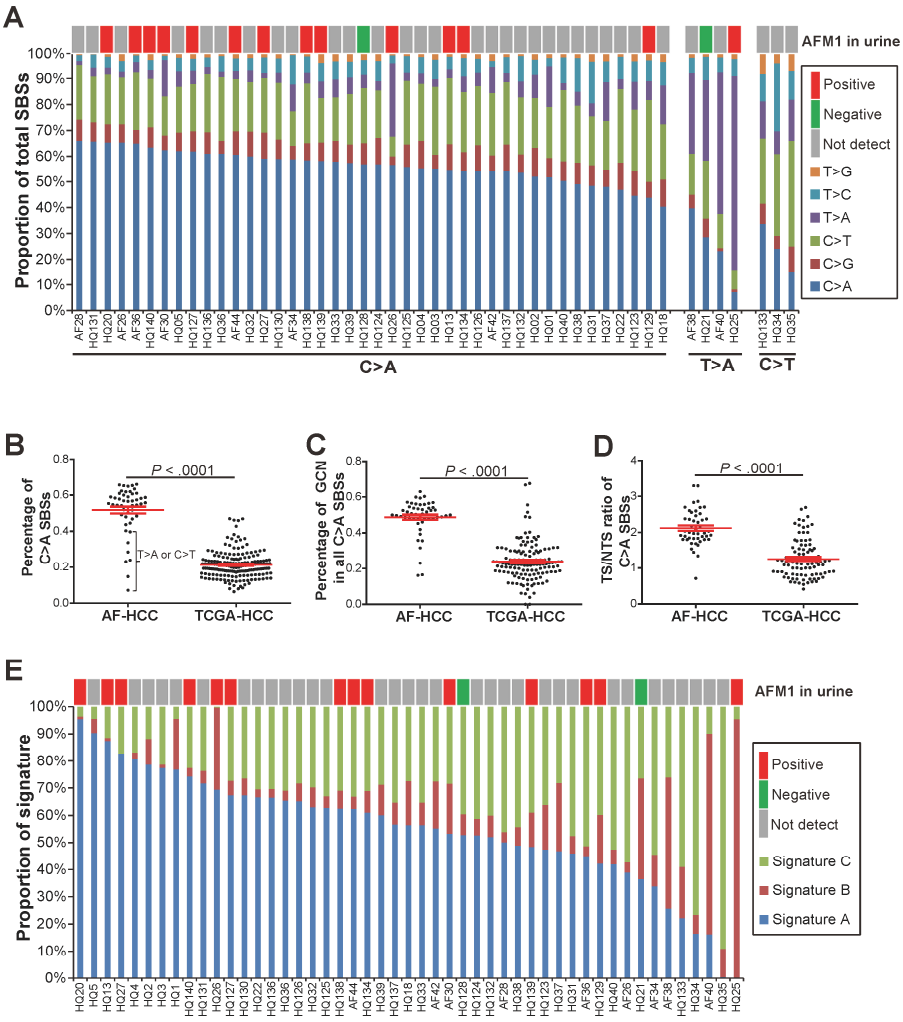

A
AFM1 in urine
Positive
Negative
Not detect
T>G
T>C
T>A
C>T
C>G
C>A
Proportion of total SBSs
C>A
T>A
C>T
B
Percentage of C>A SBSs
P < .0001
T>A or C>T
AF-HCC
TCGA-HCC
C
Percentage of GCN in all C>A SBSs
P < .0001
AF-HCC
TCGA-HCC
D
TS/NTS ratio of C>A SBSs
P < .0001
AF-HCC
TCGA-HCC
E
AFM1 in urine
Positive
Negative
Not detect
Signature C
Signature B
Signature A
Proportion of signature

Fig.3

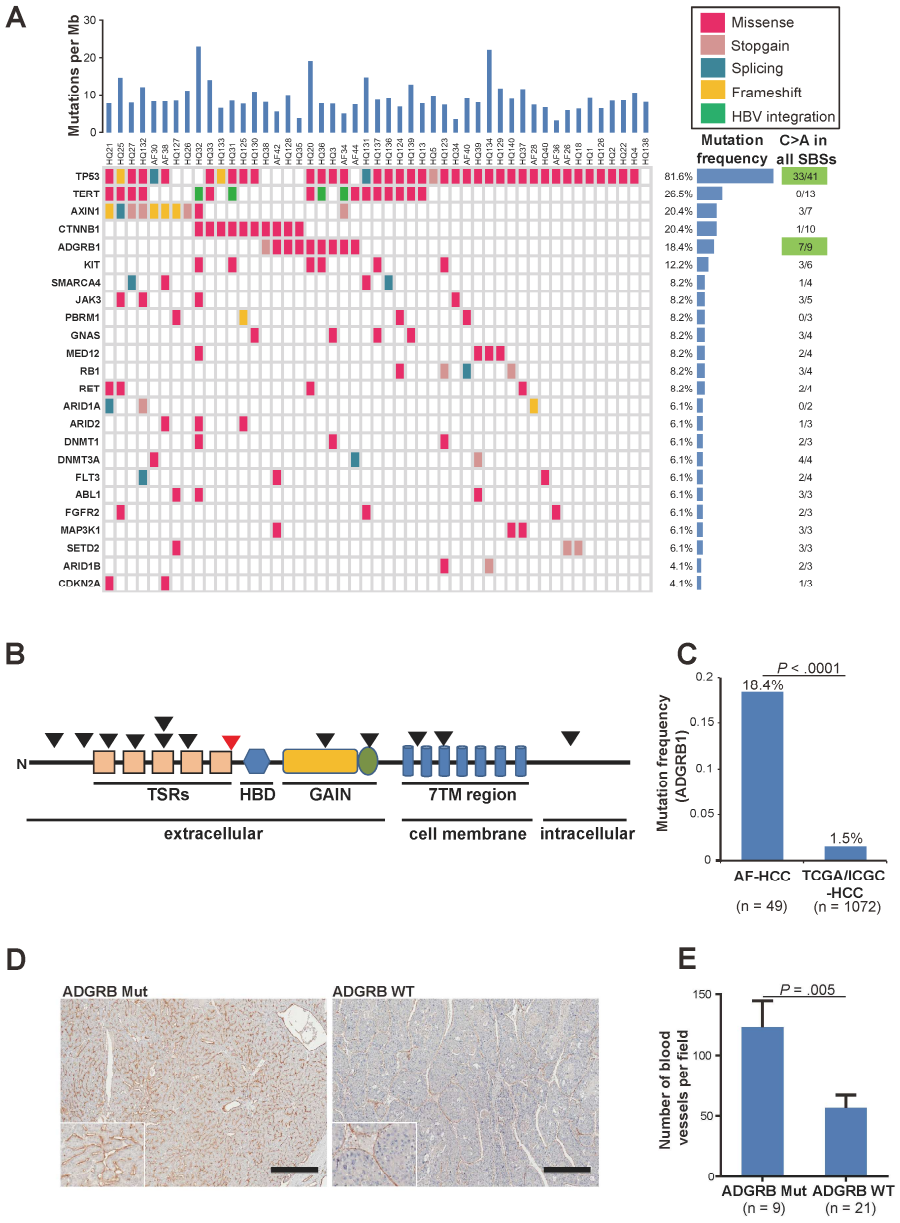

A
Mutations per Mb
Missense
Stopgain
Splicing
Frameshift
HBV integration
Mutation frequency
C>A in all SBSs
TP53
TERT
AXIN1
CTNNB1
ADGRB1
KIT
SMARCA4
JAK3
PBRM1
GNAS
MED12
RB1
RET
ARID1A
ARID2
DNMT1
DNMT3A
FLT3
ABL1
FGFR2
MAP3K1
SETD2
ARID1B
CDKN2A
B
N
TSRs
HBD
GAIN
7TM region
extracellular
cell membrane
intracellular
C
P < .0001
18.4%
1.5%
Mutation frequency (ADGRB1)
AF-HCC (n = 49)
TCGA/ICGC -HCC (n = 1072)
D
ADGRB Mut
ADGRB WT
E
P = .005
Number of blood vessels per field
ADGRB Mut (n = 9)
ADGRB WT (n = 21)

Fig.4

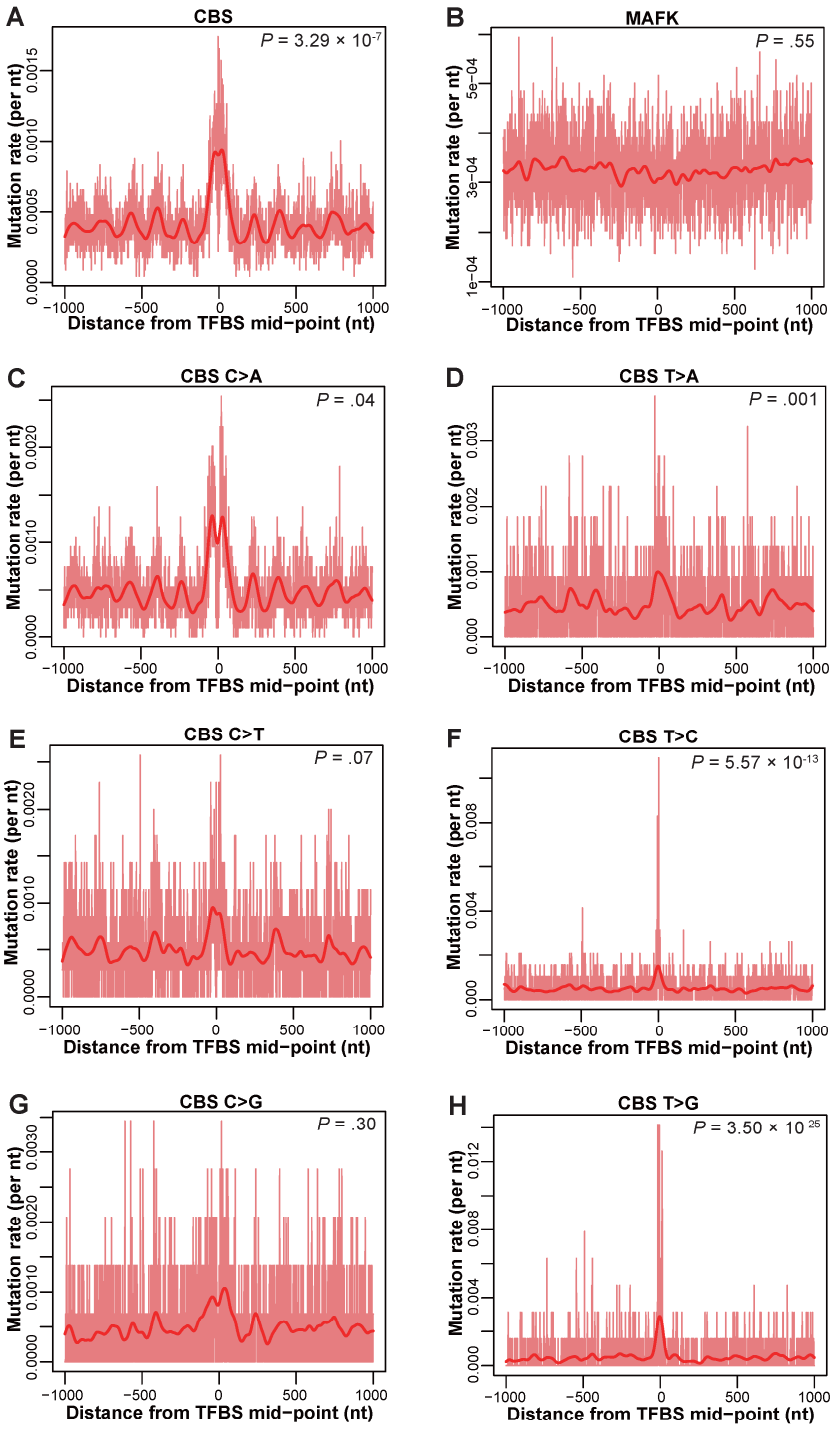

A
CBS
P = 3.29 × 10-7
B
MAFK
P = .55
C
CBS C>A
P = .04
D
CBS T>A
P = .001
E
CBS C>T
P = .07
F
CBS T>C
P = 5.57 × 10-13
G
CBS C>G
P = .30
H
CBS T>G
P = 3.50 × 10 25
Mutation rate (per nt)
Distance from TFBS mid−point (nt)

Fig.5

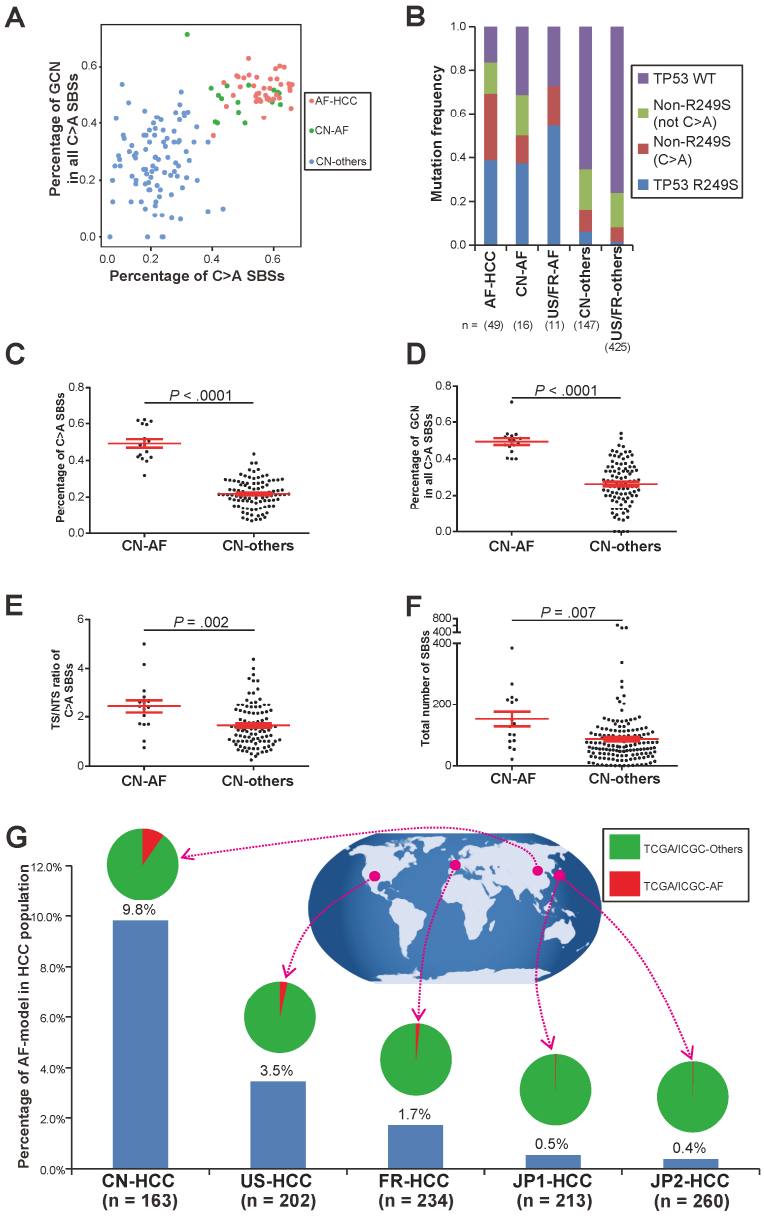
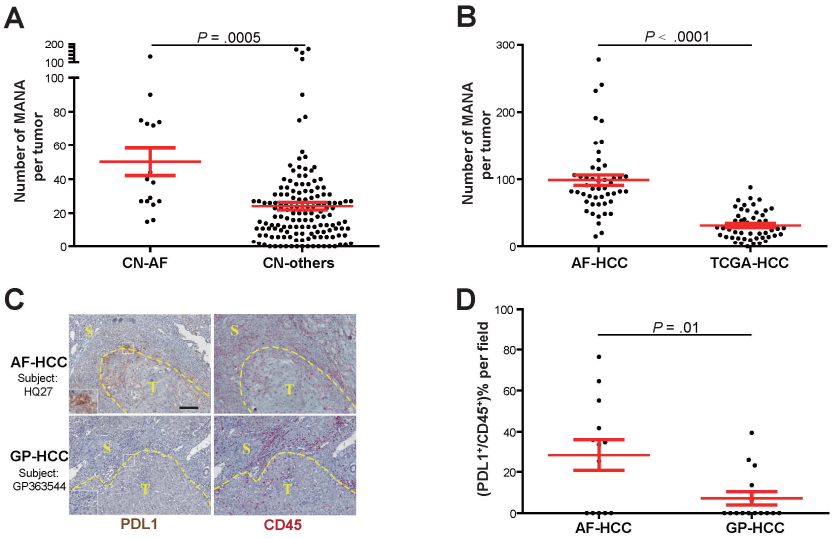

A
Percentage of GCN in all C>A SBSs
Percentage of C>A SBSs
AF-HCC
CN-AF
CN-others
B
Mutation frequency
TP53 WT
Non-R249S (not C>A)
Non-R249S (C>A)
TP53 R249S
AF-HCC
CN-AF
US/FR-AF
CN-others
US/FR-others
n = (49) (16) (11) (147) (425)
C
P < .0001
Percentage of C>A SBSs
CN-AF
CN-others
D
P < .0001
Percentage of GCN in all C>A SBSs
CN-AF
CN-others
E
P = .002
TS/NTS ratio of C>A SBSs
CN-AF
CN-others
F
P = .007
Total number of SBSs
CN-AF
CN-others
G
TCGA/ICGC-Others
TCGA/ICGC-AF
Percentage of AF-model in HCC population
9.8%
3.5%
1.7%
0.5%
0.4%
CN-HCC (n = 163)
US-HCC (n = 202)
FR-HCC (n = 234)
JP1-HCC (n = 213)
JP2-HCC (n = 260)

Fig.6


A
P = .0005
Number of MANA per tumor
CN-AF
CN-others
B
P < .0001
Number of MANA per tumor
AF-HCC
TCGA-HCC
C
AF-HCC
Subject: HQ27
GP-HCC
Subject: GP363544
PDL1
CD45
D
P = .01
(PDL1+/CD45+)% per field
AF-HCC
GP-HCC