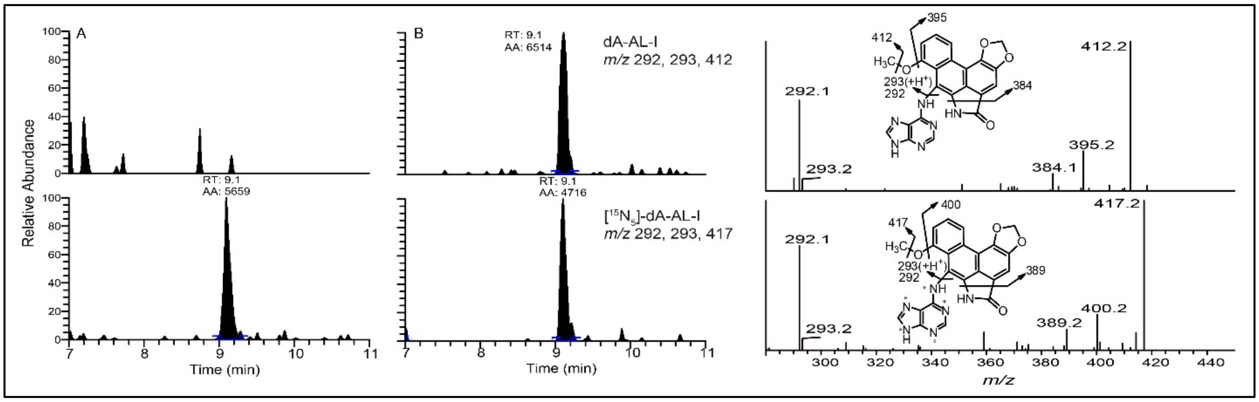
Supplementary figures.



Supplementary Figure 1. Examples of UPLC-ESI/MS^3 reconstructed ion chromatograms of dA-AL-I from human kidney cortex of patients with RCC. (A) Case from Czech with no detectable dA-AL-I; (B) Case from Romania positive for dA-AL-I. The product ion spectra at the MS^3 scan stage are shown for dA-AL-I and [$^{15}N_5$]-dA-AL-I. Asterisks (*) next to nitrogen atoms on adenine based indicate the ^{15}N isotope label. RT and AA on the mass chromatogram represent retention time and integrated area under peak respectively.

Supplementary notes.

R script for the binomial regression analysis

```r
library(xlsx)
##Loading required package: rJava
##Loading required package: xlsxjars

##Import data
d <- read.xlsx('…/<supplementary_table>.xlsx',
     sheetIndex=1,
     startRow=3,
     colClasses=c('character')
     )
#Remove footnote and reformat variables
d <- d[-nrow(d), ]
d[ ,1] <- as.character(d[, 1])
d[ ,2] <- as.character(d[, 2])
d[ ,3] <- as.character(d[, 3])
d[, 'tumor_n.A.T.T.A.'] <- as.numeric(as.character(d[, 'tumor_n.A.T.T.A.']))
d[, 'tumor_proportion.A.T.T.A.']<- as.numeric(as.character(d[, 'tumor_proportion.A.T.T.A.']))

##Change levels below detection to 0
d[which(d[ ,2] == "<0.3"), 2] <- "0"
d[ ,2] <- as.numeric(d[,2])

##Create a short name for the adduct-load column
n <- names(d)
n[2] <- 'dA.AL.I'
names(d) <- n
rm(n)

##Select the phenol-chloroform extracted samples for Romanian samples
d <- d[-which((d[,4] == "Romania") & (d[,3] != "Phenol/Chloroform")), ]
rownames(d) <- d[ ,1]

##Log of number of adducts
d$`log1p(dA.AL.I)` <- log1p(d$dA.AL.I)

##Fit binomial model on proportion, weighted by number of mutations overall
##(weights scaled to sum to the number of observations)
binreg <- glm(tumor_proportion.A.T.T.A. ~ log1p(dA.AL.I),
          weights=nrow(d)*tumor_n.all_mutations./sum(tumor_n.all_mutations.),
          family=binomial("logit"),
          data=d, model=TRUE)
summary(binreg)
```