Aristolochic acid in the etiology of renal cell carcinoma

Margaret L. Hoang1*, Chung-Hsin Chen2*, Pau-Chung Chen3, Nicholas J. Roberts1, Kathleen G. Dickman4,5, Byeong Hwa Yun6, Robert J. Turesky6, Yeong-Shiau Pu2, Bert Vogelstein1, Nickolas Papadopoulos1, Arthur P. Grollman4,5, Kenneth W. Kinzler1,**, Thomas A. Rosenquist6,**

Affiliations:

1Ludwig Center for Cancer Genetics and Therapeutics and The Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. 2Department of Urology, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan 10002. 3Department of Occupational and Environmental Medicine, National Taiwan University Hospital and Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University. 4Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794, USA. 5Department of Medicine, Stony Brook University, Stony Brook, NY 11794, USA. 6Masonic Cancer Center and Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA.

* These authors contributed equally to this study
** To whom correspondence should be addressed: kinzike@jhmi.edu, thomas.rosenquist@stonybrook.edu

Thomas Rosenquist
Department of Pharmacological Sciences, School of Medicine
Stony Brook University
BST Level 7, Room 145
Stony Brook, NY 11794-8651
1-631-444-8054, 1-631-444-3218 (fax)

Keywords: Aristolochic acid, renal cell carcinoma, mutational signature, DNA damage,

Running Title: Aristolochic acid in the etiology of renal cell carcinoma

Competing interests: Under agreements between the Johns Hopkins University, Genzyme, Exact Sciences, Sysmex-Inostics, LabCorp, PapGene, Qiagen, Invitrogen, Morphotek and Personal Genome Diagnostics, N.P., B.V. and K.W.K. are entitled to a share of the royalties received by the University on sales of products related to genes reported in this work. N.P., B.V. and K.W.K. are co-founders of PapGene, Sysmex-Inostics, and Personal Genome Diagnostics, are members of their Scientific Advisory Boards, and own PapGene and Personal Genome Diagnostics stock, which is subject to certain restrictions under Johns Hopkins University policy. The terms of these arrangements are managed by The Johns Hopkins University in accordance with its conflict-of-interest policies.

Financial Support: This research was supported, in part, by Taiwan National Science Council grants (101-2314-B-002-027-MY3) to CH Chen and (101-2314-B-002-023-MY3) to YS Pu. Additional support included The Virginia and D.K. Ludwig Fund for Cancer Research, Commonwealth Foundation, and NIH grants CA057345 (K.W.Kinzler), CA190889 (N.J. Roberts) and ES019564 (R.J. Turesky) and HHMI (B.V. Vogelstein) and Marsha and Henry Lauffer (A.P. Grollman).

Word Count: 3998 words, four figures and two tables, one supplementary section and figure and four supplementary tables, and 33 references.
Abstract

Background: *Aristolochia* species used in the practice of traditional herbal medicine contain aristolochic acid (AA), an established human carcinogen contributing to urothelial carcinomas of the upper urinary tract. AA binds covalently to genomic DNA, forming aristolactam (AL)-DNA adducts. We here investigated whether AA is also an etiologic factor in clear cell renal cell carcinoma (ccRCC).

Methods: We conducted a population-based case-control study to investigate the linkage between *Aristolochia* prescription history, cumulative AA consumption, and ccRCC incidence in Taiwan (5,709 cases and 22,836 matched controls). The presence and level of mutagenic dA-AL-I adducts were determined in the kidney DNA of 51 Taiwanese ccRCC patients. The whole exome sequences of ccRCC tumors from ten Taiwanese ccRCC patients with prior exposure to AA were determined.

Results: Cumulative ingestion of more than 250 milligrams of AA increased risk of ccRCC (OR 1.25) and we detected dA-AL-I adducts in 76% of Taiwanese ccRCC patients. Further, the distinctive AA-mutational signature was evident in six of ten sequenced ccRCC exomes from Taiwanese patients.

Conclusions: This study strongly suggests that AA contributes to the etiology of certain renal cell carcinomas.

Impact: The present study offers compelling evidence implicating AA in a significant fraction of the RCC arising in Taiwan and illustrates the power of
integrating epidemiological, molecular and genetic data in the investigation of cancer etiology.

**INTRODUCTION**

Genome-wide sequencing has the capacity to link mutational signatures to specific mutagenic agents. Notable examples include C--to-T transitions in pyrimidine dimers induced by ultraviolet radiation (1) and C-to-A transversions induced by tobacco exposure (2). A recent bioinformatics survey of tumor types occurring primarily in western populations identified at least 21 mutational signatures (3). However, for most of these putative signatures, the exogenous or endogenous agent responsible has not been identified. This gap could, in principle, be filled through molecular epidemiologic studies linking exposure to agents with the mutational signature observed.

Another remarkable example of a mutational signature associated with a specific carcinogen is provided by data on upper tract urothelial carcinomas (UTUC) in Taiwan and the Balkans. These tumors harbor a high content of A-to-T transversions, affecting adenines genome-wide within a set of specific trinucleotide sequences (4, 5). This mutational signature has been linked to aristolochic acid (AA), a nitrophenanthrene carboxylic acid found in *Aristolochia* species (6) used worldwide in the practice of traditional herbal medicine (7). Following metabolic activation, AA forms AL-DNA adducts that serve as specific biomarkers of exposure to AA (8). Remarkably, AL-DNA adducts can be detected in normal human tissues decades after exposure to AA (9-11). These adducts
were also found in the normal kidney tissues of individuals in Taiwan (12), a country with one of the highest rates of UTUC in the world. In Taiwan, prescription data reveal that one in three residents had ingested herbs containing AA (13).

In rodents exposed to AA, pro-mutagenic AL-DNA adducts are found in kidney, liver, forestomach, and bladder (14, 15). Thus, we hypothesized that AA contributes to cancers of the corresponding tissues in humans. In fact, signatures resembling the AA-mutational signature have been reported to occur in bladder cancer (BC) (16), a small fraction of hepatocellular carcinomas (HCC) in China (5, 17), Japan (18) and the US (18), and in intrahepatic cholangiocarcinomas (IHCC) from China (19), and in clear cell renal cell carcinomas (ccRCC) in China (20), Romania (21) and the endemic region of Croatia (22). AL-DNA adducts have also been detected in kidney DNA from Romanian ccRCC patients (23). Thus AA could be involved in the initiation and/or progression of tumors in these tissues. However, for HCC, IHCC and ccRCC, integrated epidemiological and molecular evidence linking AA exposure to the incidence of these cancers has not been available.

In the present study, we bridge this gap by applying both epidemiological and molecular approaches to the Taiwanese population, where a high fraction of the population has been exposed to AA (13). Specifically, we provide strong evidence linking AA exposure to ccRCC pathogenesis in this population by demonstrating an association between AA exposure and ccRCC incidence, by measuring AL-DNA adducts in the renal cortex of ccRCC patients, and by
detection of the AA-mutational signature in ccRCC tumors from patients previously exposed to AA.

MATERIALS AND METHODS

The research protocols were reviewed and approved by the Institutional Review Boards of Stony Brook University and the National Taiwan University.

Prescription database analysis

The Taiwan National Health Insurance (NHI) covers >96% of Taiwanese residents and reimburses cost of prescriptions, including Chinese herbal products. RCC cases were identified from the NHI catastrophic illness registry using ICD-9 code 189.0. A total of 5,709 RCC patients and 22,836 controls were enrolled from 1999 to 2008. The index date for each case was the date of diagnosis with RCC, and we defined the index date of control subjects as the date of RCC diagnosis of their corresponding cases. RCC patients with other cancers (ICD-9 140-208) or kidney transplant (V42.0) were excluded from this study. Each case was paired with 4 controls randomly selected from the insured population, which were matched by sex, age, income and urbanization.

The Taiwan NHI regularly reimbursed enrollees for the cost of prescribed Chinese herbs medicines containing AA, which were prescribed extensively before the ban in 2003. During our investigation period, the detailed herbal prescriptions, including regimen, dosage, duration and prescription date were uploaded to NHI for obtaining payment. All prescribed herbal medications were covered by the Taiwan NHI and required a doctor's prescription. Usage of AA-containing herbs in each case and control was determined and calculated from Taiwan NHI database from January 1997 to October 2003. Cumulative AA
exposure dosage was also calculated as described in a previous study that demonstrated a positive association between AA consumption and risk of urothelial carcinoma (24).

Known risk factors of RCCs were treated as potential confounders defined by the following diagnoses recorded between January 1, 1997, and 1 year before the diagnosis of RCCs or index dates: hypertension (ICD-9 401), diabetes (250), and hyperlipidemia (272), chronic obstructive pulmonary disease (491, 492, 496), chronic hepatitis C infection (070.7, 070.41, 070.44, 070.51, 070.54, V02.62), chronic kidney disease (585), cystic kidney disease (753.1), and kidney stones (592.0, 592.1, 592.9). Logistic regression was used to assess RCC risk based on the cumulative dose of AA. The odds ratios (ORs) and 95% CIs for RCC were calculated and estimated as crude and adjusted for covariates including sex, age, monthly income, urbanization level, hypertension, diabetes, hyperlipidemia, chronic obstructive pulmonary disease, chronic hepatitis C infection, chronic kidney disease, cystic kidney disease, kidney stones, aspirin, NSAIDs, and acetaminophen. Of these covariates only sex, monthly income, urbanization level, chronic obstructive pulmonary disease, chronic hepatitis C infection, aspirin, NSAIDs, and acetaminophen affected the results. All of these analyses were conducted using SAS statistical software (version 9.2; SAS Institute, Cary, NC).

**Genomic DNA for sequencing and adduct analysis.** RCC patients undergoing radical nephrectomy at National Taiwan University Hospital between December,
1998 and May, 2007 were enrolled. Patients with previous radiotherapy or systemic chemotherapy were excluded. Tissue specimens were sampled immediately after surgery. Renal cortical samples were taken from normal-appearing cortex far-removed from renal tumors. Tumor tissues were sampled so as to avoid inclusion of surrounding normal tissues. All fresh tissue samples were placed in aseptic vials, snap-frozen in liquid nitrogen and then stored at -80°C till DNA extraction as described (12).

**Whole Exome Sequencing** Sequencing and mutational analysis methods are in Supplementary Material and as described in Hoang, et al. (4).

**Mass spectrometric determination of AL-DNA adducts** AL-DNA adduct concentrations in five micrograms of kidney DNA were determined by ultraperformance liquid chromatography–electrospray ionization/multistage scan mass spectrometry as previously described (25, 26).

**RESULTS**

**Evidence of exposure to aristolochic acid-containing herbs in patients with ccRCC in Taiwan**

In Taiwan, 39% of the population received prescriptions for *Aristolochia*-containing remedies during the years 1997-2003 (13). The results (Table 1) indicate an adjusted OR of 1.25 (1.004-1.547) for ccRCC in persons consuming greater than 250 mg of AA during the period of 1997-2003. Each 100 mg of AA consumed contributed 1.03 (1.005-1.054) to the odds-ratio. As our study only
monitored exposure during the six-year period for which prescription records are available, it is likely that OR underestimates the impact of AA on RCC.

**Aristolactam DNA-adducts levels in renal tissues of ccRCC patients**

Molecular evidence that Taiwanese ccRCC patients were exposed to AA was obtained using a quantitative mass-spectrometric method to measure dA-AL-I-DNA adduct levels. Genomic DNA was isolated from the non-neoplastic renal cortical tissue of 51 Taiwanese RCC patients (Supplementary Table 1). AA-exposure induces dA and dG adducts derived from AAI and AAII (Fig. 1); the dA-AL-I adduct is resistant to DNA repair and can be detected in renal cortex DNA decades after exposure. We detected AL-DNA adducts in 39 of 51 (76%) patients tested (Fig. 1). In samples with detectable amounts of adducts, levels ranged from 0.3 to 258 adducts per $10^8$ bases; the average and median values were 36.3 and 20.8 adducts per $10^8$ bases, respectively. Thus, these data indicate that at least 76% of the patients in our cohort consumed AA-containing herbs. Importantly, in these individuals, dA-AL adducts were found within the same tissue that generated the tumor.

**Whole exome sequencing of ccRCCs**

We hypothesized that if AA exposure directly contributed to ccRCC tumorigenesis, then the tumor genome should harbor the characteristic AA mutational signature. To test this hypothesis, we performed whole exome sequencing on tumor and non-tumor DNA pairs from 5 men and 5 women who had been exposed to AA. Each patients selected had significant amounts of AL-DNA adducts in their renal DNA (Fig. 1). From the sequencing data, we identified
an average of 22,158 known single nucleotide polymorphisms (SNPs) per individual (Supplementary Tables 2, 3 and 4). We next estimated the neoplastic cell content of the samples by determining the average mutation allele fraction across all genes. The median estimate of neoplastic cell content was 44% (range, 26 to 78). Tumor purity in ccRCC samples is generally low, as reported in The Cancer Genome Atlas (TCGA) exome sequencing study of 417 ccRCCs (median 54% ± 14%) (27). Due to the high coverage of our whole exome sequencing, this did not pose a limitation to the identification of somatic mutations. The average high quality coverage of each base in the targeted region was 96-fold; 93% of the targeted region contained at least 10 reads. Furthermore, we observed no correlation between neoplastic cell content and the total number of somatic mutations identified for each tumor (Table 2).

Identification of somatic mutations

We used stringent criteria to identify a total of 1,204 somatic mutations with a median of 87 per tumor (range, 18 to 334) (Fig. 2A and Supplementary Table 2). For each tumor, any mutation in the VHL and PBRM1 genes, and six to nine randomly chosen mutations, were selected for validation by Sanger sequencing (Supplementary Table 3). For clonal somatic mutations with mutant fractions greater than or equal to 20%, 91% percent (46/51) of mutations cross-validated, indicating a low false positive rate. In contrast, 36 randomly chosen changes with mutant fractions below 20%, which are more likely to represent subclonal mutations, were validated in only 15 cases (41.7%). This lower validation fraction might reflect a higher false positive rate for subclonal
mutations but is more likely to reflect the lower sensitivity of Sanger sequencing. Of the total number of mutations, 94.8% (1141/1204) were single-base substitutions (SBSs) while 5.2% were indels (63/1204). Nonsynonymous mutations (protein-altering) accounted for 78.9% of total mutations with a median of 65 nonsynonymous mutations per tumor (range, 14-278). The ratio of nonsynonymous-to-synonymous mutations were not significantly different between AA-exposed ccRCC at 3.7 (950/254) and TCGA ccRCCs at 3.5 (20323/5870) (Fig. 2B, P = 0.3, two-sided Chi-square), indicating that AA exposure did not significantly alter this genome-wide parameter.

Nonsynonymous mutations were found in 853 genes, including several genes found to be significantly mutated in prior studies of ccRCC. VHL was the most frequently mutated driver gene (seven of 10 tumors) in our AA-exposed ccRCCs (Table 2). The second most frequent was PBRM1, with inactivating mutations in two out of ten tumors. VHL and PBRM1 were also the two most frequent significantly mutated driver genes in the TCGA ccRCC study and occurred at a similar mutation frequency as our cohort (53% VHL, 34% PBRM1) (27). We also observed mutations in other known ccRCC driver genes, including the tumor suppressors SETD2, BAP1, GNB2L1, and EPAS1, as well as an oncogenic mutation in PIK3CA (Supplementary Table 2). Furthermore, we compared the allelic fractions from SNPs in the tumor and normal pairs to identify regions of somatic allelic imbalance. Loss of chromosome 3p is the most frequent somatic alteration in ccRCC(28), as this region contains a number of driver genes, including VHL. Indeed we detected chromosome 3p arm loss in six
of 10 tumors (Supplementary Fig. 1). These data suggest that AA-exposed and TCGA individuals share common ccRCC driver genes.

**Mutational patterns of single base substitutions in ccRCC**

We next asked if the AA-exposed ccRCC tumors harbored the main feature of the AA mutational signature, an elevated frequency of A-to-T transversions. Indeed the fraction of A-to-T transversion mutations in the ten AA-exposed ccRCC was significantly higher (33±23%, mean±s.d.) than in the TCGA ccRCCs (10.7±4.4%) (P<0.0001, two-tailed t-test) (Fig. 3A). However, we observed a wide distribution of A-to-T transversion fraction in our ten AA-exposed ccRCCs (range 7-69%) compared to TCGA ccRCCs (range 0-26%).

We took advantage of the TCGA ccRCC cohort data to define a stringent criterion for the expected fraction of A-to-T transversions. Two standard deviations above the TCGA ccRCC cohort average was set as a threshold and tumors with A-to-T transversion fractions above this value (>20%) were designated as having elevated A-to-T. In six tumors of our AA-exposed cohort, the A-to-T fraction exceeded 20% of the SBSs (Fig. 3B). Note that 2.4% of the TCGA ccRCCs (10 of 417 tumors) has an A-to-T fraction greater than 20%. However, without the complementary AL-DNA adduct analysis to implicate AA, the elevated A-to-T fraction could have been due to other factors, including compounds whose metabolites lead to A-to-T mutations (29).

We compared the SBS mutational spectra of the following three ccRCC cohorts: (1) AA-exposed individuals with high A-to-T fractions (n=6), (2) AA-exposed individuals with low A-to-T fractions (n=4), and (3) TCGA individuals
The difference between the high A-to-T set and the TCGA dataset was primarily in the excess of A-to-T transversion mutations among the SBSs. After the A-to-T class is removed, the relative ranking of the other classes of SBSs is similar to that seen in the TCGA samples. This pattern in AA-associated ccRCC - an excess of A-to-T transversions - is reminiscent of that observed in AA-associated UTUC.

**Aristolochic acid mutational signature**

We examined 488 A-to-T mutations compiled from the six ccRCC cases with an excess fraction of A-to-T transversions for key features of the AA-mutational signature observed in AA-associated UTUC. dA-AL adducts are repaired efficiently by transcription-coupled DNA repair but only poorly by global genome repair (30). Indeed the deoxyadenosine residue mutated in A-to-T transversions was found primarily on the non-transcribed strand in these ccRCCs (Fig. 4A). In contrast, we observed no strand preference for the deoxyadenosine residue among A-to-T transversions in the TCGA dataset (Fig. 4A). The A>T strand biases in AA-associated ccRCCs and AA-associated UTUCs were 2.1-fold and 2.6-fold, respectively(4).

The sequence context of A-to-T mutations was probed by tabulating the bases neighboring the mutated deoxyadenines. Figure 4B shows the frequencies of the sixteen potential trinucleotides in which a dA residue in the central position is mutated to form an A-to-T transversion. The 5’CpApG trinucleotide includes 29% of the A-to-T mutations (141 A>T in CpApG of 488 XpApX) in the Taiwanese ccRCCs compared with only 11% (273/2575) in the TCGA dataset.
(P=2.3^{27}, \text{two-sided Chi-square}). In Taiwanese AA-associated UTUCs, the CpApG trinucleotide was also found in similar proportion, 33% (3365/10326) in AA-associated UTUCs compared to 10% (503/5250) in controls (4).

Exon splice acceptor sites almost always occur at ApG sequences and are most frequently found in a CpApG context. Consistent with an increased A-to-T preference for the CpApG context, the frequency of mutations in the splice acceptor was elevated in ccRCC compared to TCGA (2.8% versus 1.0% of substitutions, respectively) (Fig. 4C). The ratio of the number of mutations at splice acceptor to splice donor sites was 3.4-fold (24/7) in the Taiwanese ccRCCs. In contrast, there was no splice site mutation preference observed among ccRCCs in the TCGA dataset (1.1-fold, 269/251; \text{P}=0.0021 \text{two-sided Fisher's Exact Test}). The excess of splice acceptor mutations was less than that observed in AA-associated UTUC exome sequences (6.9-fold)(4). In summary, all significant features of the AA-mutational signature reported for AA-associated UTUCs were found in this set of Taiwanese ccRCCs and were highly statistically significant. We made no attempt to match our cohort to the TCGA with respect to known ccRCC risk factors such as smoking history and BMI. But as the TCGA set completely lacks the AA-mutational signature we can confidently conclude that these known risk factors, represented abundantly in the TCGA set, do not produce the mutational signature we are attributing to AA in ccRCC.

**DISCUSSION**

The widespread exposure of the Taiwanese population to herbs containing aristolochic acid has been well documented through analysis of the
national prescription database (13). Moreover, exposure to AA has been directly demonstrated via the detection of AL-DNA adducts in target tissues and by the AA-mutational signature in UTUC (12). The incidences of kidney and upper tract urinary cancer are similar in the Taiwanese population. The age specific rate (ASR, adjusted to the world standard population), as reported in the Taiwan Cancer Registry, for renal pelvis and ureter cancer (primarily UTUC) from 2003–2007 was 2.9 per 10^5 years in men and 3.3 per 10^5 years in women while the ASR for kidney cancers (primarily ccRCC) was 3.3 per 10^5 years in men and 1.9 per 10^5 years in women.

Similar to what had been previously found with urothelial carcinoma, there is a dose-dependent relationship between consumption of AA-containing herbs and ccRCC incidence. We could document exposure to AA through prescribed medicines in 38.3% of Taiwanese with ccRCC. Further, the presence of AL-DNA adducts revealed the exposure of 75% of Taiwanese ccRCC patients to this potent and highly persistent human carcinogen, indicating additional exposure beyond the estimate based on prescriptions reimbursed by NHI. This underestimate is expected in assessing life-long exposure by a seven-year window. Other contributors are potential misclassification of patients that obtained herbal remedies outside of the NHI system. This same underestimate has affected previous epidemiological studies of AA and UTUC utilizing the prescription database. Lai, et al, (24) reported that 28.7% of UTUC patients had been reimbursed for prescriptions containing AA but Chen, et al, (12) reported that 60% of UTUC patients had AL-DNA adducts. Whole exome sequencing
confirmed that the AA-mutational signature was present in six of ten ccRCC patients studied. Previous examination of the \textit{TP53} gene in Taiwanese RCC did not detect the AA-mutational signature due to limited sample size; only three \textit{TP53} mutations were detected in 25 RCCs (12). From the fraction of patients whose tumors had the AA-mutational signature and the distribution of AL-adducts in our cohort, we estimate the AA plays a role in the etiology of at least 30% of Taiwanese ccRCC.

While the AA mutational signature clearly documents exposure and a mutational impact, we cannot rule out that AA could promote tumorigenesis through genotoxic-independent mechanisms. Indeed, AA is a potent human nephrotoxin. Although subclinical renal effects of AA have not been investigated in the general population, a minority of people, estimated to be ~5%, are highly sensitive to the nephrotoxic effects of AA (31). Cellular necrosis induced by AA, followed by tissue regeneration, may provide a pro-mitogenic environment for proliferation of tumor initiating cells.

**Potential AA-mutational signature in other RCC sequences**

Recent whole exome sequencing studies of RCC have described subsets of tumors with a high incidence of A-to-T transversions. One study determined the sequences in a small cohort of Chinese RCC tumors (20). In this analysis one of ten tumors possessed an elevated frequency of A-to-T mutations. As in Taiwan, the use of \textit{Aristolochia} herbal remedies is widespread in China and many cases of aristolochic acid nephropathy have been reported there (32).
In other studies, DNA was obtained from RCC cases in England, Russia, the Czech Republic, and Romania (21) and also from the endemic region of Croatia (22). The mutation spectrum in RCCs from England, Russia, and the Czech Republic resembled that seen in the TCGA RCC study (drawn from an American population); however, the AA-mutational signature predominated a majority of the Romanian RCCs (12/14) and 4/8 Croatian RCCs. Although Romania has long been known to harbor regions of endemic nephropathy, the RCC patients whose tumors were sequenced apparently did not reside in the endemic areas. However, the therapeutic use of Aristolochia herbs is known to occur in Romania (33) and a follow-up study did establish the presence of AL-DNA adducts in Romanian RCC patients (23).

The fraction of individuals with AL-DNA adducts is similar among Taiwanese patients with UTUC and RCC(12). Although our case-control study showed that patients who took a cumulative dosage of more than 250 mg of AA had an increased RCC risk (crude OR = 1.4); a similar study reported a higher risk of urothelial carcinoma (bladder cancer and/or UTUC) among Taiwanese ingesting similar amounts of AA (crude OR = 1.9) (24). Also, the number of A-to-T transversions per exome are higher in UTUC with the AA-mutational signature than AA-related RCC (median = 188 per exome in UTUC and 46 per exome in RCC). While these results indicate that renal tissue is sensitive to the carcinogenic effects of AA, they also suggest that the urothelium may be more sensitive to these effects. An alternative explanation is that the nephrotoxic effects of AA dominate renal tumorigenesis and thus AA-induced RCC develops
only in the minority of people sensitive to the nephrotoxic effects of AA.

Regardless of the precise mechanism, the present study offers compelling
evidence implicating AA in a significant fraction of the RCC arising in Taiwan and
illustrates the power of integrating epidemiological, molecular and genetic data in
the investigation of cancer etiology.

**ACKNOWLEDGEMENTS** Chi-Jen Chen, and Yao-Hsu Yang from Department of
Urology, National Taiwan University College of Medicine and Hospital, Taipei, Taiwan
helped perform statistical analysis of the Taiwan National Health Insurance database.
REFERENCES


Table 1. Population-based case-control study of consumption of herbal products containing aristolochic acid and ccRCC incidence in Taiwan.

<table>
<thead>
<tr>
<th>Chinese herbal products containing aristolochic acid</th>
<th>Cases (n = 5709)</th>
<th>Controls (n = 22836)</th>
<th>Crude OR</th>
<th>95% CI</th>
<th>Adjusted OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3520</td>
<td>14281</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2189</td>
<td>8555</td>
<td>1.038</td>
<td>0.978 to 1.102</td>
<td>0.977</td>
<td>0.918 to 1.04</td>
</tr>
<tr>
<td>Cumulative aristolochic acid, mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3520</td>
<td>14281</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1-125</td>
<td>1937</td>
<td>7681</td>
<td>1.023</td>
<td>0.962 to 1.089</td>
<td>0.967</td>
<td>0.907 to 1.032</td>
</tr>
<tr>
<td>126-250</td>
<td>134</td>
<td>527</td>
<td>1.032</td>
<td>0.85 to 1.251</td>
<td>0.936</td>
<td>0.769 to 1.14</td>
</tr>
<tr>
<td>&gt;250</td>
<td>118</td>
<td>347</td>
<td>1.38</td>
<td>1.116 to 1.706</td>
<td>1.246</td>
<td>1.004 to 1.547</td>
</tr>
<tr>
<td>Each 100 mg increase</td>
<td></td>
<td></td>
<td>1.04</td>
<td>1.016 to 1.066</td>
<td>1.029</td>
<td>1.005 to 1.054</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; OR, odds ratio.
*Adjusted for monthly income, urbanization level, hypertension, diabetes, hyperlipidemia, chronic obstructive pulmonary disease, chronic hepatitis C infection, chronic kidney disease, cystic kidney disease, kidney stones, sickle cell disease, aspirin, NSAIDs, and acetaminophen.
Table 2. Characteristics of Taiwanese clear cell renal carcinomas

<table>
<thead>
<tr>
<th>Sample</th>
<th>Patient Age (years)</th>
<th>Gender</th>
<th>Arristolacta m-I-dA Adducts per 10^8 nts</th>
<th>Neoplastic Content</th>
<th>Number of Somatic Mutations</th>
<th>VHL Mutation Status</th>
<th>Other Driver Alterations at 3p21-25</th>
<th>Aristolochic Acid Mutational Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCC8</td>
<td>44</td>
<td>M</td>
<td>21.2</td>
<td>78%</td>
<td>323</td>
<td>11</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>RCC48</td>
<td>60</td>
<td>F</td>
<td>137</td>
<td>52%</td>
<td>225</td>
<td>14</td>
<td>164Q×X</td>
<td>POLE</td>
</tr>
<tr>
<td>RCC79</td>
<td>43</td>
<td>F</td>
<td>49.9</td>
<td>32%</td>
<td>18</td>
<td>0</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>RCC90</td>
<td>64</td>
<td>M</td>
<td>41</td>
<td>44%</td>
<td>124</td>
<td>8</td>
<td>chr3:10191543-10191551delACATC</td>
<td></td>
</tr>
<tr>
<td>RCC88</td>
<td>87</td>
<td>F</td>
<td>6.9</td>
<td>42%</td>
<td>84</td>
<td>4</td>
<td>chr3:10191551delAGATC GTCA^2</td>
<td></td>
</tr>
<tr>
<td>RCC59</td>
<td>106.1</td>
<td>M</td>
<td>45%</td>
<td>70</td>
<td>3</td>
<td>chr3:1018204delA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCC60</td>
<td>46</td>
<td>M</td>
<td>38</td>
<td>26%</td>
<td>24</td>
<td>2</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>RCC67</td>
<td>50</td>
<td>F</td>
<td>27.5</td>
<td>34%</td>
<td>34</td>
<td>8</td>
<td>chr3:1018196delA^2</td>
<td></td>
</tr>
<tr>
<td>RCC15</td>
<td>84</td>
<td>F</td>
<td>2.3</td>
<td>62%</td>
<td>8</td>
<td>7</td>
<td>chr3:101547delA^2+T</td>
<td></td>
</tr>
<tr>
<td>RCC68</td>
<td>54</td>
<td>F</td>
<td>44.7</td>
<td>29%</td>
<td>123</td>
<td>5</td>
<td>nd</td>
<td>nd</td>
</tr>
</tbody>
</table>

nts = nucleotides
nd = not detected

^1Estimated from exome sequencing data using (Distinct Mutation Count/Distinct Coverage)*2*100 averaged across all genes.

^2Identified from Sanger sequencing data.
MAIN FIGURE LEGENDS

Figure 1. Direct measurement of aristolactam (AL) DNA adducts in kidney DNA of Taiwanese ccRCC patients. Top, Diagram of metabolic activation of AA by nitroreductase (NR) and sulfotransferase (SULT). Aristolactam nitrenium ion intermediates react with DNA to generate AL-DNA adducts at deoxyadenosine (dA) or deoxyguanine (dG). Bottom, Histogram of amount of dA-AL-I adduct (7-(deoxyadenosin-N6-yl) aristolactam I) determined by UPLC-ESI/MS3 analysis of nucleotide digestions of non-tumor kidney DNA. White bars indicate all patients in the cohort (n = 51 patients) and black bars indicate the ten patients with tumors sequenced in this study.

Figure 2. Somatic mutational burden is elevated in AA-exposed ccRCCs. A, Scatter plot of the number of somatic substitutions in each ccRCC identified from whole-exome sequencing of AA-exposed Taiwanese (n=10 tumors) and the publically available The Cancer Genome Atlas (TCGA) dataset (n=417 tumors). Red lines indicate median with interquartile range. B, The distribution of mutation consequences of AA-exposed ccRCC (black bars, n=1204 mutations) and TCGA ccRCC (white bars, n=26193 mutations).

Figure 3. Subset of AA-exposed ccRCCs have elevated levels of A:T>T:A transversions. A, Scatter plot of fractions of A:T>T:A transversions, based on the six substitution types of 10 AA-exposed Taiwanese ccRCCs, compared to
417 TCGA ccRCCs. Red line for AA-exposed ccRCC is 33±23% (mean±s.d.) and, for TCGA ccRCC is 11±4.4%. B, Stacked bar chart with frequencies of each of the six substitution types in 10 AA-exposed Taiwanese ccRCC indicated on the x-axis. Tumors arranged by decreasing fraction of A-to-T transversions from left to right. Elevated level of A-to-T determined as greater than two standard deviations above A-to-T average of TCGA ccRCC in A (>20%). AA-exposed ccRCC with “high” and “low” A-to-T fractions indicated. C, Pie graphs showing mutational spectrum of AA-exposed Taiwanese ccRCCs with high (left, n=843 mutations) and low (middle, n=298 mutations), A-to-T fractions compared to TCGA ccRCC (right, n=24559). Legend for substitution types is indicated in B.

Figure 4. Aristolochic acid mutagenesis patterns in Taiwanese ccRCC tumors with greater than 20% A-to-T. A, Percentages of the six classes of substitutions indicated on x-axis for AA-exposed ccRCCs (left side, 6 tumors, n=843 mutations) compared to TCGA ccRCCs (right side, 417 tumors, n=24559 mutations). Each substitution type subdivided by non-transcribed strand (black bar) or transcribed strand (white bar). B, Percentages of A>T (same as T>A) mutations within each of the indicated trinucleotide sequences in AA-exposed ccRCCs (black bars, n = 488 A>Ts) and TCGA ccRCC dataset (white bars, n = 2575 A>Ts). The middle A (underlined) is the mutated base (A>T) in each trinucleotide sequence shown on the x axis. Note: A>T in CpApG is the same as T>A in GpTpC and is accounted for in these data. C, Bar graph showing frequency of mutations in 5’ splice donors (black bars) or 3’ splice acceptor
(white bars) in AA-exposed ccRCC compared to TCGA ccRCCs. Mutations were counted only in the canonical 5’ splice donor (GT, IVS+1 and IVS+2) and 3’ splice acceptor (AT, IVS-1, IVS-2) on the non-transcribed strand.
Figure 1

Aristolochic acid (AA)

AA-I, R = OCH₃
AA-II, R = H

Aristolactam (AL) nitrenium ion

dA-AL-I, R = OCH₃
dA-AL-II, R = H
dG-AL-I, R = OCH₃
dG-AL-II, R = H

AL-DNA adducts

Number of Taiwanese ccRCC Patients

<table>
<thead>
<tr>
<th>AL-I-dA adducts per 10⁸ bases</th>
<th>Entire cohort</th>
<th>Sequenced cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below LOQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>5 - 25</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>25 - 50</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 2

A

Number of Substitution Mutations

AA-exposed ccRCC  TCGA ccRCC

B

Mutation Consequence %

AA-exposed ccRCC  TCGA ccRCC

Missense  Frameshift  Nonsense  Splice site  In-frame indel  Nonstop  Synonymous

Nonsynonymous
Figure 3

A

% of A:T>T:A transversions out of Substitutions

AA-exposed ccRCC  TCGA ccRCC

B

% of Substitutions

AA-exposed ccRCC with high A:T>T:A  AA-exposed ccRCC with low A:T>T:A

C

AA-exposed ccRCC with high A:T>T:A  n = 6 tumors

AA-exposed ccRCC with low A:T>T:A  n = 4 tumors

TCGA ccRCC  n = 417 tumors