Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia

Alvin W. T. Ng,1,2,3* Song Ling Poon,4* Mi Ni Huang,1,2 Jing Quan Lim,4,5 Arnoud Boot,1,2 Willie Yu,1,2 Yuka Suzuki,1,2 Saranya Thangaraju,4 Cedric C. Y. Ng,4 Patrick Tan,2,6,7,8 See-Tong Pang,9 Hao-Yi Huang,10 Ming-Chin Yu,11 Po-Huang Lee,12 Sen-Yung Hsieh,10† Alex Y. Chang,13† Bin T. Teh,2,4,7,14† Steven G. Rozen1,2,3,†

Many traditional pharmacopeias include Aristolochia and related plants, which contain nephrotoxins and mutagens in the form of aristolochic acids and similar compounds (collectively, AA). AA is implicated in multiple cancer types, sometimes with very high mutational burdens, especially in upper tract urothelial cancers (UTUCs). AA-associated kidney failure and UTUCs are prevalent in Taiwan, but AA’s role in hepatocellular carcinomas (HCCs) there remains unexplored. Therefore, we sequenced the whole exomes of 98 HCCs from two hospitals in Taiwan and found that 78% showed the distinctive mutational signature of AA exposure, accounting for most of the nonsilent mutations in known cancer driver genes. We then searched for the AA signature in 1400 HCCs from diverse geographic regions. Consistent with exposure through known herbal medicines, 47% of Chinese HCCs showed the signature, albeit with lower mutation loads than in Taiwan. In addition, 29% of HCCs from Southeast Asia showed the signature. The AA signature was also detected in 13 and 2.7% of HCCs from Korea and Japan as well as in 4.8 and 1.7% of HCCs from North America and Europe, respectively, excluding one U.S. hospital where 22% of 87 “Asian” HCCs had the signature. Thus, AA exposure is geographically widespread. Asia, especially Taiwan, appears to be much more extensively affected, which is consistent with other evidence of patterns of AA exposure. We propose that additional measures aimed at primary prevention through avoidance of AA exposure and investigation of possible approaches to secondary prevention are warranted.

INTRODUCTION

Mutational signature analysis provides a molecular epidemiological tool for detecting environmental exposures that cause cancers (1–5). This has important implications for public health by providing evidence to substantiate causal links between exposures and tumors, providing opportunities for primary and secondary prevention. Mutational signature analysis may also affect clinical oncology in situations where identifiable mutagenic exposures suggest specific cancer risks or preferred treatments.

Mutational signature analysis has been particularly helpful in illuminating the epidemiology of tumors associated with aristolochic acids and their derivatives (collectively, AA). Among these compounds, the in vitro toxicity and mutagenicity of aristolochic acids and aristolactams have been most intensively studied (6–8). AAs include potent mutagens and nephrotoxins present in plants in the genera Aristolochia and Asarum, as well as related plants (6, 7). Many of these plants are used as herbal medicines (9–18). AA mutagenesis is thought to stem from the formation of bulky adducts on purines (19–21). For reasons that are imperfectly understood, but possibly related to better repair of AA-guanine adducts, more accurate translesion synthesis across AA-guanine adducts, or both, AA induces adenine-to-thymine (A>T) mutations almost exclusively (9, 20, 22–24).

In the early 1990s, inadvertent treatment with AA-containing herbs at a Belgian weight loss clinic caused kidney failure in ∼100 women (25), many of whom later developed bladder and upper tract urothelial carcinomas (UTUCs) (9). Subsequently, additional reports of kidney failure and urothelial cancers due to AA poisoning appeared, and it emerged that AA was also responsible for Balkan endemic nephropathy (9). Taiwan also emerged as a hot spot for AA exposure based on prescription records and high rates of kidney failure and UTUCs, which are likely to be partly due to AA exposure (26–30). More recently, mutational signature analysis and other lines of evidence have suggested that AA mutagenesis may be widespread in terms of both geography and types of cancer affected (10). In particular, after we and others described a distinctive mutational signature of AA exposure in the genomes of UTUCs from Taiwan (29, 30), this signature was also found in bladder carcinomas (BCs) from Taiwan and other regions (31), renal cell carcinomas (RCCs) from Taiwan and the Balkans (32–34), intrahepatic bile duct carcinomas from China (35), bile duct carcinomas from Singapore (36), and hepatocellular carcinomas (HCCs) from China, Vietnam, and other Southeast Asian countries (29, 37–39).

Inference of high rates of AA exposure in Taiwan is based on the following evidence: (i) prescription records indicating about one-third of the population exposed to AA (28), (ii) high rates of UTUCs and co-association of kidney failure and UTUCs (40, 41), (iii) presence of AA-DNA adducts associated with UTUCs and RCCs (26, 34), and (iv) presence of the AA mutational signature in UTUCs, BCs, and...
RCCs from Taiwan (29–31, 34). However, despite the high amount of AA exposure in Taiwan and reports of the AA mutational signature in HCCs from China and other areas, AA exposure in Taiwan HCCs remains unexplored.

RESULTS

Overview of somatic changes in 98 HCCs from Taiwan

To investigate the possible presence of the AA mutational signature in HCCs from Taiwan, we sequenced the exomes of 98 HCCs and matched nonmalignant tissues from two hospitals (table S1). Tumor tissue was obtained from nonconsecutive patients, and inclusion in this study was solely based on the availability of adequate DNA. Tumors were not selected based on suspicion of AA exposure.

We sequenced whole exomes, with a mean of 95% targeted tumor bases with ≥30× coverage (table S2). We detected a total of 26,805 somatic single-base substitution (SBS) mutations across the HCCs (median, 167 SBS per tumor; interquartile range, 103 to 316), with an estimated false discovery rate (FDR) of 1.9% (tables S3 and S4). We detected a total of 648 short insertions or deletions (indels; median, 6 indels per tumor; interquartile range, 3 to 9), with an estimated FDR of 3.2% (tables S3 and S5).

In total, 10,174 genes harbored nonsilent SBS mutations (table S4). Driver analysis with MutSigCV (42) and 20/20+ (43) identified 16 significantly mutated genes (tables S6 to S8). The most commonly mutated genes—TP53, CTNNB1, ALB, and AXIN1—were also the most commonly mutated in the recent TGCA (The Cancer Genome Atlas) report (37), but the proportions of tumors with mutations in these genes were higher in the Taiwan HCCs (table S6). Among these genes, it has been proposed that ALB inactivation may promote cancer development by “diverting energy into cancer-relevant metabolic pathways” (37, 44). An additional gene identified by this analysis was IRF2, which was previously reported to act as a tumor suppressor in HCC (45). Of the other genes, two have not been identified as likely drivers in previous genome- or exome-wide resequencing of HCCs, and other evidence of their roles in cancer is absent or very limited, suggesting a lack of functional roles in HCC (table S6). Previously observed genetically amplified oncogenes and deleted tumor suppressors were also amplified or deleted in the Taiwan HCCs (fig. S1). These genes included the amplified oncogene CCND1 and the deleted tumor suppressor RB1 (37, 39, 46–49).

High rates of the AA mutational signature in Taiwan HCCs

The mutational spectra of most of the HCCs from Taiwan showed marked evidence of AA exposure, in the form of high proportions of A>T:T>A mutations in the trinucleotide contexts characteristic of AA-exposed tumors and cell lines (Fig. 1, A to D, and fig. S2) (29–31), although some HCCs did not show this evidence (Fig. 1E and fig. S2). The trinucleotide contexts characteristic of AA exposure included a prominent peak at 5′-CTG-3′ (5′-CAG-3′ on the complementary strand). There was also a notable excess of A>T mutations on the transcribed strands of genes, which is characteristic of AA-induced mutations in other tumor types and in cell lines (29–35). Principal components analysis clustered the majority of the Taiwan HCCs away from other HCCs and with previously reported AA-associated UTUCs (29, 30) and BCs (31) and with AA-exposed cell lines (Fig. 1F) (29).

To systematically assess the extent of AA exposure across the 98 HCCs, we developed the mSigAct (mutational signature activity) software. mSigAct provides a signature presence test to infer whether observed mutation spectra are better explained with a contribution from the AA mutational signature [Catalogue of Somatic Mutations in Cancer (COSMIC) signature 22] than without. We developed mSigAct because, to our knowledge, current approaches, most of which are based on nonnegative matrix factorization (NMF), do not support statistical inference of the presence or absence of a signature (3, 4, 50, 51). Briefly, the mSigAct test starts by generating optimal coefficients for reconstruction of the observed spectrum using the mutational signatures previously detected in HCCs. The test first does this without the AA signature (null hypothesis) and then with the AA signature (alternative hypothesis). The test then carries out a standard likelihood ratio test on these two hypotheses. Supplementary Materials and Methods and codes S1 and S2 provide details on the test, its evaluation on synthetic data, and the code. mSigAct revealed strong evidence of AA exposure in 76 of the 98 HCCs (78% with FDR < 0.05; Fig. 2, Table 1, and table S9). Among tumors with the AA signature, there was a median of 2.26 AA signature mutations/Mb (mean, 4.94 AA signature mutations/Mb).

As a further check on the mSigAct signature presence test, we also analyzed the 98 Taiwan HCCs with the NMF procedure in (3, 4) (code S1). The signature extracted by NMF had a Pearson correlation coefficient of 0.997 and a cosine similarity of 0.997 with the AA signature (COSMIC signature 22; fig. S3 and table S10). We also used NMF to detect the presence or absence of the AA signature and compared the results for this to those from the mSigAct signature presence test. The two procedures were concordant for 90 tumors (code S1). NMF identified eight putatively AA-exposed HCCs that mSigAct did not identify (T18, T41, T50, T57, T61, T68, and T92; fig. S2). Thus, the mSigAct signature presence test was more conservative; the tumors identified by NMF but not mSigAct had very low numbers of A>T mutations (all but one ≤15) in backgrounds of relatively high numbers of other mutations, making it difficult to be confident of AA exposure (code S1). We would propose that this is the desired characteristic, that is, it is preferable to err on the side of undercalling rather than overcalling the presence of the AA signature. Furthermore, testing on synthetic data also indicated that the mSigAct signature presence test had better sensitivity and specificity (Supplementary Materials and Methods and code S2).

We examined associations between the extent of exposure to the AA signature and multiple clinical and epidemiological variables, namely, hospital, cirrhosis status, hepatitis B carrier status, hepatitis C carrier status, status as carrier of either hepatitis virus, diagnosis before or after the medicinal use of some AA-containing plants was banned in Taiwan in 2003, gender, date of diagnosis, and age at diagnosis (fig. S4). Of these, without correction for multiple hypothesis testing, AA exposure differed significantly by gender and age at diagnosis. There was a weak association of increased AA exposure with age (Spearman’s ρ = 0.28, P = 0.008). In addition, AA mutation numbers were higher in females than in males (median, 176 versus 55 AA signature mutations per HCC; P = 0.015 by two-sided Wilcoxon rank sum test). After considering multiple hypothesis testing, the Benjamini-Hochberg FDRs for both gender and age were 0.065. Although AA mutation numbers were not statistically higher in women than in men, we note other evidence of more exposure to AA-containing herbs among women: In Taiwan before the ban, exposures were 31.6 person-years per 1000 for women compared to 25.9 for men (28). We also note that, because only 10 HCCs were hepatitis-negative, these data did not offer an opportunity to investigate interactions between hepatitis and AA exposure.

Ng et al., Sci. Transl. Med. 9, eaan6446 (2017) 18 October 2017

Fig. 1. Evidence of AA exposure in Taiwan HCCs. (A and B) Sample exome spectra of individual AA-exposed UTUCs (A) and BCs (B) from Taiwan. (C and D) Sample exome spectra of individual Taiwan HCCs with high (C) and moderate (D) levels of the AA signature. (E) Sample Taiwan HCC without AA signature. In the major plots in (A) to (E), each bar indicates the proportion of mutations in a particular trinucleotide context. In the AA signature (A to D), the overwhelming majority of mutations are T:A>A:T. By convention, mutations are shown as T>A (for example) rather than A>T, although AA mutations are physical consequences of adducts on adenines that cause A>T mutations (9, 20, 22–24). In tumors strongly mutagenized by AA, the most prominent peak is at CTG>CAG (CAG>CTG on the complementary strand), indicated in (A), often with additional prominent peaks at CTA>CAA and ATG>AAG. Small plots at right in (A) to (E) show transcription strand bias. Mut count, mutation count. (F) Mutation spectra–based principal components analysis of HCCs from Taiwan, China (52), and Japan (53), plus AA-exposed UTUCs (29) and BCs (31) and an AA-exposed cell line (29). The most distinguishable features are the T>A mutations induced by AA, which are reflected in PC1. PC1 explains 35% of the variance, and PC2 explains 5.5%.
The AA mutational signature in HCCs from other regions

Given the high prevalence of the AA signature in Taiwan HCCs, we examined publicly available data comprising 1400 HCCs (Table 1). These included data from China, Japan, Korea, and several countries in Southeast Asia (37–39, 46–48, 52, 53), as well as data from North America and Europe (37, 49) as negative controls with likely rare AA exposure.

We detected the AA signature in 42 of 89 HCCs (47%) from China (Figs. 2 to 4; Table 1; fig. S5, A and B; and table S11). Among the HCCs from earlier studies (47, 52), the mSigAct signature presence test detected many more affected HCCs than we were able to identify previously (29). Overall, however, AA signature mutation burdens were lower in China (median, 0.29 AA signature mutations/Mb) than in Taiwan (median, 2.26 AA signature mutations/Mb).

We detected the AA signature in five of nine HCCs from various countries in Southeast Asia other than Vietnam (56%; fig. S5, C and D) (39). Among the HCCs with the AA signature, the median mutation burden was high (2.9 AA signature mutations/Mb). We also detected the signature in 5 of 26 HCCs from Vietnam (19%) (37), with a high median mutation burden of 3.4 AA signature mutations/Mb (fig. S5E). We also detected the AA signature in lower proportions of the HCCs from Korea and Japan (Table 1; Figs. 2, B and C, and 3; and fig. S5, F to H).

We analyzed TCGA data (37) from areas other than Vietnam in several subgroups (fig. S5I). In the largest subgroup, North America, we...
<table>
<thead>
<tr>
<th>Geographic origin</th>
<th>Number of HCCs</th>
<th>Median (all SBS mutations/Mb)</th>
<th>HCCs with AA signature</th>
<th>Weighted counts* of AA signature SBS mutations among HCCs with the signature</th>
<th>Number of nonsilent SBS mutations in known driver</th>
<th>Number of HCCs with nonsilent AA signature SBS mutations in known drivers</th>
<th>P value for difference from Taiwan in proportion of HCCs with the AA signature (Fisher’s exact test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SBS mutations/Mb</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>New data, this paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>98</td>
<td>3.33</td>
<td>76</td>
<td>0.78</td>
<td>2.26</td>
<td>4.94</td>
<td>101.6</td>
</tr>
<tr>
<td>Previously published data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>89</td>
<td>1.94</td>
<td>42</td>
<td>0.47</td>
<td>0.29</td>
<td>1.07</td>
<td>5.8</td>
</tr>
<tr>
<td>Southeast Asia (excluding Vietnam)</td>
<td>9</td>
<td>5.74</td>
<td>5</td>
<td>0.56</td>
<td>2.92</td>
<td>3.07</td>
<td>62.3</td>
</tr>
<tr>
<td>Vietnam</td>
<td>26</td>
<td>3.10</td>
<td>5</td>
<td>0.19</td>
<td>3.42</td>
<td>3.71</td>
<td>125.4</td>
</tr>
<tr>
<td>Korea</td>
<td>231</td>
<td>1.78</td>
<td>29</td>
<td>0.13</td>
<td>1.00</td>
<td>1.29</td>
<td>41.4</td>
</tr>
<tr>
<td>Japan</td>
<td>477</td>
<td>4.82</td>
<td>13</td>
<td>0.027</td>
<td>0.60</td>
<td>0.94</td>
<td>8.2</td>
</tr>
<tr>
<td>North America</td>
<td>209</td>
<td>2.36</td>
<td>10</td>
<td>0.048</td>
<td>0.99</td>
<td>2.60</td>
<td>32.9</td>
</tr>
<tr>
<td>Europe</td>
<td>230</td>
<td>1.67</td>
<td>4</td>
<td>0.017</td>
<td>0.35</td>
<td>9.72</td>
<td>15.2</td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>89</td>
<td>2.94</td>
<td>19</td>
<td>0.21</td>
<td>1.30</td>
<td>2.27</td>
<td>39.2</td>
</tr>
<tr>
<td>No information</td>
<td>30</td>
<td>2.79</td>
<td>5</td>
<td>0.17</td>
<td>0.46</td>
<td>0.57</td>
<td>13.9</td>
</tr>
<tr>
<td>Total non-Taiwan</td>
<td>1400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*We ascribe the AA signature to a proportion of each SBS as described in Materials and Methods. Because the AA signature is dominated by A:T>T:A SBS, when the activities of other signatures with these SBS mutations are low or absent, this approximates the number of A:T>T:A SBS mutations. The weighted counts are the sums of the proportions of SBS mutations ascribed to the AA signature. $1^{P} = 7.1 \times 10^{-7}$ for all Southeast Asia, including Vietnam. *Includes 10 additional TCGA HCCs: 1 Russia, 3 Brazil, 4 South Korea; 1 Brazil HCC has the AA signature.
Fig. 3. Sample spectra of HCCs with the AA signature. Display conventions are the same as in Fig. 1.
detected AA signature mutations in 10 of 209 HCCs (5%; Table 1 and fig. S5J). Among HCCs from North America, the proportion with detected AA signature mutations in 10 of 209 HCCs (5%; Table 1 and Fig. 4. Global distribution of mutagenesis associated with aristolochic acid and derivatives in HCCs. The pie chart labeled “Southeast Asia” includes both Vietnam and the other Southeast Asian HCCs. Pie chart areas are proportional to the number of HCCs in the given group.

different from non-Asian patients (8/189).

We also detected AA signature mutations in 4 of 230 HCCs from Europe (1.7%; fig. S5K). This low proportion is consistent with the rarity of reports of AA exposure in Europe outside of the Balkans and the Belgian poisoning incident in the 1990s (9, 25, 32, 33). Furthermore, the median AA mutation burden was low (0.35 AA signature mutations/Mb), although one HCC with likely DNA mismatch repair deficiency had many more mutations.

Within the TCGA data (37), there were 89 HCCs from the Mayo Clinic for which the “Country” field had no data, and almost all of these (87) had “Ethnicity” listed as Asian. Among these, 19 (21%) had the AA signature (fig. S5L). Given the high prevalence of the signature and relatively high numbers of AA signature mutations in these HCCs (median, 1.3 AA signature mutations/Mb), we speculate that some of these patients may have traveled from Asia for treatment. In addition, there were 30 HCCs from biobanks for which no Country information was available. Of these, 20 were listed as Asian, and 5 (25%) of these had the AA signature, whereas none of the non-Asian HCCs had the AA signature (fig. S5M). Finally, one of five HCCs from Brazil with non-Asian ethnicity (20%) showed the AA signature (fig. S5N).

The effects of the AA signature were especially prominent in Taiwan: A higher proportion of HCCs from Taiwan showed the AA signature than in any group other than the nine HCCs from Southeast Asia (not including the HCCs from Vietnam; Table 1). Nevertheless, this analysis of publicly available data showed widespread AA exposure in East and Southeast Asia and in self-identified Asians elsewhere.

AA signature mutations in known cancer drivers

Our initial analyses with MutSigCV and 20/20+ did not reveal any strong possibilities for previously unknown driver genes in the Taiwan HCCs, but many genes listed in the Cancer Gene Census as known cancer drivers (table S12) were affected by nonsilent mutations ascribed to the AA signature (54). Across all Taiwan HCCs, the AA signature accounted for 59% (299 of 505) of nonsilent mutations in known driver genes (table S9). Among the Taiwan HCCs, 57 had a nonsilent AA mutation in ≥1 known driver (Fig. 2C, Table 1, and table S9). Among HCCs with the AA signature, two genes, TP53 and LRPIB, were mutated frequently by both A>T and by non-A>T mutations (39 and 27 total nonsilent mutations, respectively, of which 48 and 63% were AA signature mutations; table S13). Recurrent mutations in LRPIB could be due to its large size (4599 amino acids; Uniprot accession code Q9NZR2). It was not identified as a driver in our MutSigCV and 20/20+ analysis, and experimental evidence that it can function as a tumor suppressor is limited (55, 56). Several known tumor suppressors harbored predominantly AA signature mutations (table S13). Three of these are WNT-related tumor suppressors: AXIN1, AXIN2, and APC. Three others—ARID1A, ARIDB, and SETD2—are involved in chromatin remodeling, as is the oncogene KMT2A.

Tumors with the AA signature from regions other than Taiwan also had driver genes harboring nonsilent AA signature mutations (Table 1, Fig. 2C, and table S13). For example, 19 of the 29 AA-affected HCCs from Korea and all 5 of the AA-affected HCCs from Vietnam had AA signature mutations in known driver genes.

Clonality analysis of the Taiwan HCCs that had the AA signature suggested that AA mutations are predominantly early events, which is consistent with exposure before carcinogenesis (fig. S6). However, some AA signature mutations were subclonal, indicating that AA-associated mutagenesis, and presumably AA exposure, continued during tumor development and growth. Phylogenetic analysis based on multisector sequencing of HCCs from China in (38) showed that most AA mutations were truncal (found in all regions of the tumors), but some were subclonal, suggesting additional exposure to AA after initiation of carcinogenesis. A reanalysis of HCCs treated in Singapore showed a similar pattern of predominantly truncal AA signature mutations in four of the five AA-affected tumors (table S14) (39).

Potential for immunotherapy in high-AA-burden HCCs

A large proportion of HCCs in Taiwan bore heavy burdens of AA signature mutations (Fig. 2C and Table 1) and thus may be good candidates for immune checkpoint inhibitors (57). To investigate this possibility, we predicted neoepitopes arising from nonsynonymous mutations and then predicted neoepitopes binding to patient-specific human leukocyte
antigen (HLA) types. HCCs with the AA signature had many more predicted neoantigens (median, 146.5) than the remaining Taiwan HCCs (median, 60; \( P < 2 \times 10^{-5} \) by Wilcoxon rank sum test; Fig. 5).

**DISCUSSION**

Some AA-containing herbal remedies have been officially prohibited in Taiwan since 2003, and we looked for evidence of whether this ban reduced exposure. We detected no significant difference in the prevalence of the AA signature or in the numbers of AA signature mutations in HCCs diagnosed before and after 2003. There are a number of possible, nonexclusive explanations. One possibility is that the decline in incidence of AA-associated HCCs may simply be lagging behind reduced AA exposure. There is precedent for this in tobacco-associated lung cancer. In the United States, male death rates roughly doubled in the 25 years after the 1964 Surgeon General’s Report before beginning to decline in the 1990s, presumably as a result of tobacco suppression efforts that began decades earlier (58). In this context, we also note that AA-DNA adducts are extremely persistent (59).

Another possible explanation for unchanged prevalence of AA signature or numbers of AA signature mutations after the 2003 ban would be ongoing exposure to AA-containing herbal remedies. This could arise in various ways. (i) AA-containing plants may have still been prescribed by traditional Chinese medicine practitioners after the ban; this was documented to be the case in the first year after the ban (28). (ii) The nomenclature and labeling of products is confusing and error-prone; the herbs are often bought in formulations rather than individually, and in some traditional formulations, innocuous herbs can be replaced by herbs containing AA (12, 16, 28). (iii) Plants known to contain high concentrations of aristolochic acids are easily available on the internet, sometimes labeled correctly and sometimes incorrectly (table S15) (12, 16). (iv) Some plants containing AA were not banned in Taiwan. In particular, plants in the genus *Asarum*, collectively termed *xi xin* in Mandarin Chinese, were not banned and were the most commonly included plants in prescriptions surveyed in reference (28). As an example, powdered *xi xin* products manufactured in Taiwan and China were recently recalled in Singapore because they contained aristolochic acid I (fig. S7) (60, 61). Although reports of the presence and concentrations of aristolochic acids, aristolactams, and related compounds in *xi xin* are limited, available literature suggests that the concentrations are extremely variable and sometimes high (6, 12, 14, 62). An additional area of concern are aristolactams, which are not as intensively studied as the aristolochic acids but are thought to be the immediately mutagenic metabolites of aristolochic acids that interact directly with DNA (21).

There is strong evidence that the mutational signature that we and others have consistently observed is caused by aristolochic acids, possibly in conjunction with related compounds. Nevertheless, we cannot exclude the formal possibility that chemicals unrelated to aristolochic acids, aristolactams, and derivatives might also induce a mutational signature resembling the AA signature. However, at present, no such chemical is known, and many groups have independently and reproducibly detected a consistent AA mutational signature in UTUCs, BCs, RCCs, and bile duct cancers (table S10) (3, 29–37, 49, 63). Evidence that this signature is caused by AA and related compounds includes the mutational spectra of AA-treated cell lines (29) and the signature’s association with AA-DNA adducts (30, 34) and with AA-related nephropathy in Taiwan and the Balkans. Furthermore, animal studies have shown that AA adducts, and presumably AA mutagenesis, occur in the liver (64–66).

To summarize the findings of this study, mutational signature analysis implicated AA exposure in 78% of HCCs from Taiwan. The AA signature was much more prevalent, and the number and proportion of mutations were notably higher in Taiwan than in most other regions. At the same time, AA exposure was found in cohorts from all Asian countries examined and in 22% of Asian patients treated at the Mayo Clinic. Across Taiwan HCCs, 299 of 505 nonsilent mutations in known driver genes were ascribed to the AA signature. Among the 76 AA-affected Taiwan HCCs, 57 had ≥1 nonsilent AA mutation in a known driver, and among the 133 AA-affected HCCs from elsewhere, 56 had ≥1 nonsilent AA mutation in a known driver gene, suggesting an active role for AA in the origins of these HCCs.

The findings here indicate that exposure to aristolochic acids and their derivatives is geographically widespread, implying substantial opportunities for primary and secondary prevention (Fig. 4). Medicinal use of AA-containing plants is only lightly regulated in many jurisdictions. The plants are not banned outright in China (67), and even in Taiwan, to the best of our understanding, only specific plants, rather than any plant and product containing AA or its derivatives, are regulated. Strikingly, *xi xin*, the most commonly prescribed herb before 2003, is not banned (28). In the United States, sale of AA-containing herbs is unregulated provided that they are correctly labeled and there are no claims of health benefits (68). Furthermore, plants containing aristolochic acid and its derivatives are readily available for sale on the internet (table S15).

In light of the wide availability of AA-containing plants, education and public awareness are paramount for primary prevention. In addition, the traditional nomenclature is confusing, making it difficult for consumers and suppliers to be sure of plant identification or of the constituents of multiherb preparations; there is ample evidence that mislabeling is common (12, 16, 25, 28). This latter point suggests that more thorough methods for testing herbal products, such as chromatographic fingerprinting, combined with regulatory oversight of the supply chain could also help reduce exposure (69). Secondary prevention might take the form of enhanced screening for AA-associated cancers or for kidney disease in patients suspected or known to be exposed to AA.

**MATERIALS AND METHODS**

**Study design**

This was designed as an exploratory retrospective study, because it was not known in advance whether HCCs from Taiwan would show evidence of the AA mutational signature or what the signature’s prevalence might be. As noted above, tumor tissue was obtained from nonconsecutive patients, and inclusion of HCCs from Taiwan in this study was solely based on availability of adequate DNA and was not selected based on suspicion of AA exposure. After we discovered prevalent AA signature mutations in HCCs from Taiwan, we extended the study to publicly available somatic mutation data from 1400 tumors.

**Patients and preparation of clinical samples**

HCCs were diagnosed and classified by histopathological examination of surgically excised tumors, according to the World Health Organization classification system. Snap-frozen liver tumor tissues and matched normal samples (whole blood) from patients with HCC were obtained from Chang Gung Memorial Hospital Taiwan (21 patients) and National Taiwan University (77 patients). The human samples were sourced ethically with informed consent, and their research use was in accordance with the protocols approved by the Chang Gung Memorial Hospital.
(103–6534C) and National Taiwan University Institutional Review Boards and Human Research Ethics Committees. Table S1 provides clinicopathological data and information on sequencing.

**Whole-exome sequencing and mutation identification**

Samples were captured with the Agilent SureSelect V5 exome panels. Paired 101–base pair reads were generated on HiSeq 2500 sequencers. BWA-MEM aligned reads to the human reference genome (hg19) (70), SAMtools removed polymerase chain reaction duplicates (71), and Qualimap 2 computed quality control metrics (72). Candidate somatic mutations were initially identified by three callers: GATK (73), Strelka (74), and MuTect (75). SBS called by ≥2 callers and small indels called by both GATK and Strelka were curated for downstream analysis. Examination of ≥1% of SBS calls and ≥1% indel calls from each tumour-normal pair (with ≥1 mutation from each tumour-normal pair) in IGV indicated an FDR of 6/312 (1.9%) of SBS calls and 3/93 (3.2%) of indel calls (table S3). Indels were rare, and their length distribution was unremarkable (fig. S8).

**Sources of publicly available HCC somatic mutation data**

These were as follows: (i) whole-genome sequence (WGS) from 78 HCCs from China (47, 52), in which we had previously noted the AA signature in 11 tumors (29), downloaded read data from ftp.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/ERP/ERP0001/ERP0001196/ and realigned and called as described (76); (ii) whole-exome sequence (WES) from 11 HCCs from China (38), variant calls downloaded from the supplementary information for that paper (http://cancerres.aacrjournals.org/highwire/filestream/334829/field_highwire_adjunct_files/1/173141_1_supp_3805627_mj6gp5.xls); (iii) WES and WGS from nine HCCs in patients from various Southeast Asian countries, treated in Singapore (39), variant calls downloaded from the supplementary information for that paper; (iv) WES from 231 HCCs from Korea (48), downloaded from cBioPortal (www.cbioportal.org); 8 February 2017 (77); (v) WGS from 264 HCCs from Japan (53), downloaded from the International Cancer Genome Consortium (ICGC) data portal (https://dcc.icgc.org/; March 2015, release 18, before publication of the paper); (vi) 213 WES from Japan, in which a mutational signature that, in retrospect, appears to be a merge of the AA signature with another signature had been reported (46), downloaded from the ICGC data portal (https://dcc.icgc.org/; March 2015, release 18); (vii) WES from 230 HCCs from France, Spain, and Italy, which likely are not regions of widespread AA exposure (49), downloaded from the ICGC data portal (https://dcc.icgc.org/; March, 2015, release 18); (viii) WES from 364 HCCs from North America, Vietnam, and a few other regions (fig. S51) (37), downloaded from the supplementary information from that publication.

**Analysis of mutational signatures**

We used the R (www.r-project.org) function prcomp to compute the first two principal components over 96-channel mutation spectra (Fig. 1F). The sources of the data were as follows: UTUCs and AA-exposed cell lines (29), AA BCs (31), and HCCs from Japan (53) and China (47, 52). Supplementary Materials and Methods and code S1 describe in detail the mSigAct signature presence test and its comparison to the NMF approach in (3, 4).

**Allocating mutational signature contributions to mutations ("weighted mutation counts")**

For a given tumor, we allocated the partial contribution of each signature to each type of mutation as follows. Let $t_1 \ldots t_{96}$ be the 96 strand invariant mutation types in trinucleotide context, namely, ACA>AAA, ACA>AGA, ACA>A TA, CCA>CAA, \ldots, TTT>TAT, TTT>TCT, and TTT>TGT (see also the labels at the bottom of Fig. 1E). Let $n_k$ be the number of mutational signatures, and let $e_1 \ldots e_n$ be the exposures of the tumor to each of the signatures. Let $\pi(t_j, S_i)$ be the proportion of mutation type $t_j$ in signature $S_i$ with $1 = \sum_{j=1}^{96} \pi(t_j, S_i)$. Then, in a given tumor, we define the partial contribution of $S_i$ to each instance of $t_j$ as

$$\pi(t_j, S_i) \cdot \frac{e_i}{\sum_{k=1}^{n} \pi(t_j, S_k) \cdot e_k}$$

in which the denominator ensures that the partial credits sum to 1.

**AA signature mutations in driver genes**

We used MutSigCV (42) and 20/20+ (43) to identify candidate drivers in the Taiwan AA HCCs (tables S6 to S8). The MutSigCV preprocessor was used to generate the MAF file, and MutSigCV was run with default parameters. 20/20+ was run with default parameters except using the “pretrained 2020plus_100k.Rdata” classifier.

To analyze nonsilent mutations in known cancer driver genes, we used the Cancer Gene Census (http://cancer.sanger.ac.uk/census; downloaded 3 January 2017) and selected the 159 genes listed as “oncogene” or “TSG” (tumor suppressor gene) in the “Role in cancer” column (table S12) (54).

**Neoantigen prediction**

Nonsynonymous somatic variants were annotated by wANNOVAR (http://wannovar.wglab.org/) (78), and a custom script generated all possible 9–amino acid sequences containing the mutated residue. In silico HLA typing of individual tumors was carried out using OptiType for major histocompatibility complex class I genes and used for HLA allele–specific peptide binding predictions (79). NetMHC4.0 and NetMHCPan2.8 were used to predict peptide binding (80, 81). Rank parameters >2 were considered nonbinding, and those ≤2 were considered binding, as suggested in (80).

**SUPPLEMENTARY MATERIALS**

www.sciencetranslationalmedicine.org/cgi/content/full/9/412/eaan6446/DC1

Materials and Methods

Fig. S1. GISTIC analysis of significant amplifications and deletions in Taiwan HCCs.

Fig. S2. Mutational spectra of all 98 individual Taiwan HCCs.

Fig. S3. Comparison of COSMIC signature 22 with AA mutational signature extracted from all Taiwan HCCs.

Fig. S4. Associations between the number of AA signature mutations and clinical and epidemiological variables.

Fig. S5. Mutational spectra of all individual HCCs with the AA signature from publicly available data.

Fig. S6. Examples of clonal and subclonal AA SBS mutations in Taiwan HCCs.

Fig. S7. Two recall notices from Singapore for xin products containing aristolochic acid I.

Fig. S8. Length distributions of small indels in 98 Taiwan HCC exomes.

Fig. S9. Workflow for generating synthetic mutation data for testing.

Fig. S10. Receiver operating characteristics of LA-NMF for AA signature detection.

Fig. S11. Receiver operating characteristics for AA detection by mSigAct and LA-NMF.

Table S1. Clinicopathological parameters and statistics on sequencing for 98 HCCs and matched normal tissues from Taiwan.

Table S2. Percent targeted bases at ≥30x coverage.

Table S3. FDR estimated from IGV screenshots.

Table S4. Somatic SBS mutations in Taiwan HCCs.

Table S5. Somatic indel mutations in Taiwan HCCs.

Table S6. Drivers identified by MutSigCV and 20/20+ in 98 Taiwan HCCs.

Table S7. MutSigCV output for 98 Taiwan HCCs.

Table S8. 20/20+ output for 98 Taiwan HCCs.
Table S9. AA signature mutations and effects on driver genes in 98 Taiwan HCCs.
Table S10. Comparison of LA-NMF
Table S11. List of AA signature
Table S12. Known oncogene and tumor suppressor drivers from COSMIC Cancer Gene Census.
Table S13. Nonsilent mutations in known cancer driver genes plus genes identified by MutSigCV or 20+iso.
Table S14. Subclonality analysis of AA mutations in published HCC multiregion sequencing studies.
Table S15. Likely AA-containing plants for sale on the internet.
Table S16. Selecting the negative binomial dispersion parameter for mSigAct.

REFERENCES AND NOTES

Table S1. Comparison of LA-NMF-extracted AA signatures with COSMIC 22.

Table S2. Analysis and tests of HCCs with mSigAct and the NMF procedure from publicly available HCC spectra.

Table S3. True- and false-positive rates for detection of the AA signature by mSigAct and LA-NMF.

Table S4. Comparison of detection of the AA signature by mSigAct and LA-NMF on 1400 publicly available HCC spectra.

Table S5. Likely AA-containing plants for sale on the internet.

Table S6. Selecting the negative binomial dispersion parameter for mSigAct.

Table S7. True- and false-positive rates for detection of the AA signature by mSigAct and LA-NMF.

Table S8. Comparison of detection of the AA signature by mSigAct and LA-NMF on 1400 publicly available HCC spectra.

Table S9. AA signature mutations and effects on driver genes in 98 Taiwan HCCs.

Table S10. Comparison of LA-NMF

Table S11. List of AA signature

Table S12. Known oncogene and tumor suppressor drivers from COSMIC Cancer Gene Census.

Table S13. Nonsilent mutations in known cancer driver genes plus genes identified by MutSigCV or 20+iso.

Table S14. Subclonality analysis of AA mutations in published HCC multiregion sequencing studies.

Table S15. Likely AA-containing plants for sale on the internet.

Table S16. Selecting the negative binomial dispersion parameter for mSigAct.


Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia

Aristolochic acids, an herbal compound found in many traditional medicines, had been previously linked to kidney failure, as well as cancers of the urinary tract. Because of these known toxicities, herbs containing this compound have been restricted or banned in some countries, but it is still available on the internet and in alternate formulations. By analyzing numerous samples from Taiwan and other countries in Asia and elsewhere, Ng et al. demonstrated the effects of aristolochic acid in hepatocellular carcinoma, a much more common tumor type. The authors showed that the use of this drug remains widespread in Asia and particularly in Taiwan, and that it appears to increase the risk of multiple cancer types.