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 Tomic3, Evangelos Xylinas4, Geoffroy Durand5, Stephanie Villar1, Nathalie Forey6, Florence Le Calvez-Kelm7, Catherine Voegele2, Kresimir Karlovíc8, Maja Mišić9, Damir Dittrich3, Igor Dolgalev6, James McKay5, Shahrokh F. Shariat4, Viktoria S. Sidorenko1, Andrea Fernandes7, Adriana Heguy6, Kathleen G. Dickman7,8, Magali Olivier1, Arthur P. Grollman7,8, Bojan Jelaković3, 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); 1873–81. ©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-mass spectrometry (UPLC-ESI-MS/MS); both
applicable to formalin-fixed paraffin-embedded (FFPE) tissues (16, 18, 19). However, the 32P-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>T transversions within the 5′-Pyr-A-Pur-3′ sequence context (enriched for 5′-CpApG-3′), preferentially located on the nontranscribed strand (8–10, 20, 21). In cancers not associated with AA, such A>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× in contrast with the conventional 100× 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-micromolar sections, cut with the Leica RM 2145 microtome. The tumor cell areas were measured by ImageJ software (27). Ten DNA isolation from paraf...
Nonnegative matrix factorization (NMF) decomposes mutational patterns based on factorization of one matrix ($n \times m$) into 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 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 mutations 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 (≥33% 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)/G(C) pApgG-3′ 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′-CpApgG-3′ context). Similar correlation and 0.5% enrichment of the 5'-CpApgG-3′ context. Thus, the specific and unique features of the AA signature can be reliably detected in FFPE tumor samples by superfamily 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 commoditization 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, 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'-XGpG-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 \( \geq 5 \times \) 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>G 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-IES platform on a distinct sequencing platform

To validate the LC-IES performance using another sequencing chemistry and platform, we analyzed four additional EN UTUCs positive for AL-DNA adducts and p53 A>T mutations (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 5500d 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 Supp Fig. S2A). The signature remained detectable at ultra-low coverage (<4.6×), when considering only the 3 to 9 read interval (Supp 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, MLL2 (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 AHNK, 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
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 experimentally.

Helicase activity, chromatin modification, cell-cycle, cell signaling (MAPKKK/RAS and PI3K cascades), functional classes of cell adhesion, cell-matrix contact, cell migration, cell-cycle, cell signaling (MAPKKK/RAS and PI3K cascades, mTOR pathway), pathways of WNT, insulin and ERBB signaling, nucleotide excision repair, and the DNA-dependent ATPase and helicase activity, chromatin modification and histone binding related to gene-expression regulation, with dozens to hundreds of mutated genes per category (Supplementary Table S3). This observation suggests massive deregulation and/or destabilization of key homeostatic pathways by the high A>T mutation loads.

Next, for the 83 recurrently mutated cancer genes (Fig. 3B; Supplementary Table S4), we observed enrichment of GO and KEGG categories related to regulation of transcription, chromatin/histone modification, and categories of DNA damage response and DNA repair (Supplementary Table S5). These included numerous previously established cancer driver genes (TP53, AHNAC, ARID1B, ATRX, BLM, CHD2, CHD3, CHD8, CHD9, CHEK2, CLTC, ERBB4, FN1, HIUWE1 IARS2, KALRN, LRRK2, ML22, NEB, RXRA, SMCHD1, SPEG, STAG2, SYNE1, TRIO; refs. 35–39). Thus, the LC-WES analysis of AA-exposed urothelial tumors and associated data mining can reveal biologic information contents, particularly upon meta-analysis with data from different populations characterized by identical etiology and tumor types.

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, Olympus, Wolff, Pierre Fabre, and Sanochenia, has ownership interest (including patents) in prostate and bladder cancer biomarkers, and is a consultant/advisory board member for Astellas, Ipsen, Sanochenia, Olympus, Wolff, 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. Tomić, E. Xylinas,

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): X. Castells, M. Ardin, K. Tomić, C. Voegele, I. Dolgalev, J. McKay, S.F. Shariat, M. Olivier, J. Zavadil

Writing, review, and/or revision of the manuscript: X. Castells, S. Karanović, M. Ardin, K. Tomić, E. Xylinas, F. Le Calvez-Kelm, C. Voegele, S.F. Shariat, K.G. Dickman, M. Olivier, A.P. Grollman, B. Jelaković, J. Zavadil

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Castells, S. Karanović, K. Tomić, S. Villar, K. Karlović, D. Dittrich, I. Dolgalev, A. Fernandes, K.G. Dickman, M. Ardin, M. Mišić

Study supervision: B. Jelaković, J. Zavadil

Acknowledgments

The authors thank the Genetic Platform, IARC, and the Genome Technology Center, NYU Langone Medical Center for expert assistance with exome sequencing; the authors thank Christine Carreira, Dr. Behnoush Abedi-Ardekani, and Dr. Elisabeta Kuhn for assistance with the biospecimen evaluation and processing. The authors thank the staff of the General Hospital Slavonski Brod, the University Hospital Centre Zagreb and of the Weill Medical College of Cornell University, New York, for expert assistance and support. The authors are grateful to Drs. Monica Hollstein and Michael Kornjak for critical comments on the article.

Grant Support

IARC Regular Budget; grant P01 ES04068 (to A.P. Grollman) from the NIH/NIEHS; R03-TW007042 grant (to A.P. Grollman) from the NIH Fogarty International Center; grant (to A.P. Grollman) from Henry and Marsha Laufer; grant no. 108-000000-0329 (to B. Jelakovic) from the Croatian Ministry of Science, and grant 04/38 (to B. Jelakovic) from the Croatian Foundation for Science; and grant NIH/NCI P30 CA016087-33, which partially supports the NYU Genome Technology Center.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 18, 2015; revised August 11, 2015; accepted September 8, 2015; published OnlineFirst September 17, 2015.

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 2015;24:1873-1881. 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

Cited articles
This article cites 41 articles, 10 of which you can access for free at:
http://cebp.aacrjournals.org/content/24/12/1873.full#ref-list-1

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.