Low-Coverage Exome Sequencing Screen in Formalin-Fixed Paraffin-Embedded Tumors Reveals Evidence of Exposure to Carcinogenic Aristolochic Acid

Xavier Castells1, Sandra Karanović2, Maude Ardin1, Karla Tomić3, Evangelos Xylinas4, Geoffroy Durand5, Stephanie Villar1, Nathalie Forey6, Florence Le Calvez-Kelm7, Catherine Voegelin5, Krešimir Karlović2, Maja Mišić2, Damir Dittrich5, Igor Dolgalev5, James McKay5, Shahrkh F. Shariat4, Viktoria S. Sidorenko7, Andrea Fernandes7, Adriana Heguy5, Kathleen G. Dickman7,8, Magali Olivier1, Arthur P. Grollman7,8, Bojan Jelaković2, and Jiri Zavadil1

Abstract

Background: Dietary exposure to cytotoxic and carcinogenic aristolochic acid (AA) causes severe nephropathy typically associated with urologic cancers. Monitoring of AA exposure uses biomarkers such as aristolactam-DNA adducts, detected by mass spectrometry in the kidney cortex, or the somatic A>T transversion pattern characteristic of exposure to AA, as revealed by previous DNA-sequencing studies using fresh-frozen tumors.

Methods: Here, we report a low-coverage whole-exome sequencing method (LC-WES) optimized for multisample detection of the AA mutational signature, and demonstrate its utility in 17 formalin-fixed paraffin-embedded urothelial tumors obtained from 15 patients with endemic nephropathy, an environmental form of AA nephropathy.

Results: LC-WES identified the AA signature, alongside signatures of age and APOBEC enzyme activity, in 15 samples sequenced at the average per-base coverage of approximately 10×. Analysis at 3 to 9× coverage revealed the signature in 91% of the positive samples. The exome-wide distribution of the predominant A>T transversions exhibited a stochastic pattern, whereas 83 cancer driver genes were enriched for recurrent non-synonymous A>T mutations. In two patients, pairs of tumors from different parts of the urinary tract, including the bladder, harbored overlapping mutation patterns, suggesting tumor dissemination via cell seeding.

Conclusions: LC-WES analysis of archived tumor tissues is a reliable method applicable to investigations of both the exposure to AA and its biologic effects in human carcinomas.

Impact: By detecting cancers associated with AA exposure in high-risk populations, LC-WES can support future molecular epidemiology studies and provide evidence-base for relevant preventive measures. Cancer Epidemiol Biomarkers Prev; 24(12); 1–9. ©2015 AACR.

Introduction

The International Agency for Research on Cancer (IARC) classified aristolochic acid (AA) as a group 1 carcinogen (1). Exposure to AA, following intake of Aristolochia herbaceous plants as traditional medicines or due to consumption of bread from flour contaminated by Aristolochia seeds, can lead to AA nephropathy (AAN). AAN is a progressive tubulointerstitial nephropathy with high risk of developing upper tract urothelial carcinoma (UTUC), refs. 2–5. In addition, recent studies proposed AA as a factor contributing to the development of hepatocellular (6–8), renal cell (9, 10) and urinary bladder carcinomas (11), and intrahepatic cholangiocarcinoma (12). Given this growing spectrum of AA-associated tumor types, AA exposure detection methods for screening of disease-risk populations are of key importance.

Following metabolic activation of AA, aristolactam (AL)-DNA adducts accumulate in the proximal tubules of the renal cortex and can be measured by32P-postlabeling (14, 17) or by ultra-performance-liquid chromatography-electrospray ionization-multistage scan mass spectrometry (UPLC-ESI-MS/MS), both
applicable to formalin-fixed paraffin-embedded (FFPE) tissues (16, 18, 19). However, the $^{32}$P-postlabeling method lacks specificity, and access to the UPLC-ESI-MS/MS methodology and its optimization for biomaterial of low quantity are limiting factors.

DNA sequencing established a characteristic AA mutational signature marked by accumulation of A$^\circ$T transversions within the 5'-Pyr-A-Pur-3' sequence context (enriched for 5'-GpaG-3'), preferentially located on the nontranscribed strand (8–10, 20, 21). In cancers not associated with AA, such A$^\circ$T transversions are infrequent (22, 23).

We exploited the unique features of the AA mutational signature to devise a sensitive method for AA exposure detection, based on low-coverage whole-exome sequencing (LC-WES, at approximately 10$^{-5}$ in contrast with the conventional 100$^{-5}$ coverage), optimized for analysis of tumor-specific DNA of limited quantity and integrity extracted from archived FFPE tissues. The studied urothelial tumor samples originated from a well-characterized population residing in the endemic nephropathy (EN) regions of Croatia and Bosnia and Herzegovina (13), with EN being thus far the only recognized environmental form of AAN (4, 24). For the first time, we report in the urothelial tumors of EN patients the genome-wide signatures of AA, age, and APOBEC cytidine deaminase activity, thereby extending previous mutational analyses of this population based solely on the mutations of the TP53 tumor-suppressor gene (4, 13, 25). In addition, we demonstrate the ability of LC-WES to elucidate the impact of the AA mutation spectra on key homeostatic biologic pathways and to reveal possible mechanisms of tumor dissemination along the urinary tract.

Materials and Methods

Patients and tumor samples

Exposure to AA was investigated in 15 patients with urothelial tumors, diagnosed with EN following established criteria (13, 26). As controls, UTUC samples were obtained from 4 patients from a metropolitan area of the United States, unlikely exposed to AA. All specimens were FFPE-converted in the histopathologic laboratories of the participating centers. The involved anatomical sites were renal pelvis, ureter, and bladder (ICD-10 codes C65, C66, and C67, respectively). Clinicopathologic features and Aristolochia exposure history are listed in Supplementary Table S1. The study protocols included patients' informed consent and were approved by the IARC Ethics Committee and the Institutional Review Boards of the participating institutions.

DNA isolation from paraffin sections

Hematoxylin and eosin preparations of the paraffin block sections were used to identify tumor tissue free of necrotic areas. The tumor cell areas were measured by ImageJ software (27). Ten micromolar sections, cut with the Leica RM 2145 microtome (Leica Microsystems), were used to macrodissect the tumor-enriched areas and isolate genomic DNA yielding 1 to 2 ng/mm$^2$ per sample. Before DNA isolation, slides were deparaffinized for 5 minutes in 100% xylene, kept for 5 minutes in absolute ethanol, 5 minutes in 85% ethanol, 5 minutes in 75% ethanol and stored in milliQ water. DNA isolation was done using the QIAamp DNA FFPE Tissue Kit (Qiagen) following the manufacturer's protocol. DNA yields and concentrations were measured using the Picogreen assay (LifeTechnologies) and Fluoroskan Ascent FL microplate fluorometer (Thermo Fisher Scientific). DNA purity was evaluated by the NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific), and DNA integrity assessed by 0.8% agarose gel electrophoresis.

AL-DNA adduct analysis and TP53 resequencing

DNA was isolated from the renal cortex and tumor tissues by standard phenol-chloroform extraction techniques. The level of AL-DNA adducts in the renal cortex DNA (10–20 μg) was determined using $^{32}$P-postlabeling PAGE, as previously described (13). The TP53-specific mutations were identified using the AmpliChip p53 Research Test (Roche Molecular Diagnostics), sensitively detecting all single base-pair substitutions and single-base deletions (13).

WES library preparation, exome capture, and sequencing

Two hundred and fifty (250) ng of genomic DNA was sheared by the adaptive focused acoustics method (Covaris, Inc.) to obtain approximately 300 bp fragments, with water temperature of 4°C, one cycle at 175 Watt peak power, duty factor 10 and 200 cycles per burst. Resulting fragment size was assessed using the 2100 Bioanalyzer and High Sensitivity DNA Kit (Agilent Technologies). The sheared DNA was converted into libraries using the Kapa HT Library Preparation Kit (Kapa Biosystems). Briefly, the fragmented DNA was subjected to end repair reaction followed by poly-A-tailing and adapter ligation, excess adapters removed by Agen- court AMPure XP beads (Beckman Coulter). Eight cycles of PCR were performed to amplify the libraries with correct adapters on both ends. Four libraries (250 ng each) were pooled per exome capture with the Nimblegen SeqCap EZ Exome reagent. Exome-enriched mixes were PCR-amplified in 10 cycles, post-enrichment libraries pooled in 420 μL of water to a final concentration of 6 pmol/L. This volume was divided and loaded in two lanes of the rapid run mode flow cell for cluster generation and sequencing on the HiSeq2500, in a paired-end 50 bp cycle run. Multiplexing 16 samples per run resulted in the target coverage of approximately 10$^{-5}$.

Four additional EN UTUC samples and two UTUC samples from the metropolitan United States were analyzed in a validation assay using the SOLiD 5500XL sequencer (LifeTechnologies). See Supplementary Methods for details. Raw HiSeq2500 sequencing data were deposited to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) repository (ID SRP042035) to become available from the NCBI’s dbGaP database. The annotated list of single-base substitutions (HiSeq2500 data) is provided in Supplementary Table S2.

Sequencing data analysis

FastQ reads were aligned to the human genome (hg19) using Burrows-Wheeler Aligner. Realignment and base quality score recalibration was done by the Genome Analysis Toolkit (GATK) and the duplicate-read removal by Picard. GATK HaplotypeCaller was used to call variants subsequently annotated on the RefSeq Gene transcript contents by ANNOVAR (28). Polymorphisms present in normal population and removed from our data originated from these collections: 1,000 genomes (1,000 g, http://www.1000genomes.org/), Exome Sequencing Project (ESP, http://exome.gs.washington.edu/) and the SNP database build 137 (dbSNP, http://www.ncbi.nlm.nih.gov/SNP/). We removed variants with frequency above 0.1% in either the 1,000 g or ESP databases, or annotated in the dbSNP database, or present in a
custom germline variant catalog built from 560 cases from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/). Variants mapping to repetitive sequences contained in the genomic segmental duplication database (29) alongside variants with ≥90% homology with multiple regions were excluded. R functions were developed to compute the mutation type distributions and strand bias. The strand bias significance was determined by the Pearson \( \chi^2 \) test. These tertiary analysis parameters were computed in two separate coverage ranges, ≥3x with no defined maximum, and between 3 and 9x to emulate ultra-low coverage.

Mutational signature analysis using nonnegative matrix factorization
Nonnegative matrix factorization (NMF) decomposes mutational patterns based on factorization of one matrix \((n \times m)\) in two matrices \(W(n \times r)\) and \(H(r \times m)\) with the constraint that all three matrices must be composed of nonnegative elements (30). The \( r \) is the rank of factors to be extracted from the input matrix, corresponding to the number of signatures. The input matrix contained one column per patient (only HiSeq2500 data considered) and in rows the frequency of mutations types in 96 possible two-base sequence contexts. The R package NMF (31) was used to extract mutational signatures. The correlation between the extracted signatures and previously published ones (7, 21, 22, 32) and/or available in COSMIC (23) was computed as the inner product of the two signatures (vectors) divided by the product of their norms.

Functional analysis of tumor-specific nonsynonymous mutations
To examine the biologic impact of the gene mutants in the AA signature-positive samples, analysis was performed using the DAVID tool (33), with two input gene lists: (i) genes harboring nonsynonymous (missense, stop-gain, or stop-loss) SBS and (ii) genes nonsynonymously mutated in the EN dataset and in AA signature-positive samples from at least one of the two published datasets on UTUC in Taiwanese patients (8, 21). The list was further narrowed by classifying the mutated genes as established oncogenes or tumor suppressors listed by the Gene Set Enrichment Analysis (GSEA) database (34) and/or a cancer driver genes defined by recent seminal studies (35–39).

Results and Discussion
Low-coverage detection of AA exposure signature in urothelial tumors of EN patients
We applied HiSeq2500 LC-WES to genomic DNAs isolated from FFPE urothelial tumors from 11 EN cases, two of whom had concurrent UTUC and bladder carcinoma, and from two U.S. patients providing non-EN control samples (13 patients and 15 tumors in total, see Supplementary Table S1). Features of AA signature had been described earlier, as follows: mutational load of ≥40 SBS or ≥10 A>T in exonic positions, high proportion of A>T (>35% of all SBS types or as the predominant type) with a strand bias of ≥1.25, and ≥33% enrichment of A>T in the 5’-C/T pApG-3’ sequence context (8, 21). We used analogous criteria for the AA signature (≥50 SBS/sample of which ≥15% are A>T SBS, of which ≥20% are in the 5’-CpApG-3’ context), applying more stringent statistical analysis of the strand bias ratio combined with a cutoff of ≥1.5 (9, 11). Under these criteria the AA signature was readily observed in 10 of the 13 analyzed EN tumor samples, with 33% to 77% of A>T transversions per sample (Fig. 1), a nontranscribed strand bias of 2.0 to 3.3 and the 5’-C-G-3’ context enrichment above 19% (mean 24.6%, SD = 4.9, range of 19.1%–27.4%; Table 1). In contrast, A>T mutations and their enrichment in the 5’-CpApG-3’ context are generally low in cancers of non-AA etiology, based on our analysis of 7,160 tumors of 52 cancer types in the COSMIC database (average 5.8% A>T, range 0%–12.1%, of which 10% are in the 5’-CpApG-3’ context). Similarly, the average percentage of A>T in the 5’-CpApG-3’ context in TCGA urothelial carcinoma data (only bladder data available) is 10.8% (0%–50%), whereas the mean percentage of all A>T mutations is low (average of 3.9%, range 0.8%–8.3%).

A weaker signature marked by 18.7% A>T, strand bias of 2.1 and the 5’-CpApG-3’ context proportion of 12.5% was observed in the bladder tumor sample (EN-01-B) of a patient with a concurrent AA signature-positive UTUC (EN-01-RP, see Table 1 and Supplementary Table S1). Two EN samples (EN-06 and EN-07) and the two non-EN controls were found negative for the AA signature, with A>T transversions present at 4% to 8%. In the case of EN-07 (bladder carcinoma with no history of UTUC), despite the presence of AL-DNA adducts in the patient’s renal cortex, the mutational profile (Fig. 1) suggested AA-unrelated etiology. Among the AA signature-positive samples, we detected an average of 1,142 (range 349–2,707) mutations per tumor (~18 SBS/sample per exome megabase (Mb)) whereas the mutation rate in the control and negative samples (including the weaker AA-signature bladder cancer) was on average 357 (range 258–440) mutations per sample (~6 per sample/exome Mb). As shown in Table 1, the predominant A>T transversions substantially contributed to the high SBS counts.

Thus, LC-WES analysis of the EN UTUC generates results consistent with previous reports on the highly mutagenic potential of AA (8, 21, 40), and our results justify the use of exome sequencing for reliable detection of exposure to AA in archived FFPE material.

LC-WES identifies AA signature at ultra-low coverage
We next investigated whether the AA signature can be identified at ultra-low coverage. Upon considering 3 to 9 non-duplicate per-base reads, mutation counts in the AA-associated samples decreased to 233 per tumor on average (~4 per sample/exome Mb) and to 67 per tumor (1/sample/exome Mb) in the negative samples and the weakly positive bladder tumor (EN-01-B). The 10 tumors shown in Fig. 1 (two top rows) still exhibited the AA signature at ultra-low coverage (Supplementary Fig. S1), with the strand bias ratios between 1.7 and 4.7, and retained prominent enrichment of the 5’-CpApG-3’ context (>25%). Thus, the specific and unique features of the AA signature can be reliably detected in FFPE tumor samples by superficial coverage sequencing.

These results open an attractive opportunity for retrospective analyses of archived pathologic specimens from the regions of AA exposure risk. In comparison with the 32P-postlabeling and mass spectrometry adduct detection techniques, the LC-WES approach is based on a commodity technology that generates genome-wide information. LC-WES is also very sensitive, using low input DNA amounts (250 ng compared with 5–10 µg required for adduct analysis). Finally, it can indicate exposure to AA when neither AL-DNA adducts nor mutations in TP53 are detected, as we demonstrate for the AA signature-positive cases EN-01, EN-03, EN-04, and EN-11 (Fig. 1; Supplementary Table S1).
AA-associated urothelial tumors harbor three major mutational signatures

NMF extracts individual mutational signatures from complex alteration patterns observed in primary tumors, reflecting thus the specific effects of etiologic factors (7, 9, 22, 41). NMF was used to describe the AA signature in human UTUC, bladder, liver, and renal carcinomas (7, 9, 11) and in experimental *in vitro* system designed to model mutational signatures of carcinogens (32). Here, in the EN urothelial tumors, the NMF approach identified three distinct signatures, the AA-specific signature (Signature 22; ref. 42), the signature related to age (C:G>T:A in the 5’-XpCpG-3’ context, Signature 1; ref. 22), and the Signatures 2 and/or 13 associated with the cytidine deaminase activity of the APOBEC enzymes (Fig. 2; ref. 22). All three signatures are currently listed in the COSMIC database (23). Furthermore, NMF aided in classifying the EN-01-B bladder tumor as positive due to nonnegligible sample contribution to the AA signature (18%), in contrast with the negative samples (EN-06-RP and EN-07-B, with contributions of 0% for both) and non-EN controls (each 0% contribution, see Table 1 and Fig. 2B). The identified EN UTUC AA signature

Figure 1.
SBS alterations in urothelial tumors analyzed by LC-WES. The distribution of six SBS types and their trinucleotide context are shown for variants detected at ≥3× per-base coverage. The doughnut charts correspond to individual samples (sample ID on top), ordered from high to low percentage of A>T. Total SBS counts per sample are provided in the center of each graph. The numbers outside the chart sections denote each mutation type percentage. The suffixes -B, -RP, and -U stand for bladder, renal pelvis, and ureter, respectively. Add+: sample positive or negative for aristolactam-DNA adducts; TP53+/-, mutated (+) or wild-type (-) TP53 gene. The heatmaps summarize relative frequencies of the six mutation types (C>A stands for C:G>A:T, etc.) across the 16 possible trinucleotide contexts listed at the bottom. Red, high frequency; yellow, low frequency.
correlated highly (>90%) with the COSMIC Signature 22 (23), derived from AA-associated primary UTUC tumors from Taiwanese patients (8, 21), and with the AA signature modeled in vitro (Fig. 2C; ref. 32). The other EN tumor signatures matched their COSMIC counterparts with 72% similarity (age) and 70% and 64% similarity (APOBEC, Signature 2 and 13, respectively; Fig. 2C).

Validation of the LC-WES performance on a distinct sequencing platform

To validate the LC-WES performance using another sequencing chemistry and platform, we analyzed four additional EN UTUCs positive for AL-DNA adducts and p53 A>T mutations (EN-12, EN-13, EN-14, and EN-15), and two control UTUCs from U.S. patients (Non–EN-03 and Non–EN-04), on the SOLiD 5500xl sequencer. At the average 14.5× coverage, we observed the AA signature in all EN samples, although in samples EN-13-RP and EN-14-RP, the A>T transversion was the second most abundant mutation type following C>T (see Supplementary Fig. S2A). The signature remained detectable at ultra-low coverage (~4.6×), when considering only the 3 to 9 read interval (Supplementary Fig. S2B).

Chromosomal distribution of the AA-specific mutations and recurrently mutated cancer driver genes

In the A>T enriched samples, A>T transversions were randomly distributed along the sequenced regions, with linear correlation between A>T SBS counts and chromosome size ($R^2 = 0.9$; Fig. 3A). Similar correlation was maintained in the minimum coverage interval of 3 to 9× (data not shown). This result was confirmed by the analysis of the Taiwan UTUCs (8, 21) in which a similar, although less linear trend was observed ($R^2 = 0.61$–0.64). These findings suggest a stochastic A>T mutation distribution within the gene/transcription units represented by the exome.

Despite this apparently random pattern, we identified 83 cancer driver genes carrying protein sequence-altering A>T SBSs, that were recurrently mutated across the three datasets of the AA signature–positive tumor samples (this study, $n = 10$, and the two previously reported Taiwanese sets of $n = 18$ (21) and $n = 9$ (8)). These findings are summarized in Fig. 3B and in Supplementary Table S4. The recurrently mutated genes included numerous known drivers and chromatin-associated factors such as TP53, ARID1B, ATRX, CREBBP, CHD2, CHD5, CHD8, FAT1, KDM6A, ML22 (KMT2D), SETBP1, TRRAP. TP53 was the most frequently mutated gene [17/37 (46%) samples] with all its mutations being A>T transversions. Fifteen samples exhibited mutations in the histone methyl-transferase KMT2D (MLL2), with varying SBS types, suggesting that secondary mutation processes possibly linked due to high mutational loads and increased genomic instability. Further systematic investigations should be undertaken to establish possible recurrent alterations in particular genes and pathways in UTUC across studies of different populations/geographical areas. For instance, data in Fig. 3 and in Supplementary Table S3 indicate that TP53, CREBBP, and LRRK2 are mutated mostly in the Taiwanese samples whereas mutations in the AHNAK, ATRX, SMCHD1, and XIRP2 genes are enriched in the EN UTUC samples. Other factors contributing to these differences merit further investigations, including varying modes of AA exposure (low-dose chronic intake in the EN regions as compared with higher-dose, (sub)acute exposures resulting from the use of traditional herbal medicines in Asia) and disease susceptibility due to the patients’ genetic background.

Biologic impact of the AA-signature

Using NIH DAVID, we performed Gene Ontology (GO) and KEGG pathway analyses of the genes harboring nonsynonymous A>T mutations in the AA-signature–positive samples analyzed by HiSeq2500 ($n = 10$). We identified gene targets from the

Table 1. Summary of the SBS detected at ≥ 3× per-base coverage and of the AA-signature analysis results

<table>
<thead>
<tr>
<th>Case ID</th>
<th>Total SBS</th>
<th>SBS per Mbp</th>
<th>A &gt; T SBS per Mbp</th>
<th>A &gt; T (%)</th>
<th>A-T context (%)</th>
<th>CAG context (%)</th>
<th>A &gt; T ratio (NTR/Tr)</th>
<th>SB A &gt; T (%)</th>
<th>SB P value</th>
<th>SB FDR Q value</th>
<th>Contribution to Sig 22 (AA)</th>
<th>AA signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN-01-RP</td>
<td>608</td>
<td>9.1</td>
<td>376</td>
<td>5.6</td>
<td>61.8</td>
<td>26.1</td>
<td>3.3 (272/82)</td>
<td>0.0</td>
<td>380</td>
<td>63%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-01-B</td>
<td>295</td>
<td>4.5</td>
<td>56</td>
<td>0.8</td>
<td>18.7</td>
<td>12.5</td>
<td>2.1 (37/18)</td>
<td>0.015</td>
<td>1</td>
<td>53%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-02-RP</td>
<td>984</td>
<td>14.7</td>
<td>585</td>
<td>8.7</td>
<td>59.5</td>
<td>26.7</td>
<td>3.1 (424/136)</td>
<td>0.0</td>
<td>514</td>
<td>52%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-02-U</td>
<td>889</td>
<td>13.3</td>
<td>529</td>
<td>7.9</td>
<td>59.5</td>
<td>26.1</td>
<td>3.3 (384/118)</td>
<td>0.0</td>
<td>457</td>
<td>51%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-03-RP</td>
<td>807</td>
<td>12.0</td>
<td>443</td>
<td>6.6</td>
<td>54.9</td>
<td>25.0</td>
<td>2.0 (278/140)</td>
<td>0</td>
<td>1</td>
<td>514</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-04-RP</td>
<td>639</td>
<td>9.5</td>
<td>241</td>
<td>3.6</td>
<td>57.7</td>
<td>19.5</td>
<td>2.5 (157/68)</td>
<td>4.4E-09</td>
<td>618</td>
<td>40%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-05-RP</td>
<td>549</td>
<td>9.2</td>
<td>115</td>
<td>1.7</td>
<td>33.0</td>
<td>19.1</td>
<td>2.8 (79/28)</td>
<td>1.3E-06</td>
<td>380</td>
<td>63%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-06-RP</td>
<td>352</td>
<td>5.3</td>
<td>15</td>
<td>0.2</td>
<td>4.3</td>
<td>0.0</td>
<td>5.5 (11/2)</td>
<td>0.027</td>
<td>1</td>
<td>0%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>EN-07-B</td>
<td>440</td>
<td>6.6</td>
<td>35</td>
<td>0.5</td>
<td>8.0</td>
<td>11.4</td>
<td>1.2 (15/13)</td>
<td>0.850</td>
<td>1</td>
<td>0%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>EN-08-U</td>
<td>1551</td>
<td>23.1</td>
<td>1195</td>
<td>17.8</td>
<td>77.0</td>
<td>23.4</td>
<td>2.8 (840/303)</td>
<td>0</td>
<td>139</td>
<td>90%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-09-U</td>
<td>2221</td>
<td>33.1</td>
<td>1701</td>
<td>25.4</td>
<td>76.6</td>
<td>27.4</td>
<td>2.7 (1885/431)</td>
<td>0</td>
<td>193</td>
<td>87%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-10-U</td>
<td>2707</td>
<td>40.4</td>
<td>1486</td>
<td>22.2</td>
<td>54.9</td>
<td>26.2</td>
<td>2.8 (1052/377)</td>
<td>0</td>
<td>2706</td>
<td>100%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EN-11-U</td>
<td>669</td>
<td>10.0</td>
<td>309</td>
<td>4.6</td>
<td>46.2</td>
<td>26.9</td>
<td>2.5 (210/85)</td>
<td>3.7E-13</td>
<td>297</td>
<td>44%</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Non-EN-01-RP</td>
<td>385</td>
<td>5.7</td>
<td>24</td>
<td>0.4</td>
<td>6.2</td>
<td>0.0</td>
<td>0.6 (9/14)</td>
<td>0.400</td>
<td>1</td>
<td>0%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Non-EN-02-RP</td>
<td>258</td>
<td>3.9</td>
<td>14</td>
<td>0.9</td>
<td>5.4</td>
<td>14.3</td>
<td>5.0 (10/2)</td>
<td>0.043</td>
<td>1</td>
<td>0%</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Suffixes -B, -RP, and -U indicate bladder, renal pelvis, and ureter, respectively. Abbreviations: SBS, single base substitution; Mbp, megabase pair; A > T, A > T transversion(s); SB A > T, strand bias, the ratio of A > T transversions on the non-transcribed versus transcribed strand (the number of respective transversions is shown in brackets); CAG context (%), percentage of A > T transversions in the most frequent context reported for the AA signature; NTR/Tr, ratio of A>T variants on nontranscribed versus Transcribed strand; SB P value and FDR q value, measures of significance of the strand bias, see Materials and Methods. Contribution to Sig 22 (AA) (EN-01, EN-02, EN-04, EN-05, EN-06, EN-07, EN-08, EN-09, EN-10, EN-11, Non-EN-01-RP, Non-EN-02-RP, Non-EN-03-RP, Non-EN-04-RP, Non-EN-05-RP, Non-EN-06-RP, Non-EN-07-RP, Non-EN-08-RP, Non-EN-09-RP, Non-EN-10-RP, Non-EN-11-RP) by the NMF-determined mutational signature shown as the number of SBS and the corresponding percentage (in brackets) of the total SBS per sample. AA signature, positivity for AA signature considering the co-occurrence of ≥ 50 SBS and ≥ 15% A>T, ≥ 20% CAG context, strand bias (SB) and its significance (SB P and/or q value) and mutation load contribution to Sig 22 (AA), as described previously (9).

*Zero values correspond to a P value below 2 × 10^-16.
*Lower percentage in the 5’-CpApG-3’ context and nonsignificant q value; supported by the NMF analysis.

www.aacjournals.org Cancer Epidemiol Biomarkers Prev; 24(12) December 2015 OF5

Published OnlineFirst September 17, 2015; DOI: 10.1158/1055-9965.EPI-15-0553

Downloaded from cebp.aacjournals.org on June 20, 2017. © 2015 American Association for Cancer Research.
Figure 2.
Mutational signatures determined by NMF. Results are shown for urothelial carcinoma samples sequenced on HiSeq2500. A, contribution of each mutation type to signatures of AA, age/CpG, and APOBEC. The x-axis represents the trinucleotide sequence contexts, with the 5'-flank base in the first row and the 3'-flank in the second. B, contributions of the studied urothelial tumors to the individual signatures shown in A. C, correlation between NMF-identified EN UC signatures and previously described COSMIC signatures 1, 2/13, and 22 (22), signature 22 identified in Taiwan UTUC samples (8, 21), and signature 22 AA Exp, modeled experimentally in vitro (32).

Overlapping mutation patterns in distinct tumors from same patients
Two EN patients had synchronous urothelial tumors in distinct anatomical sites (renal pelvis and bladder, samples EN-01-RP and EN-01-B; and renal pelvis and ureter, samples EN-02-RP and EN-02-U). By using LC-WES, we investigated the common genetic origins of these synchronous tumor pairs. In patient EN-01, the overlapping SBS were enriched for C>T mutations (42%) followed by A>G (20%), and only 7.7% of the overlapping SBS were A>T transversions affecting the coding sequence of mere 3 non-cancer genes (VWA3B, KDM3B, and ACIN1). However, the A>T SBS were enriched among the mutations unique to the renal pelvis and to the bladder tumor (77% and 28%, respectively, Supplementary Fig. S3A), suggestive of a common precursor carrying mainly non-A>T driver mutations, giving rise to two tumor progenies subsequently accumulating distinct patterns of A>T alterations in either anatomical site. The distinct AA signature in the bladder tumor is in keeping with a recent study of Asian bladder cancer patients in whose tumors the AA signature manifested without the involvement of upper tract or a history of renal disease (11). In contrast, the tumors in the renal pelvis and ureter of patient EN-02 shared the majority of mutations contributing to a prominent AA signature, suggesting a
common precursor carrying mostly A>T alterations (Supplementary Fig. S3B). This genetic relationship between same-patient tumors suggests cell seeding along the tract as the basis for tumor dissemination. However, further investigations of a larger multiple-tumor case series and with the use of deep sequencing is needed to further elucidate the exact mechanisms of multifocal and recurrent tumorigenesis in the urinary tract of AA-exposed patients.

In summary, we report successful detection of the genome-wide AA signature in urothelial tumors of EN patients, using archived FFPE specimens and a customized low-coverage exome sequencing. The described technique is a cost-effective screening tool potentially applicable to molecular epidemiology studies aiming at identifying cancers associated with AA exposure. This ability of the LC-WES and its applicability to archived biomaterial may be exploited in future systematic studies on AAN and associated cancers, in support of established or future disease prevention programs.

Disclosure of Potential Conflicts of Interest

S.F. Shariat has received honoraria from speakers bureau from Astellas, Takeda, Ipsen, Janssen, Sanoﬁ, Wolff, Pierre Fabre, and Sanochemia, has ownership interest (including patents) in prostate and bladder cancer biomarkers, and is a consultant/advisory board member for Astellas, Ipsen, Sanochemia, Olympus, Wolf, and Janssen. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: A.P. Grollman, B. Jelaković, J. Zavadil
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X. Castells, S. Karanović, K. Tomic, E. Xylinas,
References


Low-Coverage Exome Sequencing Screen in Formalin-Fixed Paraffin-Embedded Tumors Reveals Evidence of Exposure to Carcinogenic Aristolochic Acid

Xavier Castells, Sandra Karanovic, Maude Ardin, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst September 17, 2015.

Updated version  Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-15-0553

Supplementary Material  Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2015/09/17/1055-9965.EPI-15-0553.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.