# Tissue- and sex-specific DNA damage tracks aging in rodents and humans

Axel Guilbaud[1,2], Farzan Ghanegolmohammadi[2Ψ], Yijun Wang[1,2Ψ], Jiapeng Leng[1,2], Alexander Kreymerman[1], Jacqueline Gamboa Varela[1,2], Jessica Garbern[1], Hannah Elwell[1], Fang Cao[1], Elisabeth M. Ricci-Blair[1], Cui Liang[3], Seetharamsingh Balamkundu[3], Charles Vidoudez[4], Michael S. DeMott[2], Kenneth Bedi[5], Kenneth B. Margulies[5], David A. Bennett[6], Abraham A. Palmer[7,8], Amanda Barkley-Levenson[7], Richard T. Lee[1,*], Peter C. Dedon[2,3,*].

[1] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, 02138, USA

[2] Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

[3] Singapore-MIT Alliance for Research and Technology, Antimicrobial Resistance Interdisciplinary Research Group, Campus for Research Excellence and Technological Enterprise, Singapore 138602, Singapore

[4] Harvard Center for Mass Spectrometry, Harvard University, Cambridge, MA 02138

[5] University of Pennsylvania Cardiovascular Institute, Philadelphia, PA, United States.

[6] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

[7] Department of Psychiatry, University of California San Diego, La Jolla, CA, 92093, USA

[8] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, 92093, USA

[*] Corresponding authors: Richard Lee, Richard_Lee@harvard.edu; Peter Dedon, pcdedon@mit.edu

[Ψ] Contributed equally

## Summary

DNA damage causes genomic instability underlying many human diseases. Traditional approaches to DNA damage analysis provide minimal insights into the spectrum of disease-driving DNA lesions and the mechanisms causing imbalances in damage formation and repair. Here we used untargeted mass spectrometry-based adductomics[1] to discover 114 putative DNA lesions and modifications consistently detected in humans and two independent analyses in rats, showing species-, tissue-, age-, and sex-biases. As evidence of methodologic rigor, 10 selected adductomic signals were structurally validated as epigenetic marks: 5-MdC, 5-HMdC, 5-FdC; DNA damage products: $N^2$-CMdG, 1,$N^6$-εdA, 3,$N^4$-εdC, $M^1$dG, $O^6$/$N^2$-MdG, and 8-Oxo-dG; and established analytical artifacts: cyclobutane dimers of 2'-deoxycytosine. With steady-state levels of putative DNA adducts integrating multiple cell types in each tissue, there was strong age-dependent variation for many putative adducts, including $N^2$-CMdG, 5-HMdC, and 8-Oxo-dG in rats and 1,$N^6$-εdA in human heart, as well as sex biases for 67 putative adducts in rat tissues. These results demonstrate the potential of untargeted adductomic analysis for defining DNA adducts as disease determinants, assigning substrates to DNA repair pathways, discovering new metabolically-driven DNA lesions, and quantifying inter-individual variation in DNA damage and repair across populations.


## Main text

***Refinement and validation of untargeted adductomics***. Traditional targeted approaches to analyzing DNA adducts and structural alterations have revealed the existence of dozens of types of damage caused by both endogenous metabolic processes and exogenous environmental agents[2,3]. However, targeted approaches have yielded little information about

50 the true spectrum of DNA adducts and lesions driving aging and disease in human tissues.
51 Here we took an untargeted low-resolution mass spectrometry-based "adductomic"
52 approach[1,4-8] to characterize known DNA lesions and modifications and discover new ones
53 unique to individual rat and human tissues and the behavior of these damage products as a
54 function of age and sex. The adductomics approach involves an untargeted discovery phase in
55 which the 2'-deoxynucleoside components of hydrolyzed DNA are analyzed by multiple
56 reaction monitoring (MRM) of collision-induced dissociation (CID) products losing 2'-
57 deoxyribose (*m/z* 116) in 1 Dalton increments across the range of *m/z* 225-525 (*i.e.,* "stepped
58 MRM"). This approach provides greater sensitivity (≤ 10 fmol; *vide infra*) than simple neutral
59 loss scanning. After filtering data to remove known artefacts and other noise (**Fig. 1a**), the
60 remaining putative DNA adduct signals serve as a basis set for subsequent MRM-based LC-
61 MS/MS analysis of individual DNA samples to quantify the putative adducts. For all studies
62 in rat, mouse, and human tissues, care was taken to avoid adventitious damage during DNA
63 isolation and processing, with addition of antioxidants and deaminase inhibitors[9].
64 **Supplementary Table 1** details the adductomics discovery phase using DNA isolated from
65 four tissues from groups of four male and four female Brown-Norway rats ranging age from
66 1- to 26-months-old, in two completely independent studies performed on different mass
67 spectrometers with different rat cohorts. To provide the broadest coverage of adduct variation,
68 DNA samples from each tissue were pooled across all ages to create four samples subjected to
69 adductomics discovery analysis. The 300 mass spectrometer signals from adduct discovery
70 for each organ were reduced to 92 and 94 mass spectrometer signals in the two independent
71 rat studies and assigned as putative DNA adducts and modifications on the basis of several
72 filtering criteria (**Fig. 1a**): (1) MS fragmentation released a mass consistent with 2'-
73 deoxyribose, (2) signals were detected in at least one tissue at a level that was at least three-
74 times the average background signal intensity and were detected in all individual samples of

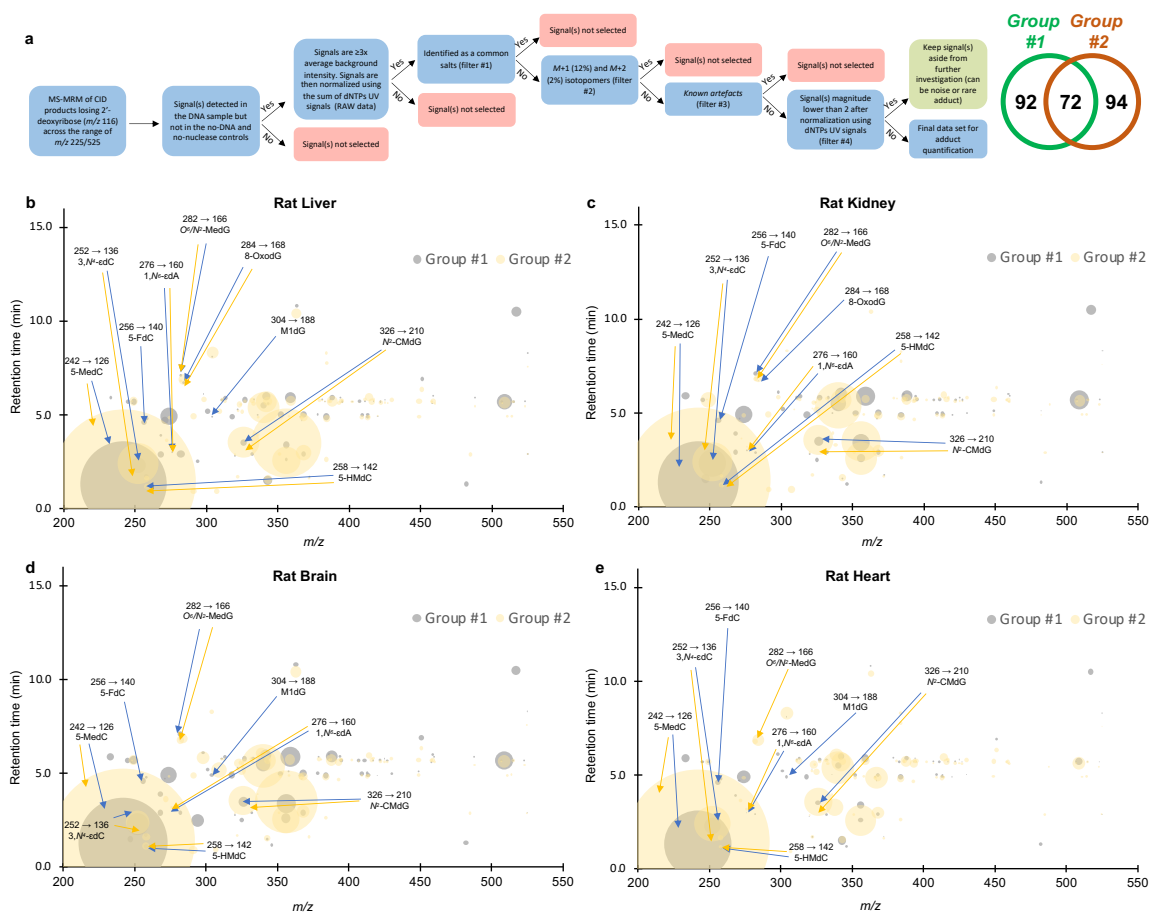



**Figure 1. Adductomic analysis in two different sets of rat tissues.** Adductomics involves two stages of data acquisition: (1) discovery of putative adducts by stepped MRM analysis of pooled DNA samples and (2) subsequent targeted analysis of a high-confidence set of adductomic signals representing putative DNA adducts in each tissue sample. (**a**) For the first step of identifying putative DNA adducts, the stepped MRM signals must be curated to remove artifacts, noise, and other spurious signals. Here we developed a decision tree using different analytical filters to remove common salts, isotopomers, known artifacts, and potential noise (signal intensity <2) to obtain a final data set of DNA adducts for subsequent targeted analysis. The Venn diagram shows the final sets of 92 and 94 putative adducts identified in discovery studies for tissues from two different rat cohorts, yielding a total set of 114 putative DNA adducts and modifications for subsequent quantitative analysis in each sample. (**b**) Using the curated set of adducts, DNA adductome profiles were generated for individual DNA samples from rat liver (**b**), kidney (**c**), brain (**d**), and heart (**e**) in rats ranging in age from 1 to 26 months for rat group #1 (blue) and 4 to 26 months for rat group #2 (yellow) months. DNA adducts that were further identified using chemical standards are labelled with names and CID transitions. Each DNA adduct, defined or putative, is represented by its retention time (min, y-axis), mass-to-charge ratio (*m/z*, x-axis) and relative abundance (size of the bubble). Signals from the four canonical DNA nucleobases are not represented in the plot. Blue bubbles and arrows are DNA adduct signals observed in group #1 while yellow are those observed in group #2.

each rat or human tissue, (3) signals were not detected in no-DNA and no-nuclease controls (*i.e.,* all reagents except DNA or nuclease/phosphatase), (4) signals representing salts (*e.g.,* $Na^+$, $K^+$) were removed, (5) signals representing putative isotopomers of stronger signals (*m/z* +1 and +2) were removed, (6) known artifacts were removed, (7) signals with normalized

signal intensities ≤ 2.0 and not shared between the two independent analyses were removed. It is important to note that the mass spectrometer signals could represent co-eluting mixtures of chemicals with the same *m/z* value and cannot be assumed to represent individual 2'-deoxyribonucleosides until the signals are structurally validated. The power of the adductomics method lies in the ability to discover all detectable damage products and modifications in cells or tissues and then quantify their variance among tissues and conditions, with subsequent structural analysis confirming the identity of the putative DNA damage products or modifications as the first step in mechanistic understanding of disease-driving genetic toxicology.

As evidence of the rigor of the discovery method, 72 of putative adducts (63%) were shared between the two independent cohorts of rats analyzed with two different mass spectrometer systems (Agilent 6490 and 6495) (**Fig. 1a**, **Supplementary Table 1**). We then attempted to assign chemical structures to each of the 114 putative adducts according to literature data[2,5-7,10-15], subsequent high-resolution mass spectrometric analysis of HPLC fractions, or by comparison to synthetic standards. The adductomic analysis revealed 66 previously undescribed putative 2'-deoxyribonucleoside species as well as 48 species with fragmentation and *m/z* values consistent with known damage products (average normalized signal intensities and tentative chemical identities in **Supplementary Table 1** and **Extended Data Figure 1**). These results are summarized in the bubble plots in **Figure 1**, depicting the identity and quantity of putative DNA damage products in liver (**Fig. 1a**), kidney (**Fig. 1b**), brain (**Fig. 1c**), and heart (**Fig. 1d**) tissues from the two rat cohorts (blue and yellow bubbles).

That the untargeted adductomic method revealed true DNA damage products and modifications is supported by confirmation of the structures of 10 signals as known DNA

104 lesions based on comparison to chemical standards (**Extended Data Fig. 1c**). For example,
105 we detected a putative adduct with *m/z* of 242→126 in all tissues, which was validated as the
106 epigenetic marker 5-methyl-2'-deoxycytidine (5-MdC; **Fig. 2**). The fact that the signal
107 intensity for the *m/z* of 242→126 5-MdC modification is several orders-of-magnitude higher
108 than other putative adducts in all tissues (**Fig. 1, Supplementary Table 1**) is consistent with
109 its presence as ~1% of all dC in the human genome[16,17]. Also consistent with the quantitative
110 rigor of the method, we detected an age-dependent signal with *m/z* 326→210 at much lower
111 levels than 5-MdC and identified it as $N^2$-carboxymethyl-2'-deoxyguanosine ($N^2$-CMdG)
112 (**Fig. 2**, **Supplementary Table 1**) – an advanced glycation end-product resulting from the
113 reaction between glyoxal and 2'-deoxyguanosine – by chemical synthesis of a standard for

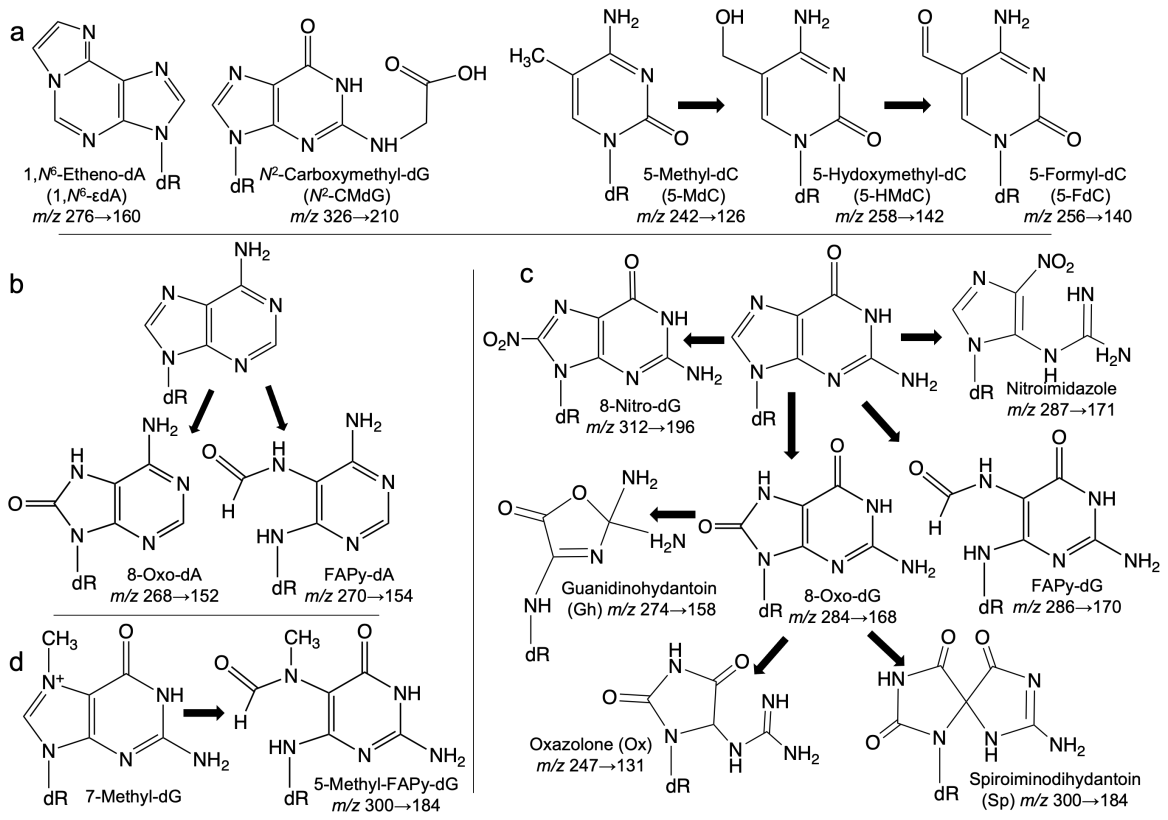



**Figure 2. Known DNA damage products and modifications found or predicted among adductomic signals.** (**a**) Using synthetic standards, adductomic signals were validated as the DNA damage products *1,N6*-εdA, $N^2$-CMdG, and 8-Oxo-dG (panel **c**), among others, as well as the epigenetic marks 5-MdC and 5-HM-dC. (**b-d**) Oxidation of canonical 2'-deoxynucleosides and DNA damage products leads to a host of secondary products predicted to occur *in vivo,* including FAPy-dA (*m/z* 271.1→155.1), guanidinohydantoin (*m/z* 274.1→158.1), 8-nitro-dG (*m/z* 313.1→197.1), and oxazolone (*m/z* 247.1→131.1). See **Supplementary Table 1** for adductomic signal data.

114 comparison by high-resolution LC-MS analysis (**Extended Data Fig. 1c**). Similar structural
115 validation was achieved for DNA damage arising from oxidation (8-hydroxy-2'-
116 deoxyguanosine, 8-Oxo-dG) and alkylation (1,$N^6$-etheno-2'-deoxyadenosine, 1,$N^6$-εdA; 3,$N^4$-
117 etheno-2'-deoxycytosine, 3,$N^4$-εdC; $N^1$-methyl-2'-deoxyguanosine, $M^1$dG; $O^6$- or $N^2$-methyl-
118 2'- deoxyguanosine, $O^6$-or $N^2$-MdG), as well as physiological epigenetic marks (5-
119 hydroxymethyl-2'-deoxycytidine, 5-HMdC; 5-formyl-2'-deoxycytidine, 5-FdC). The rigor of
120 the method was further validated with a putative adduct having an *m/z* 455→339
121 (**Supplementary Table 1**). High-resolution $MS^4$ Orbitrap mass spectrometric analysis
122 revealed the structure of this adduct as the cyclobutane dimer of 2'-deoxycytosine (dC=dC;
123 **Extended Data Fig. 2**). That this adduct co-eluted with dC is consistent with its formation as
124 an artifact of ionization and not as a pre-existing damage product in DNA.
125
126 Either due to lack of formation or levels below detection limits, we did not detect DNA
127 halogenation products (*e.g.*, 5-chloro-2'-deoxycytidine, 8-chloro-2'-deoxyguanosine, 8-
128 chloro-2'deoxyadenosine, 2-chloro-2'-deoxyadenosine) or DNA deamination products 2'-
129 deoxyxanthosine (dX; from 2'-deoxyguanosine), 2'-deoxyinosine (dI; from 2'-
130 deoxyadenosine), and 2'-deoxyuridine (dU; from 2'-deoxycytidine). dX is relatively unstable
131 and depurinates during mass spectrometric ionization[9], while pyrimidine nucleosides such as
132 dU do not ionize well[18], which may explain our inability to detect these damage products in
133 the set of 114 detected species. Rough upper boundaries for the limits of quantification (LOQ)
134 by the stepped MRM adductomics method can be inferred from targeted analyses of 8-Oxo-
135 dG, $N^2$-CMdG, 5-HMdC and 1,$N^6$-εdA by isotope-dilution LC-MS (**Extended Data Fig. 1**),
136 which revealed LOQ values as follows (compound, fmol injected, lesions per $10^9$ nt): 8-Oxo-
137 dG, 0.9, 49; $N^2$-CMdG, 3.8, 209; 5-HMdC, 4.2, 231; 1,$N^6$-εdA , 2.6, 143. The 114 species
138 identified by this validated adductomic method are thus very likely to be DNA damage

139 products and modifications meriting further characterization given their strong associations
140 with tissue, sex, and age in mice, rats, and humans.
141
142 ***Tissue- and sex-specific DNA damage and modification in rats***. The observation of 114
143 putative DNA adducts identified in two independent studies of four tissues from male and
144 female rats that were 1- to 26-months-old raised the question of tissue-, sex-, and age-
145 specificity of the individual adduct species. At the simplest level, a comparison of adductomic
146 signal intensities (**Extended Data Fig. 3**) shows substantial variation among brain, heart,
147 kidney, and liver. The Venn diagram in **Figure 3a** quantifies the distributions of these adducts
148 among the tissues, revealing that the epigenetic marks (5-MdC, 5-HMdC, 5-FdC) were
149 detected in every tissue, as expected. While there were a few putative adducts uniquely
150 detected in each tissue, 60-70% of the signals were detected in all tissues (52 of 92 in group
151 #1 and 68 of 94 in group #2) (**Supplementary Table 1**). Among the adducts shared in all
152 tissues, a principal components analysis (PCA) revealed strong tissue-specific variations in
153 the levels of the shared adducts (**Extended Data Fig. 4a,c**), with a Bayesian statistical
154 approach (**Extended Data Figure 4b,e**) revealing 5 clusters common to both groups of rats
155 (**Extended Data Figure 4c,f**). This similar clustering for the two independent adductomics
156 experiments underscores the potential biological meaning of the datasets. A further Gaussian
157 mixture model clustering of the 52 shared putative DNA adducts from group #1 revealed that
158 signals from heart tissue were uniquely segregated into two clusters, with Cluster III being
159 heavily biased to sex (5 female) and age (≥16 mo) (**Fig. 3b**). A significant number of sex-
160 biased DNA adducts and modifications were observed in the four tissues from group #1, as
161 shown in **Figure 3c-f** and **Supplementary Table 2** (brain, 1; heart, 30; kidney, 34; and liver,
162 2; all with FDR = 1%). Among the epigenetic marks, significant sex differences were
163 observed for 5-HMdC in heart and kidney tissues, and for 5-FdC in kidney, with no difference

observed for 5-MdC in any tissue (**Supplementary Table 2**). While clusters were strongly homogenous in group #1 rats, it was not possible to analyze group #2 rats for sex differences since there were no females in 2 of 4 groups (**Extended Data Figure 3f-i**).

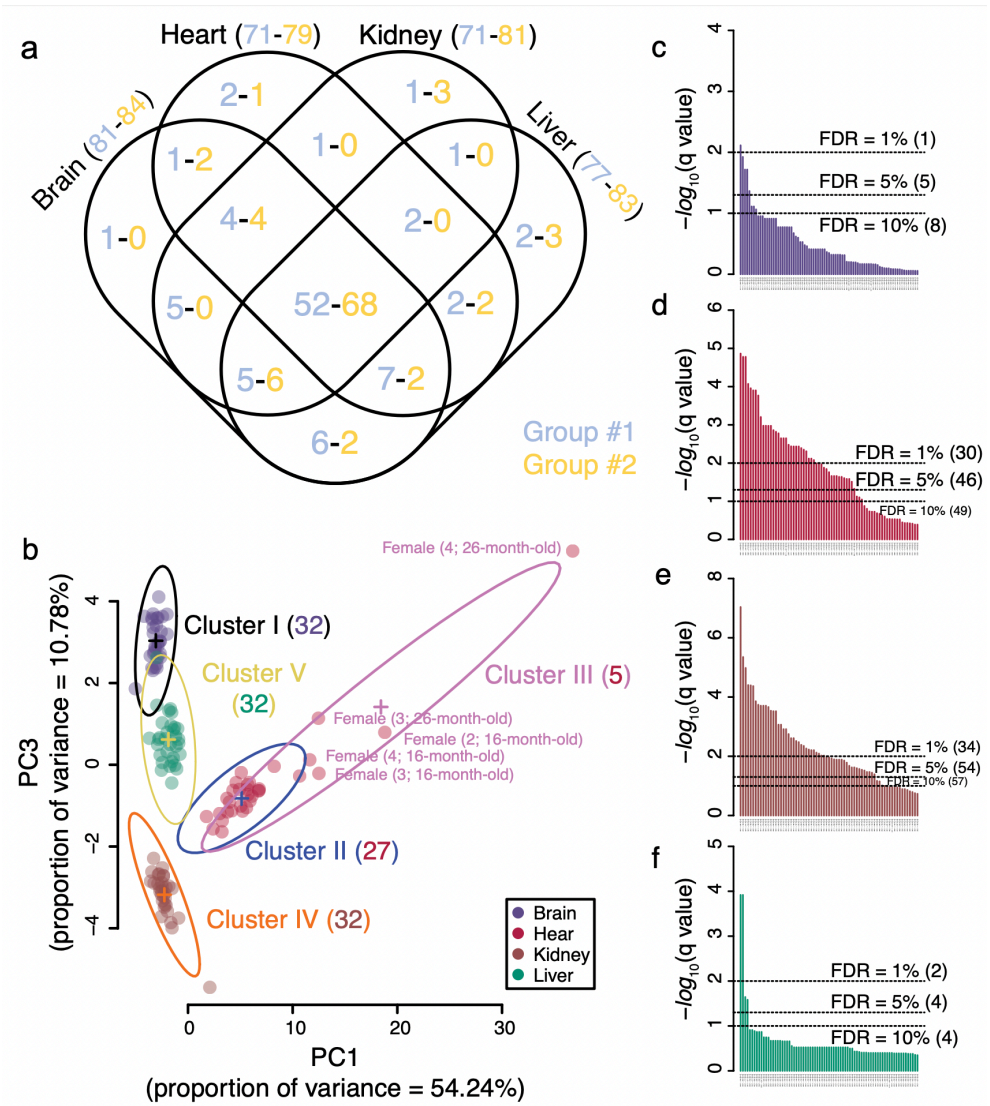



**Figure 3. DNA damage is strongly tissue- and gender-specific in rats**. (**a**) Venn diagram of the putative DNA adducts shared among the four rat tissues. Blue and yellow numbers (X-Y) correspond to the 92 putative adducts identified in group #1 rats and 94 in group #2 rats. Data are from **Supplementary Table 1**. (**b**) Based on multivariate statistical analysis of tissue, sex, and age associations of the 52 shared adducts in group #1 rats (**Extended Data Fig. 4a-c**), an EVV model (ellipsoidal distributions with equal volume and variable shape and orientation axes) revealed 5 underlying Gaussian distributions (Clusters I-V). The number of members in each cluster is shown in parentheses. Members of Cluster III comprise a striking group of 5 heart samples from mainly older female rats (labeled with sample # and age). (**c-f**) Bar plots of q-values for comparisons of putative DNA adduct levels in tissues from female versus male group #1 rats: (**c**) brain (81 adducts), (**d**) heart (71 adducts), (**e**) kidney (71), and (**f**) liver (77). q-Values were calculated from p-values obtained with the Wilcoxon rank sum test pn data from **Supplementary Tables 1** and **2**. Dotted lines represent the false discovery rates (FDR) at 10, 5, and 1%, with values in parentheses representing the number of significant adducts at each threshold.

168 ***Adductome analysis in rat tissues reveals age-dependent DNA damage and modification.***

169 While most DNA adducts in rat tissues were unchanged with age, several adducts showed

170 strong age dependence, including 5-HMdC, $N^2$-CMdG, $N^2$-methyl-dG, 8-Oxo-dG, and 1,$N^6$-

171 εdA (**Extended Data Fig. 5, Supplementary Table 3**). To validate the adductomic results,

172 we performed sensitive and specific absolute quantification of these adducts, using isotope

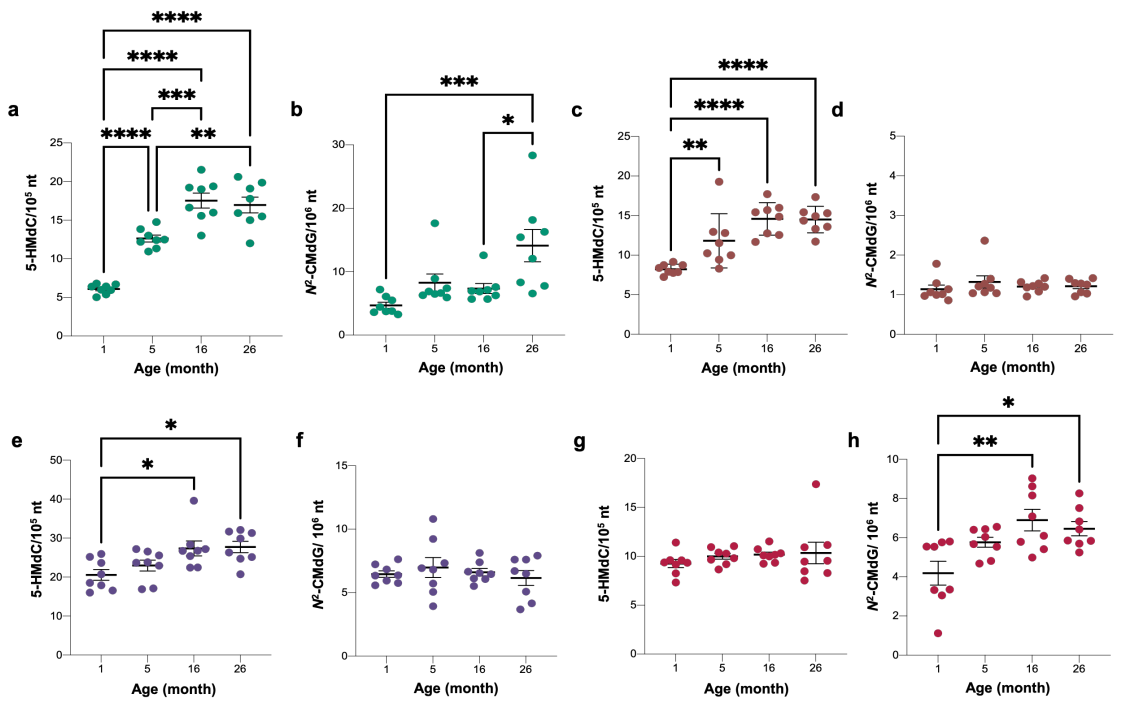



**Figure 4. Quantitative analysis of age-dependent DNA damage products in rat tissues.** Isotope dilution chromatography-coupled triple quadrupole mass spectrometric analysis was used to quantify 5-HMdC (**a, c, e, g**) and $N^2$-CMdG (**b, d, f, h**) levels in liver (**a, b**), kidney (**c, d**), brain (**e, f**), and heart (**g, h**) from rats at 1-, 5-, 16- and 26-months-old. Box and whisker plots show mean ± SEM for N=8. One-way ANOVA with Bonferroni's multiple comparisons test was used to evaluate differences among the ages. *$p$<0.05, **$p$<0.01, ***$p$<0.001 and **** $p$<0.0001 indicate statistically significant differences.

173 dilution triple-quadrupole mass spectrometry, as shown in **Figure 4**, **Extended Data Figure**

174 **6**, and **Supplementary Table 3**. The epigenetic mark 5-HMdC was previously observed to

175 increase with age in mouse liver, kidney and brain[16,17] and in human tissues[19]. Our study

176 confirmed this increase in rats aged 1 to 26 months in liver (2.8-fold, **Fig. 4a**), kidney (1.7-

177 fold, **Fig. 4c**), and brain (1.4-fold, **Fig. 4e**), while it was unchanged in heart ($p$= 0.62) (**Fig.**

178 **4g**). Similarly, we confirmed a previous observation of age-dependent increases in 8-Oxo-

179 dG in rat liver[20], with a 1.8-fold increase over 26 months of age ($p$=0.0001; **Extended Data**

180 **Fig. 6e**). These results validate the quantitative robustness of the adductomic analyses.
181 Similar analyses were not performed on the group #2.
182
183 We also found age-dependent increases in $N^2$-CMdG, an advanced glycation end-product
184 (AGE) that results from a reaction between glyoxal and 2'-deoxyguanosine (dG)[21] and that
185 has been in mouse hepatoma cells, 293T human kidney cells, and calf thymus DNA[12,22]. In
186 rat, $N^2$-CMdG increased in 3-fold in liver (**Fig. 4b**) and 1.5-fold in heart (**Fig. 4h**) over 1-26
187 months but remained unchanged in kidney (**Fig. 4d**) and brain (**Fig. 4f**). While mutations
188 caused by glyoxal-induced DNA adducts can be prevented by nucleotide excision DNA repair
189 (NER), mismatch DNA repair (MMR), and the protein DJ-1 *in vitro*[23,24], metabolically-
190 derived glyoxal, methylglyoxal, and other reactive aldehydes are removed by the glyoxalase
191 system of GLO1 and GLO2 in humans to prevent DNA adduct formation. In this system,
192 GLO1 converts glyoxal, methylglyoxal, and glutathione (GSH) to *S*-glycolylglutathione that
193 is cleaved by GLO2 to glycolate, D-lactate, and GSH[25]. Here, we quantified the effect of
194 GLO1 expression on $N^2$-CMdG levels in brain, heart, and liver tissues from mice with *Glo1*
195 knock-down[26] or over-expression[27] and wild-type littermates. Surprisingly, the only change
196 observed in $N^2$-CMdG levels involved a significant increase in liver in GLO1 over-expressing
197 mice compared to wild-type (**Extended Data Fig. 6f-j**). These data indicate either that $N^2$-
198 CMdG is formed from sources other than glyoxal or that GLO1 does not significantly affect
199 glyoxal levels in mice, either of which has implications for the observed age-dependent
200 accumulation of $N^2$-CMdG.
201
202 ***5-HMdC and 1,N$^6$-εdA are age-dependent DNA adducts in human heart.*** The observed
203 results in rats raised the question of age-dependent DNA damage and modification in humans.
204 To this end, we performed adductomic analyses with DNA from non-failing human

205

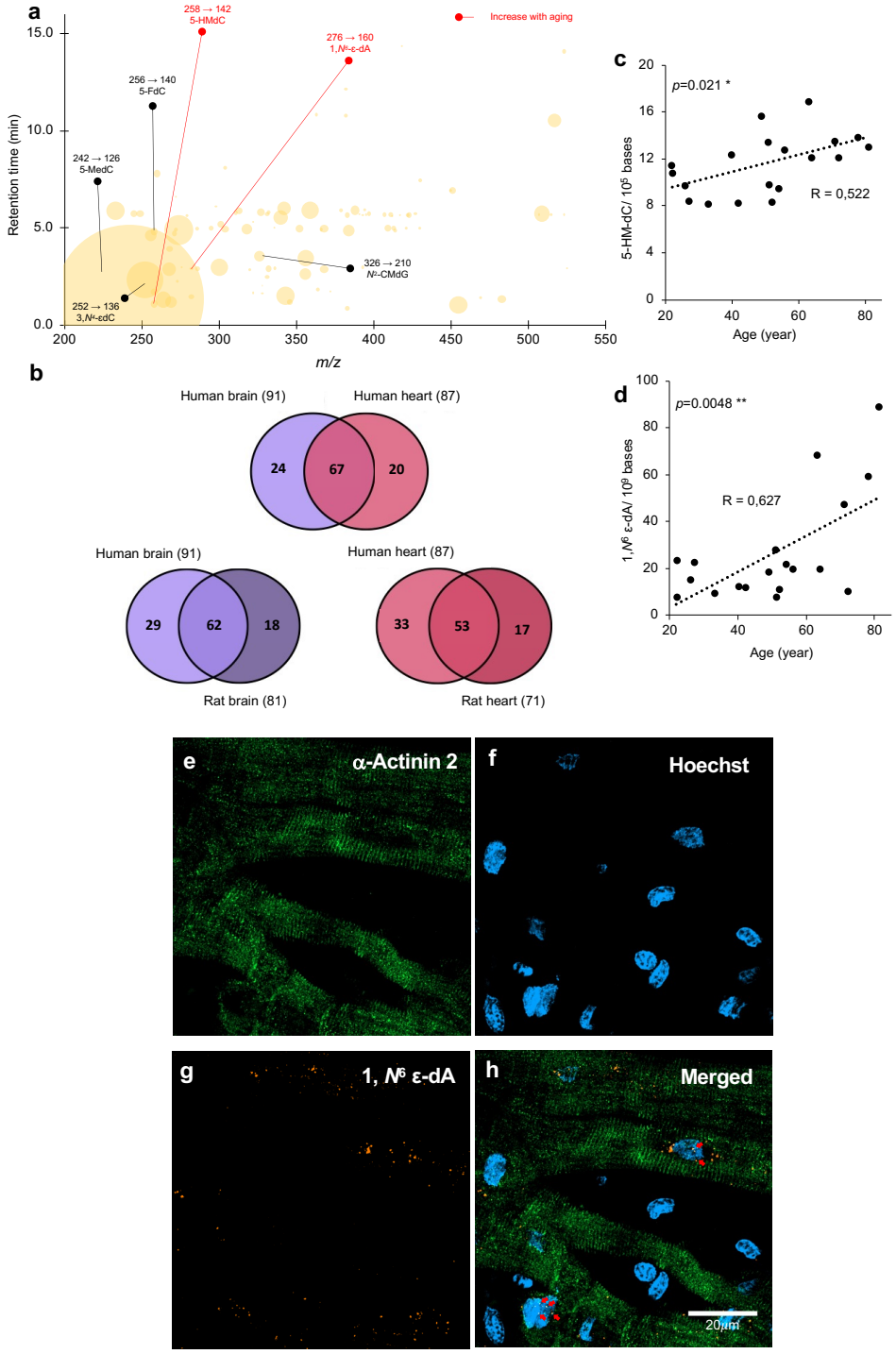



**Figure 5. Discovery of age-tracking DNA adducts in human heart.** (**a**) The bubble plot shows the DNA adductome map in non-failing human heart. Mass spectrometry transitions are noted above the names of adducts verified by chemical standards. Adducts that increase with age are highlighted in red**.** (**b**) Venn diagrams showing overlap among the adductomic signals detected in human and rat heart and brain. (**c, d**) Levels of 5-HMdC and 1,$N^6$-εdA, 5-HMdC were quantified in 19 human non-failing heart samples from donors ranging in age from 22- to 81-years-old by isotope dilution chromatography-coupled triple quadrupole mass spectrometric analysis. Pearson's correlation test was used to establish correlation between 5-HMdC and 1,$N^6$-εdA levels and age; *$p$<0.05 and **$p$<0.01. (**e-h**) Confocal imaging of sectioned and immune-stained human non-failing heart samples reveals the presence of scattered and sparse 1,$N^6$-εdA throughout cardiac tissue. 1,$N^6$-εdA is present within cardiac cells, as seen by orange spots found within green (α-Actinin 2 labeled) areas. Red arrows denote 1,$N^6$-εdA spots within nuclei (labeled with Hoechst), as shown in the merged figure panel (**h**). Scale bar represents a length of 20 $\mu$m.

206 myocardial tissues comprised of left ventricular myocardial samples from 17 people ranging
207 in age from 22- to 81-years-old[27,28] and from 10 human brain tissue samples from persons
208 without cognitive impairment ranging in age from 71- to 89-years-old (**Supplemental Table**
209 **5**). The analysis revealed 111 total putative DNA adducts and modifications, with 91 detected
210 in brain and 87 in heart (**Fig. 5**, **Supplementary Table 3**). The Venn diagrams in **Figure 5b**
211 reveal that rats and humans share signals for 62 putative adducts in brain and 53 in heart.
212 Among these, adductomics revealed that 5-HMdC and 1,$N^6$-εdA all increased with age in
213 human hearts (**Fig. 5**, **Supplementary Table 3**), with this age dependence corroborated by
214 isotope dilution LC-MS/MS that shows a significant and positive correlation for 5-HMdC ($p$=
215 0.021) and 1,$N^6$-εdA ($p$= 0.0048) in non-failing human hearts (**Fig. 5c,d**; **Supplementary**
216 **Table 3**). As a product of reaction of dA with electrophilic products of lipid peroxidation[29],
217 1,$N^6$-εdA levels are elevated in human diseases such as Wilson's disease, hemochromatosis,
218 and cirrhosis[30], and show age-dependent increases in liver and brain in the OXYS rat model
219 of aging due to oxidative stress[31]. **Figure 5e-h** shows 1, $N^6$ ε-dA concentrated mainly in select
220 cardiomyocyte nuclei, raising the question of the age of these cells in the local tissue
221 environment.
222

223 ## Discussion

224

225 In spite of the overwhelming evidence that imbalances in hundreds of types of DNA damage
226 and dozens of DNA repair enzymes cause and drive many human diseases, such as in
227 cancer[32], the underlying mechanisms that lead to age-, tissue-, and sex-specificity of genetic
228 toxicity are poorly understood. Here, we took a first step in unraveling this problem with a
229 systems-level estimate of the spectrum of DNA damage products and modification in rat and
230 human tissues. Such an analysis is possible only with the emergence of convergent

231 adductomic technology consisting of untargeted mass spectrometric detection and
232 quantification of virtually any type of nucleobase damage in DNA[6] coupled with statistical
233 modeling to visualize patterns and covariations among the putative DNA damage products.
234 This systems-level untargeted DNA adductomic approach serves as a discovery tool by
235 revealing all detectable forms of DNA damage, including new chemical species, followed by
236 providing relative quantification of adduct levels across different conditions.
237
238 ***Validation of the adductomic method***. As an untargeted method, there is always concern
239 about detecting artifacts caused by non-DNA contaminants that produce a low-resolution CID
240 fragmentation mimicking loss of 2'-deoxyribose. Such artifacts are mitigated by starting with
241 highly purified DNA and accounting for contaminants in non-DNA components of the
242 analyzed samples. Each analytical run must include controls lacking buffer, enzymes, and
243 DNA to identify contaminants. Our adductomic method helped us to achieves these
244 requirements. Selected signals were reproducibly present at 3-times the background level in
245 all 32 samples of each analyzed rat tissue (*i.e.,* 8 samples of each tissue at 4 different ages;
246 22% average error), with more than half of the detected adducts being shared between rats
247 and humans in brain and heart tissues (**Fig. 5b**) in two different sets of rat. The adductomic
248 method was validated (1) by *post hoc* identification of 10 detected adducts using chemical
249 standards and (2) by high-resolution structural analysis of products, such as the tentatively
250 identified cyclobutane dimer of dC. Of the remaining 114 unidentified species, about half had
251 low-resolution *m/z* values consistent with known DNA damage structures (**Supplementary**
252 **Table 1**). More interesting after filtering our data, we were able to find 63% of shared adducts
253 between the two studies showing the robustness of our approach. Thus, there is a high
254 probability that most signals detected in our adductomic analyses are indeed DNA adducts
255 that merit identification and definition of formation and repair mechanisms to understand their

256 roles in pathobiology and the basis for our observed age, tissue, and sex biases. Of course, the
257 adductomic method is not as sensitive as isotope-dilution LC-MS/MS and will not detect
258 labile nucleobase species, such as rapid depurination of 8-nitro-2'-deoxyguanosine arising
259 during inflammatory stresses[2], or species arising from oxidation of the 2'-deoxyribose
260 moiety[33]. There are also complexities in the adductomic method, such as MS ionization of
261 adducts as salts (*e.g.,* possible ammonium salt of 5-Cl-dC for *m/z* 279-163; **Supplementary**
262 **Table 1**) and rearrangements of free 2'-deoxyribonuclesides (*e.g.,* sugar tautomers unique to
263 FAPy-dG; *vide infra*). Despite these constraints, the strength of the adductomic approach lies
264 in the insights gained from mining the systems-level datasets for correlative patterns that
265 reveal behaviors of classes and groups of DNA adducts across different conditions of health
266 and disease. By first identifying adductomic patterns, then identifying the chemical species
267 driving the patterns, and finally doing the biochemical detective work, we can define the
268 DNA damage products that truly drive human disease.
269
270 ***Mining adductomic data for chemical classes of DNA adducts.*** There are many ways to
271 mine adductomic datasets. One initial approach is to identify a putative DNA damage species
272 that correlates with a condition of interest, such as the age-dependent increases in 8-Oxo-dG
273 (**Extended Data Fig. 6e**), and then explore other mechanistically related species, such as the
274 large and well-studied class of purine damage products arising during inflammation and other
275 oxidative and nitrosative stresses (**Fig. 2**)[2]. Taking this path, we identified 8-Oxo-dG among
276 the 114 adducts using a standard but failed to detect signals accordant with 8-Oxo-dA (*m/z*
277 268→152) or FAPy-dA (*m/z* 270→154). This is consistent with the observation of 10-fold
278 lower levels of 8-Oxo-dA than 8-Oxo-dG and the relative amounts of 8-Oxo-dA, 8-Oxo-dG,
279 FAPy-dA, and FAPy-dG in cells and tissues subjected to irradiation and other oxidative
280 stresses (**Fig. 2**)[34-36]. While we did detect two signals suggestive of the known anomeric or

281 sugar tautomer forms of FAPy-dG (*m/z* 286-170)[37-39], the observed HPLC retention times of
282 2.9 and 4.9 min for *m/z* 286-170 (**Supplementary Table 1**) are inconsistent with
283 expectations-based retention times of ~0.7 and ~0.9 min for 5-methyl-FAPy-dG (**Fig. 2**). We
284 also observed signals suggestive of the other oxidized purine nucleobase products noted in
285 **Figure 2**: oxazolone (Oz; *m/z* 247→131), nitroimidazole (NitroIm; *m/z* 287→171),
286 spiroiminodihydantoin (Sp; *m/z* 300→184; same as FAPy-dG), and guanidinohydantoin (Gh;
287 *m/z* 274→158). While 8-Oxo-dG and putative Oz were variably present in rat and human
288 tissues, the putative NitroIm signal was only present in rat liver tissue and a signal consistent
289 with Gh was present in all rat and human tissues at high levels. The observation with Gh runs
290 contrary to published *in vitro* studies of DNA damage caused by a variety of oxidants in
291 which Gh rose at levels nearly two orders-of-magnitude lower than 8-Oxo-dG in all cases[40].
292 However, the biochemical environment in cells and tissues has the potential to strongly
293 influence the chemical reactions involved in DNA damage formation and partitioning, which
294 leaves the door open to the identity of *m/z* 274→158 as Gh. Based on the HPLC retention
295 times of chemical standards, we ruled out Sp and 5-methyl-FAPy-dG as the adduct present at
296 *m/z* 300→184: ~0.95 min for Sp (consistent with its hydrophilicity) and ~0.7 and ~0.9 min for
297 the anomeric or tautomeric pair of FAPy-dG species (**Fig. 2**). The adductomic data thus
298 provides the foundation for a host of testable hypotheses about the *in vivo* behavior of entire
299 classes of adducts arising by similar chemical mechanisms: Are *in vitro* observations
300 recapitulated *in vivo*? How does the cellular environment alter chemical partitioning? What
301 are the true oxidants causing damage to dG and dA *in vivo*? Biochemical detective work,
302 structural validation, and subsequent targeted analyses can then be used to test these models.
303 Structural validation using high-resolution mass spectrometry and chemical synthesis is well-
304 practiced and entirely feasible, as we demonstrated with the cyclobutane dimer of dC and $N^2$-
305 CMdG. One can also define the role of specific DNA repair mechanisms by comparing

306 adductomic profiles between wild-type and DNA repair mutant animal models or across
307 human populations with defined inter-individual repair capacities.
308
309 ***DNA adduct biases among tissues.*** The power of comparing adductomic profiles across
310 different conditions is illustrated by our observations of tissue-, age- and sex-specific biases in
311 adduct spectra. Strong tissue-specific biases were observed in rat for 114 putative and
312 validated DNA adducts (**Fig. 3; Extended Data Figures 3, 4d-f**), with a variety of tissue-
313 specific or at least tissue-enriched DNA damage products (**Supplementary Table 1**). The
314 normalized signal intensities for adducts noted in **Supplementary Table 1** represent
315 quantitative metrics that allow relative comparisons of adduct levels across tissues given the
316 precision of the LC-MS signals, with appropriate caution that the signal intensities do not
317 reflect absolute amounts of the putative adducts due to differences in ionization and detection
318 efficiencies. After removing the large signal for epigenetic mark 5-MdC, summing the signal
319 values in a tissue represents a crude measure of the DNA adduct load in the tissue (number of
320 adduct types and percentage of total adduct load in parentheses): rat brain, 3517 (81; 28%);
321 rat heart, 1458 (71; 12%); rat kidney, 3574 (71; 29%); rat liver, 3790 (77; 31%); human brain,
322 6142 (91; 63%); human heart, 3584 (87; 37%). The DNA adduct loads were higher in the
323 tissues analyzed in group #2: rat brain, 9483 (85; 24%); rat heart, 7085 (84; 18%); rat kidney,
324 10552 (90; 26%); rat liver, 13080 (90; 32%). The higher normalized signal intensities
325 observed for the group #2 rat tissues compared to group #1 reflect higher sensitivity of the
326 Agilent 6495C MS used analyze group #2 compared to the Agilent 6490 MS used for group
327 #1. While there is little meaningful information that can be derived from the adduct load data,
328 it is notable that both rat and human heart have the lowest level of adducts relative to other
329 tissues and that the two rat groups shared a similar order of adduct loads
330 (heart<brain<kidney<liver). The complexity of the adductomic datasets, as mixtures of

331 potentially all classes of lesions type and repair, precludes comparisons to published studies
332 showing tissue-specific variation in adduct repair for one or a few representative adducts or a
333 single repair class[41-43]. For example, Gosponodov *et al.* estimated NER rates among rat
334 kidney (0.1 lesions/kb/hr), liver (0.1 lesions/kb/hr), brain (0.08 lesions/kb/hr), and heart (0.04
335 lesions/kb/hr)[44], while Langie *et al.* found lower BER activity in mouse brain compared to
336 liver[45]. Cell replication rate could also account for tissue differences. For example, repair of
337 8-Oxo-dG and FAPy-dG occurs by OGG1 base-excision repair throughout the cell cycle and
338 by MUTYH removal of A at 8-Oxo-dG:A mispairs during DNA replication[46]. MUTYH
339 activity is highest in rapidly dividing cells (*e.g.*, gastrointestinal tract) due to replication-
340 coupled repair of 8-Oxo-dG:A mismatches. However, in addition to other repair metrics, there
341 is no correlation between tissue-specific levels or loads of adducts and cell turnover rates in
342 humans: hepatocyte cells (200-400 d)[47], endothelial kidney cells (>1 y)[48], cardiomyocytes
343 (~50 y)[49] and neurons (lifespan)[50]. Even less is known about tissue-specific metabolism that
344 causes DNA adducts. For example, mitochondrial density and energy activity correlate with
345 production of reactive oxygen species and associated oxidative damage to lipids and
346 carbohydrates, to form DNA-reactive alkylating agents, as well as to DNA and RNA
347 directly[41,42]. Brain and liver are lower in both metrics compared to heart and kidney[41,42], but,
348 again, differences among the tissues in adduct load and the number of types of adducts in
349 these tissues do not correlate with these metrics. These attempts to explain the observed
350 biased tissue distribution of known and putative DNA adducts point to potential value in
351 performing adductomic analyses in genetically engineered mouse models for the various
352 DNA repair pathways, by identifying true substrates of DNA repair enzymes, quantifying the
353 contribution of repair to adduct levels in each tissue, and to the age- and sex-dependence of
354 adduct levels.
355

356 ***DNA adduct biases among different cell types in tissues.*** One of the limitations of
357 performing adductomic analyses on whole tissues is the potential for biases among specific
358 cell types comprising the tissue. Enzymatic disaggregation of tissues may be the optimal
359 approach for isolating cell populations, with the potential for adequate DNA for adductomic
360 analyses, though DNA damage artifacts during processing are likely. Heat-induced artifacts
361 are highly likely to occur with laser-capture microdissection, with DNA yield (30 ng/mm$^2$)
362 and low resolution also limiting its utility[51]. While the 5-10 μm resolution of imaging mass
363 spectrometry may provide adequate resolution in tissues, the method is likely to be limiting in
364 terms of sensitivity (1 fmol)[52]. Immunohistochemistry with highly specific antibodies
365 ultimately optimizes chemical specificity with resolution. Here we used immunostaining to
366 detect 1,$N^6$-εdA in the nuclei of cardiomyocytes in human ventricular myocardial samples,
367 which indicated that accumulation may be limited to a small number of cells. While this may
368 result from a sensitivity threshold and all cells show some level of the adduct, substantial
369 accumulation of DNA adducts in a small number of cells might indicate damage specific to
370 cell subtypes or to the oldest cells in the tissue, with the latter indicating compromised
371 genomic integrity during senescence.
372
373 ***Age dependence of DNA adducts.*** Comparative analysis of DNA adducts in rats across the
374 age span of 1 to 26 months revealed that a subset of DNA adducts did not change, several
375 adducts showed significant age-dependent accumulation and decreases with age in a tissue-
376 and sex-specific manner (**Fig. 4**, **Extended Data Fig. 6**). This suggests shifts in the balance
377 between adduct formation and repair mechanisms over the life of the animal. We observed
378 significant age-dependent increases of $N^2$-CMdG in rat liver and heart (**Fig. 4**). $N^2$-CMdG and
379 the related $N^2$-(1-carboxyethyl)-2'-deoxyguanosine ($N^2$-CEdG) are advanced glycation end-
380 products resulting from the reaction of glyoxal and methylglyoxal, respectively, with 2'-

381 deoxyguanosine in DNA. This is not surprising since the methylglyoxal and glyoxal are
382 reactive α-oxoaldehydes formed in abundance during the glucose metabolism[53] in liver[54] and
383 there is evidence for age-dependent decreases in base excision repair (BER) mechanism in
384 mouse liver[55], with $N^2$-CMdG likely repaired by the BER enzyme alkyladenine glycosylase
385 (Aag)[56]. In studies examining age-dependent mechanisms of formation, we observed no effect
386 of glyoxalase I activity on $N^2$-CMdG levels. It is known that glyoxalase I prefers
387 methylglyoxal as a substrate over glyoxal and protects against $N^2$-CEdG[57,58]. Interestingly, we
388 detected a signal consistent with $N^2$-CEdG at *m/z* 340→224 and found strong covariance of
389 this signal with that for $N^2$-CMdG (**Extended Data Fig. 6k, Supplementary Table 1**).
390
391 Adductomic analysis of human heart tissue also revealed age-dependent DNA damage and
392 epigenetic marks. We observed a significant age-dependent accumulation of 5-HMdC, which
393 results from oxidation of 5-MdC[19], in non-failing heart tissue, which parallels a published
394 observation of increases in 5-HMdC with age in grey and white matter of the human
395 cerebrum[17]. We also found age-dependent increase in the 1,$N^6$-εdA damage product. While
396 this contrasts the inconsistent changes in 1,$N^6$-εdA in rat heart (**Extended Data Fig. 6d**), it is
397 consistent with the idea of cardiomyocytes becoming more susceptible to oxidative stress with
398 age[59], with cardiac DNA repair mechanisms tending to be less efficient with age[60], and with
399 the observation of age-dependent increases in the 4-hydroxy-2-nonenal (HNE) lipid
400 peroxidation product that reacts with DNA to form 1,$N^6$-εdA[61,62]. The immunohistochemical
401 evidence for concentration of 1,$N^6$-εdA in subpopulations of cardiomyocyte nuclei suggests a
402 possible correlation between adduct formation and cell age or senescence.
403
404 ***Sex dependence of DNA adducts.*** The complexity of the network of determinants of DNA
405 adduct spectra is increased by our observation of sex-specific differences in adduct levels in

406 all four of the tissues studied (**Fig. 3c-f, Supplementary Table 2**). Interestingly, while the
407 epigenetic mark 5-MdC did not show a sex bias in any of the rat tissues, its oxidation products
408 5-HMdC and 5-FdC were significantly higher or trending higher in female kidney and heart
409 (**Supplementary Table 2**). While sex-based biases in metabolism may account for the
410 observation that females show higher levels of smoking-related DNA adducts in human lung
411 tissue[63] and in mice treated with 2-amino-3-methylimidazo [4,5-f]-quinoline, a potent food
412 mutagen[64], our observations with 5-HMdC and 5-FdC suggest sex differences in epigenetic
413 regulation of gene expression. The variety of unknown putative DNA adducts showing sex
414 biases merits further investigation of their structures and definition of their potential roles in
415 cell biology, aging, and pathology.
416
417 ***Conclusions.*** Adductomics proved to be an effective approach for discovering individual
418 DNA damage products as well as for systems-level analysis of the behaviors of dozens of
419 putative DNA adducts as a function of tissue type, age, and sex. Such covariate behavior can
420 be highly informative about local environment of DNA repair and metabolism, especially
421 when combined with large metabolomics[65], proteomics[66,67], and other datasets. While age-
422 sex-, and tissue-specific variations in the levels of 114 putative and method-validated DNA
423 adducts provide important new biological insights into genetic toxicology and epigenetics, the
424 adductomic approach illustrated here is a powerful approach to (1) assigning substrates to
425 DNA repair pathways using genetically engineered animal models, (2) understanding
426 mechanisms of DNA damage formation by monitoring classes of DNA lesions and
427 manipulating metabolic pathways, and (3) quantifying inter-individual variation in DNA
428 damage and repair across populations and pathologies. These efforts will be enhanced by
429 advancing the mass spectrometry and data analysis tools for adductomic profiling[67].
430

431 **Main Text References**

432


433 1 Kanaly, R. A. *et al.* Development of the adductome approach to detect DNA damage
434 in humans. *Antiox Redox Signal* **8**, 993-1001, (2006).
435 2 Dedon, P. C. & Tannenbaum, S. R. Reactive nitrogen species in the chemical biology
436 of inflammation. *Arch Biochem Biophys* **423**, 12-22, (2004).
437 3 Yousefzadeh, M. *et al.* DNA damage-how and why we age? *Elife* **10**, (2021).
438 4 Demir, E., Turna, F., Kaya, B., Creus, A. & Marcos, R. Mutagenic/recombinogenic
439 effects of four lipid peroxidation products in Drosophila. *Food Chem Toxicol* **53**, 221-
440 227, (2012).
441 5 Carrà, A. *et al.* Targeted High Resolution LC/MS3 Adductomics Method for the
442 Characterization of Endogenous DNA Damage. *Front Chem* **7**, 658, (2019).
443 6 Kanaly, R. A., Matsui, S., Hanaoka, T. & Matsuda, T. Application of the adductome
444 approach to assess intertissue DNA damage variations in human lung and esophagus.
445 *Mut Res* **625**, 83-93, (2007).
446 7 Guidolin, V. *et al.* Identification of New Markers of Alcohol-Derived DNA Damage
447 in Humans. *Biomolecules* **11**, 366, (2021).
448 8 Walmsley, S. J., Guo, J., Wang, J., Villalta, P. W. & Turesky, R. J. Methods and
449 Challenges for Computational Data Analysis for DNA Adductomics. *Chem Res*
450 *Toxicol* **32**, 2156-2168, (2019).
451 9 Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from
452 deamination, oxidation and reaction with products of lipid peroxidation by liquid
453 chromatography isotope dilution tandem mass spectrometry. *Nat Protoc* **3**, 1287-1298,
454 (2008).

455 10 Belous, A. R., Hachey, D. L., Dawling, S., Roodi, N. & Parl, F. F. Cytochrome P450
456 1B1-mediated estrogen metabolism results in estrogen-deoxyribonucleoside adduct
457 formation. *Cancer Research* **67**, 812-817, (2007).
458 11 Huang, H. *et al.* DNA Cross-links Induced by trans-4-Hydroxynonenal. *Environ*
459 *Molec Mutagen* **51**, 625-634, (2010).
460 12 Pluskota-Karwatka, D., Pawlowicz, A. J., Tomas, M. & Kronberg, L. Formation of
461 adducts in the reaction of glyoxal with 2'-deoxyguanosine and with calf thymus DNA.
462 *Bioorg Chem* **36**, 57-64, (2008).
463 13 Carvalho, V. M. *et al.* Novel 1,N(6)-etheno-2'-deoxyadenosine adducts from lipid
464 peroxidation products. *Chem Res Toxicol* **13**, 397-405, (2000).
465 14 Sousa, P. F. M., Martella, G., Åberg, K. M., Esfahani, B. & Motwani, H. V.
466 nLossFinder—A Graphical User Interface Program for the Nontargeted Detection of
467 DNA Adducts. *Toxics* **9**, 78, (2021).
468 15 Chou, P.-H. *et al.* Detection of lipid peroxidation-induced DNA adducts caused by 4-
469 oxo-2(E)-nonenal and 4-oxo-2(E)-hexenal in human autopsy tissues. *Chem Res*
470 *Toxicol* **23**, 1442-1448, (2010).
471 16 Zarakowska, E. *et al.* Oxidation Products of 5-Methylcytosine are Decreased in
472 Senescent Cells and Tissues of Progeroid Mice. *J Gerontol Series A: Biol Sci Med*
473 *Sci***73**, 1003-1009, (2018).
474 17 Wagner, M. *et al.* Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-
475 Formylcytosine in Human and Mouse Brain Tissues. *Angew Chemie Intnl Ed* **54**,
476 12511-12514, (2015).
477 18 Cai, W. M. *et al.* A Platform for Discovery and Quantification of Modified
478 Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA
479 Modifications. *Methods Enzymol* **560**, 29-71, (2015).

480 19 López, V., Fernández, A. F. & Fraga, M. F. The role of 5-hydroxymethylcytosine in
481 development, aging and age-related diseases. *Ageing Res Rev* **37**, 28-38, (2017).
482 20 Nie, B. *et al.* Age-Dependent Accumulation of 8-Oxoguanine in the DNA and RNA in
483 Various Rat Tissues. *Oxid Med Cell Long* **2013**, 303181, (2013).
484 21 Olsen, R. *et al.* Reaction of glyoxal with 2'-deoxyguanosine, 2'-deoxyadenosine, 2'-
485 deoxycytidine, cytidine, thymidine, and calf thymus DNA: identification of DNA
486 adducts. *Chem Res Toxicol* **18**, 730-739, (2005).
487 22 Wang, H., Cao, H. & Wang, Y. Quantification of N2-carboxymethyl-2'-
488 deoxyguanosine in calf thymus DNA and cultured human kidney epithelial cells by
489 capillary high-performance liquid chromatography-tandem mass spectrometry coupled
490 with stable isotope dilution method. *Chem Res Toxicol* **23**, 74-81, (2010).
491 23 Murata-Kamiya, N., Kamiya, H., Kaji, H. & Kasai, H. Nucleotide excision repair
492 proteins may be involved in the fixation of glyoxal-induced mutagenesis in
493 Escherichia coli. *Biochem Biophys Res Comm* **248**, 412-417, (1998).
494 24 Richarme, G. *et al.* Guanine glycation repair by DJ-1/Park7 and its bacterial
495 homologs. *Science* **357**, 208-211, (2017).
496 25 Thornalley, P. J. Glyoxalase I--structure, function and a critical role in the enzymatic
497 defence against glycation. *Biochem Soc Trans***31**, 1343-1348, (2003).
498 26 El-Osta, A. *et al.* Transient high glucose causes persistent epigenetic changes and
499 altered gene expression during subsequent normoglycemia. *J Exper Med* **205**, 2409-
500 2417, (2008).
501 27 Distler, M. G. *et al.* Glyoxalase 1 increases anxiety by reducing GABAA receptor
502 agonist methylglyoxal. *J Clin Investig* **122**, 2306-2315, (2012).

503 28 Caporizzo, M. A., Chen, C. Y., Bedi, K., Margulies, K. B. & Prosser, B. L.
504 Microtubules increase diastolic stiffness in failing human cardiomyocytes and
505 myocardium. *Circulation* **141**, 902-915, (2020).
506 29 Bartsch, H. & Nair, J. Oxidative stress and lipid peroxidation-derived DNA-lesions in
507 inflammation driven carcinogenesis. *Cancer Detect Prev* **28**, 385-391, (2004).
508 30 Bartsch, H. & Nair, J. Accumulation of lipid peroxidation-derived DNA lesions:
509 potential lead markers for chemoprevention of inflammation-driven malignancies. *Mut*
510 *Res* **591**, 34-44, (2005).
511 31 Nair, J. *et al.* Age-dependent increase of etheno-DNA-adducts in liver and brain of
512 ROS overproducing OXYS rats. *Biochem Biophys Res Comm* **336**, 478-482, (2005).
513 32 Basu, A. K. DNA Damage, Mutagenesis and Cancer. *Int J Mol Sci* **19**, (2018).
514 33 Dedon, P. C. The chemical toxicology of 2-deoxyribose oxidation in DNA. *Chem Res*
515 *Toxicol* **21**, 206-219, (2008).
516 34 Liu, D. *et al.* Evidence that OGG1 glycosylase protects neurons against oxidative
517 DNA damage and cell death under ischemic conditions. *J Cereb Blood Flow Metab*
518 **31**, 680-692, (2011).
519 35 Pouget, J. P., Douki, T., Richard, M. J. & Cadet, J. DNA damage induced in cells by
520 gamma and UVA radiation as measured by HPLC/GC-MS and HPLC-EC and Comet
521 assay. *Chem Res Toxicol* **13**, 541-549., (2000).
522 36 Ma, B. *et al.* Simultaneous determination of 8-oxo-2'-deoxyguanosine and 8-oxo-2'-
523 deoxyadenosine in human retinal DNA by liquid chromatography nanoelectrospray-
524 tandem mass spectrometry. *Sci Rep* **6**, 22375, (2016).
525 37 Patro, J. N., Haraguchi, K., Delaney, M. O. & Greenberg, M. M. Probing the
526 configurations of formamidopyrimidine lesions Fapy.dA and Fapy.dG in DNA using
527 endonuclease IV. *Biochemistry* **43**, 13397-13403, (2004).

528 38 Christov, P. P., Banerjee, S., Stone, M. P. & Rizzo, C. J. Selective Incision of the
529 alpha-N-Methyl-Formamidopyrimidine Anomer by Escherichia coli Endonuclease IV.
530 *J Nucleic Acids* **2010**, (2010).
531 39 Lukin, M. *et al.* Novel post-synthetic generation, isomeric resolution, and
532 characterization of Fapy-dG within oligodeoxynucleotides: differential anomeric
533 impacts on DNA duplex properties. *Nucleic Acids Res* **39**, 5776-5789, (2011).
534 40 Cui, L. *et al.* Comparative analysis of four oxidized guanine lesions from reactions of
535 DNA with peroxynitrite, singlet oxygen, and gamma-radiation. *Chem Res Toxicol* **26**,
536 195-202, (2013).
537 41 Kappler, L. *et al.* Linking bioenergetic function of mitochondria to tissue-specific
538 molecular fingerprints. *Am J Physio-Endocrin Metab* **317**, E374-E387, (2019).
539 42 Fernandez-Vizarra, E., Enriquez, J. A., Perez-Martos, A., Montoya, J. & Fernandez-
540 Silva, P. Tissue-specific differences in mitochondrial activity and biogenesis.
541 *Mitochondrion* **11**, 207-213, (2011).
542 43 Sun, S., Osterman, M. D. & Li, M. Tissue specificity of DNA damage response and
543 tumorigenesis. *Cancer Biol Med* **16**, 396-414, (2019).
544 44 Gospodinov, A., Ivanov, R., Anachkova, B. & Russev, G. Nucleotide excision repair
545 rates in rat tissues. *Eur J Biochem* **270**, 1000-1005, (2003).
546 45 Langie, S. A. *et al.* The effect of oxidative stress on nucleotide-excision repair in
547 colon tissue of newborn piglets. *Mutat Res* **695**, 75-80, (2010).
548 46 Markkanen, E. Not breathing is not an option: How to deal with oxidative DNA
549 damage. *DNA Repair* **59**, 82-105, (2017).
550 47 Malato, Y. *et al.* Fate tracing of mature hepatocytes in mouse liver homeostasis and
551 regeneration. *J Clin Invest* **121**, 4850-4860, (2011).

552 48 Aird, W. C. Phenotypic heterogeneity of the endothelium: II. Representative vascular
553 beds. *Circulation Research* **100**, 174-190, (2007).
554 49 Parmacek, M. S. & Epstein, J. A. Cardiomyocyte Renewal. *New Engl J Med* **361**, 86-
555 88, (2009).
556 50 Magrassi, L., Leto, K. & Rossi, F. Lifespan of neurons is uncoupled from organismal
557 lifespan. *Proc Natl Acad Sci USA* **110**, 4374-4379, (2013).
558 51 Aaltonen, K. E., Ebbesson, A., Wigerup, C. & Hedenfalk, I. Laser capture
559 microdissection (LCM) and whole genome amplification (WGA) of DNA from
560 normal breast tissue --- optimization for genome wide array analyses. *BMC Res Notes*
561 **4**, 69, (2011).
562 52 Aichler, M. & Walch, A. MALDI Imaging mass spectrometry: current frontiers and
563 perspectives in pathology research and practice. *Lab Invest* **95**, 422-431, (2015).
564 53 Brownlee, M. Biochemistry and molecular cell biology of diabetic complications.
565 *Nature* **414**, 813-820, (2001).
566 54 Rui, L. Energy Metabolism in the Liver. *Compr Physiol* **4**, 177-197, (2014).
567 55 Intano, G. W., Cho, E. J., McMahan, C. A. & Walter, C. A. Age-related base excision
568 repair activity in mouse brain and liver nuclear extracts. *J Gerontol Series A Biol Sci*
569 *Med Sci* **58**, 205-211, (2003).
570 56 You, C. *et al.* Roles of Aag, Alkbh2, and Alkbh3 in the Repair of Carboxymethylated
571 and Ethylated Thymidine Lesions. *ACS Chem Biol* **11**, 1332-1338, (2016).
572 57 Frischmann, M., Bidmon, C., Angerer, J. & Pischetsrieder, M. Identification of DNA
573 adducts of methylglyoxal. *Chem Res Toxicol* **18**, 1586-1592, (2005).
574 58 Yuan, B., Cao, H., Jiang, Y., Hong, H. & Wang, Y. Efficient and accurate bypass of
575 N2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase in vitro and in
576 vivo. *Proc Natl Acad Sci U S A* **105**, 8679-8684, (2008).

577 59 North, B. J. & Sinclair, D. A. The Intersection Between Aging and Cardiovascular
578 Disease. *Circul Res* **110**, 1097-1108, (2012).
579 60 Bernhard, D. & Laufer, G. The aging cardiomyocyte: a mini-review. *Gerontology* **54**,
580 24-31, (2008).
581 61 Strait, J. B. & Lakatta, E. G. Aging-associated cardiovascular changes and their
582 relationship to heart failure. *Heart Failure Clin* **8**, 143-164, (2012).
583 62 Zhang, H. & Forman, H. J. 4-Hydroxynonenal-mediated signaling and aging. *Free*
584 *Radic Biol Med* **111**, 219-225, (2017).
585 63 Kiyohara, C. & Ohno, Y. Sex differences in lung cancer susceptibility: a review.
586 *Gender Medicine* **7**, 381-401, (2010).
587 64 Zu, H. X. & Schut, H. A. Sex differences in the formation and persistence of DNA
588 adducts of 2-amino-3-methylimidazo[4,5-f]quinoline (IQ) in CDF1 mice.
589 *Carcinogenesis* **12**, 2163-2168, (1991).
590 65 Guo, J. *et al.* Development of a DNA Adductome Mass Spectral Database. *Chem Res*
591 *Toxixol* **33**, 852-854, (2020).
592 66 Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283
593 e219, (2020).
594 67 Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy
595 human tissues. *Mol Syst Biol* **15**, e8503, (2019).
596

597 ## Figure legends

598

599 **Figure 1. Adductomic analysis in two different sets of rat tissues.** Adductomics involves
600 two stages of data acquisition: (1) discovery of putative adducts by stepped MRM analysis of
601 pooled DNA samples and (2) subsequent targeted analysis of a high-confidence set of
602 adductomic signals representing putative DNA adducts in each tissue sample. (**a**) For the first
603 step of identifying putative DNA adducts, the stepped MRM signals must be curated to
604 remove artifacts, noise, and other spurious signals. Here we developed a decision tree using
605 different analytical filters to remove common salts, isotopomers, known artifacts, and
606 potential noise (signal intensity <2) to obtain a final data set of DNA adducts for subsequent
607 targeted analysis. The Venn diagram shows the final sets of 92 and 94 putative adducts
608 identified in discovery studies for tissues from two different rat cohorts, yielding a total set of
609 114 putative DNA adducts and modifications for subsequent quantitative analysis in each
610 sample. (**b**) Using the curated set of adducts, DNA adductome profiles were generated for
611 individual DNA samples from rat liver (**b**), kidney (**c**), brain (**d**), and heart (**e**) in rats ranging
612 in age from 1 to 26 months for rat group #1 (blue) and 4 to 26 months for rat group #2
613 (yellow) months. DNA adducts that were further identified using chemical standards are
614 labelled with names and CID transitions. Each DNA adduct, defined or putative, is
615 represented by its retention time (min, y-axis), mass-to-charge ratio (*m/z*, x-axis) and relative
616 abundance (size of the bubble). Signals from the four canonical DNA nucleobases are not
617 represented in the plot. Blue bubbles and arrows are DNA adduct signals observed in group
618 #1 while yellow are those observed in group #2.

619

620 **Figure 2. Known DNA damage products and modifications found or predicted among**
621 **adductomic signals.** (**a**) Using synthetic standards, adductomic signals were validated as the

622 DNA damage products *1,N*$^6$-εdA, *N*$^2$-CMdG, and 8-Oxo-dG (panel **c**), among others, as well
623 as the epigenetic marks 5-MdC and 5-HM-dC. (**b-d**) Oxidation of canonical 2'-
624 deoxynucleosides and DNA damage products leads to a host of secondary products predicted
625 to occur *in vivo,* including FAPy-dA (*m/z* 271.1→155.1), guanidinohydantoin *(m/z*
626 274.1→158.1), 8-nitro-dG (*m/z* 313.1→197.1), and oxazolone (*m/z* 247.1→131.1). See
627 **Supplementary Table 1** for adductomic signal data.
628
629 **Figure 3. DNA damage is strongly tissue- and gender-specific in rats**. (**a**) Venn diagram of
630 the putative DNA adducts shared among the four rat tissues. Blue and yellow numbers (X-Y)
631 correspond to the 92 putative adducts identified in group #1 rats and 94 in group #2 rats. Data
632 are from **Supplementary Table 1**. (**b**) Based on multivariate statistical analysis of tissue, sex,
633 and age associations of the 52 shared adducts in group #1 rats (**Extended Data Fig. 4a-c**), an
634 EVV model (ellipsoidal distributions with equal volume and variable shape and orientation
635 axes) revealed 5 underlying Gaussian distributions (Clusters I-V). The number of members in
636 each cluster is shown in parentheses. Members of Cluster III comprise a striking group of 5
637 heart samples from mainly older female rats (labeled with sample # and age). (**c-f**) Bar plots
638 of q-values for comparisons of putative DNA adduct levels in tissues from female versus male
639 group #1 rats: (**c**) brain (81 adducts), (**d**) heart (71 adducts), (**e**) kidney (71), and (**f**) liver (77).
640 q-Values were calculated from p-values obtained with the Wilcoxon rank sum test pn data
641 from **Supplementary Tables 1** and **2**. Dotted lines represent the false discovery rates (FDR)
642 at 10, 5, and 1%, with values in parentheses representing the number of significant adducts at
643 each threshold.
644
645 **Figure 4. Quantitative analysis of age-dependent DNA damage products in rat tissues.**
646 Isotope dilution chromatography-coupled triple quadrupole mass spectrometric analysis was

647 used to quantify 5-HMdC (**a, c, e, g**) and $N^2$-CMdG (**b, d, f, h**) levels in liver (**a, b**), kidney **(c,**
648 **d**), brain (**e, f**), and heart (**g, h**) from rats at 1-, 5-, 16- and 26-months-old. Box and whisker
649 plots show mean ± SEM for N=8. One-way ANOVA with Bonferroni's multiple comparisons
650 test was used to evaluate differences among the ages. *$p$<0.05, **$p$<0.01, ***$p$<0.001 and
651 **** $p$<0.0001 indicate statistically significant differences.
652
653 **Figure 5. Discovery of age-tracking DNA adducts in human heart.** (**a**) The bubble plot
654 shows the DNA adductome map in non-failing human heart. Mass spectrometry transitions
655 are noted above the names of adducts verified by chemical standards. Adducts that increase
656 with age are highlighted in red**.** (**b**) Venn diagrams showing overlap among the adductomic
657 signals detected in human and rat heart and brain. (**c,d**) Levels of 5-HMdC and 1, $N^6$ ε-dA, 5-
658 HMdC were quantified in 19 human non-failing heart samples from donors ranging in age
659 from 22- to 81-years-old by isotope dilution chromatography-coupled triple quadrupole mass
660 spectrometric analysis. Pearson's correlation test was used to establish correlation between 5-
661 HMdC and 1,$N^6$-εdA levels and age; *$p$<0.05 and **$p$<0.01. (**e-h**) Confocal imaging of
662 sectioned and immune-stained human non-failing heart samples reveals the presence of
663 scattered and sparse 1,$N^6$-εdA throughout cardiac tissue. 1,$N^6$-εdA is present within cardiac
664 cells, as seen by orange spots found within green (α-Actinin 2 labeled) areas. Red arrows
665 denote 1,$N^6$-εdA spots within nuclei (labeled with Hoechst), as shown in the merged figure
666 panel (**h**). Scale bar represents a length of 20 $\mu$m.
667
668 ## Methods
669
670 ***Materials.*** Tris-hydrochloride (Tris HCl), magnesium chloride ($MgCl_2$), acetic acid, sodium
671 acetate, formic acid, butylated hydroxytoluene (BHT), tetrahydrouridine (THU), coformycin,

672 DNAse I, benzonase, phosphatase alkaline, desferrioxamine mesylate salt, glyoxal, 4-Thio-2'-
673 deoxyuridine, 6-thio-2'-deoxyguanosine, 5-Methyl-2'-deoxycytidine, 2-Chloro-2'-
674 deoxyadenosine, 8-Hydroxy-2'-deoxyguanosine, 5-hydroxymethyl-2'-deoxyuridine, 5-Fluoro-
675 2'-deoxycytidine and 6-thio-2'-deoxyguanosine were obtained from Sigma-Aldrich (Saint-
676 Louis, MO, USA). Phosphodiesterase I was purchased from Affymetrix (Cleveland, Ohio,
677 USA). [$^{15}$N(U)]-labeled 2'-deoxyguanosine was purchased from Cambridge Isotopes
678 (Andover, MA, USA). Isopropanol and acetonitrile were purchased from VWR Scientific
679 (Franklin, MA, USA). Deionized water was further filtered through MilliQ systems (Millipore
680 corporation, Bedford, MA, USA) and used in the whole experiment. Anti-sarcomeric protein
681 α-Actinin 2 and ProLong Gold Antifade were purchased from Thermofisher Scientific
682 (Waltham, MA, USA) while 1, $N^6$ ε-dA Antibody and TrueBlack Plus Lipofuscin
683 Autofluorescence Quencher came from Novus Biological (Centennial, CO, USA) and
684 Biotium (Fremont, CA, USA) respectively. 5-Hydroxymethyl-2'-deoxycytosine, 1, $N^6$ ε-2'-
685 deoxyadenosine, 5-formyl-2'-deoxycytidine, 7-deaza-2'-deoxyxanthosine, 2'-deoxycytidine-5-
686 carboxylic acid sodium salt, $O^6$-nethyl-2'-deoxyguanosine, 8-oxo-2'-deoxyadenosine, $N^2$-
687 methyl-2'-deoxyguanosine and $N^6$-methyl-2'-deoxyadenosine were purchased from Berry &
688 Associates. 2'-Deoxyxanthosine, 8-chloro-2'-deoxyadenosine and 8-chloro-2'-
689 deoxyguanosine were purchased from Biosynth Carbosynth (San Diego, CA, USA). [$^3$D]-5-
690 HMdC, [$^{15}N_5$]-1, $N^6$ ε-dA, and [$^{15}N_5$]-8-oxo-2'-deoxyguanosine were purchased from Toronto
691 Research Chemical (Toronto, Canada). 5'-Ethynyl-2'-deoxycytidine and 5'-chloro-2'-
692 deoxycytidine were obtained from Cayman Chemical Company (Ann Arbor, MI, USA) and
693 Biolog (Hayward, CA, USA), respectively.
694
695 ***Synthesis of $N^2$-CMdG and [$^{15}$N(U)]-$N^2$-CMdG.*** For unlabeled $N^2$-CMdG, 1 mmol (0.267 g)
696 of 2'-deoxyguanosine or [$^{15}$N(U)]-labeled 2'-deoxyguanosine was reacted with 5 mmol of

697 glyoxal (0.290 g) in 1 M NaOAc/HOAc buffer pH 5.5 and the mixture was heated at 37 °C
698 overnight. The final $N^2$-CMdG product was purified from this mixture by HPLC with a
699 Hypersil GOLD C18 column (3.0 × 250 mm, 5 µm, Thermo Scientific) on an Agilent 1200
700 HPLC system. Binary mobile phase flow rate was 1 mL/min (A – 5mM ammonium acetate in
701 water, B – acetonitrile; increased to 40% of B for 20 min; increased to 90% of B in 0.1 min
702 and hold at 90% of B for 10 min; decreased to 0% of B in 1 min and re-equilibrate for 10
703 min). UV ($\lambda_{max}$, pH 7) = 260 nm. $[M+H]^+_{obs}$ = 326.1103, $[M+H]^+_{calc}$ = 326.1111, error =
704 2.39 ppm. MS/MS (ESI+): 210 (nucleobase + H), 164 (nucleobase – $CO_2$ + H), 117 (oxonium
705 ion of 2'-dR).
706
707 ***Animal procedures.*** All animal experiments were carried out by the local animal care and use
708 committee and respected the principals of animal experimentation (protocol number: 16-05-
709 271, National Institute on Aging). For the aging study, Brown Norway wild-type rats at 1, 5,
710 16 and 26 months old (4 females and 4 males per group of age) or 4, 9, 17 and 26 months old
711 (4 females and 4 males for groups at 9 and 17 months old; 8 males for groups at 4 and 26
712 months old) were purchased from the National Institute of Aging colony housed at Charles
713 River. All rats had *ad libidum* access to a standard chow diet (irradiated LabDiet Prolab
714 Isopro RMH 3000 5P75; LabDiet, St. Louis, MO) and water. All animals were maintained
715 under a 12-hour light/dark cycle at 22 °C. After two days acclimatization, all rats were
716 euthanized using isoflurane and cervical spine dislocation. Liver, kidneys, brain, and heart
717 were removed, washed with ice-cold PBS, and immediately plunged into liquid nitrogen and
718 stored at -80 °C.
719
720 For the transgenic mice, two models were used.: Reduced GLO1 activity in liver (n=11 and
721 18), brain (n=10 and 15) and heart (n=8 and 12) was modeled using Glyoxalase I Knock

722 Down (Glo1-KD) mice on a C57/BL6 background and their wild-type littermates[26]. Increased
723 GLO1 in brain (n=5 and 5) and heart (n=7 and 7) was modeled using Glyoxalase I over-
724 expression mice on a C57/BL6 background and their wild-type littermates[27]. Frozen tissues
725 were directly obtained from University of California San Diego. The ages for the knockdown
726 and the overexpressing mice were 217-273 and 180-292 days, respectively.
727
728 ***Human samples.*** Procurement of 19 frozen samples of non-failing left ventricular myocardial
729 tissue from human donors, deemed unsuitable for transplant, were obtained from University
730 of Pennsylvania (Philadelphia, PA USA)[68]. Samples were immediately stored at -80 °C until
731 DNA extraction. The ages of the 9 female and 10 male subjects ranged from 22 to 81 years.
732 Donor demographics are detailed in **Extended Data Table 1**. Samples of human brain tissue
733 from participants in the Religious Orders Study or Rush Memory and Aging Project
734 (ROSMAP) without cognitive impairment (control subjects, CT) ranging in age from 71 to 89
735 years were obtained from the Rush Alzheimer's Disease Center (Chicago, IL, USA)[69] and
736 were stored at -80 °C. Both studies were approved by an Institutional Review Board of Rush
737 University Medical Center. Each participant signed an informed consent. Anatomic Gift Act,
738 and Repository Consent to allow their data to be repurposed. NCI was defined as previously
739 reported[70]. Donor demographics and definitions of clinical and neuropathological criteria are
740 detailed in **Extended Data Table 1.**
741
742 ***DNA extraction, purification, and hydrolysis.*** DNA extraction from rat organs (livers,
743 kidneys, brains, and hearts), mouse organs (livers, brains, and hearts) and human hearts and
744 brains was performed with a commercial kit (Sigma-Aldrich, 11814770001) according to the
745 manufacturer's protocol, with the addition to all solutions of appropriate inhibitors and
746 antioxidants to minimize adventitious damage: 5 µg/ml coformycin, 50 µg/ml

747 tetrahydrouridine (THU), 100 μM desferrioxamine, 100 μM butylated hydroxytoluene
748 (BHT)[16]. DNA was then quantified by spectrophotometry at 260 nm (NanoDrop Lite
749 Spectrophotometer, Thermofisher scientific). Samples were finally stored at -80 °C.
750
751 For DNA adductome analysis, a mix of genomic DNA (containing 5 μg of each sample to
752 cover adductome in every age) was dried by using speed vacuum and then reconstituted in 90
753 μL of 10 mM Tris-HCl buffer pH 8, 1 mM $MgCl_2$, 10 μg/mL coformycin, 50 μg/mL THU, 1
754 mM desferrioxamine, 1 mM BHT and digested using 10 U benzonase, 5 U DNAse I, 17U
755 alkaline phosphatase, 0.1 U phosphodiesterase I. A blank sample was also prepared using the
756 same procedure without the DNA. The digestion was allowed to occur overnight at 37 °C.
757 Reactions were then worked up using 10 KDa exclusion filters and centrifuged at 12,800 × g
758 for 10 min to remove proteins and then analyzed by LC-MS/MS.
759
760 For adduct quantification analysis, 20 μg of genomic DNA was dried by using speed vacuum
761 and then reconstituted in 27 μL containing 10 mM Tris-HCl buffer pH 8, 1 mM $MgCl_2$, 10
762 μg/mL coformycin, 50 μg/mL THU, 1 mM desferrioxamine, 1 mM of BHT and digested
763 using 10 U benzonase, 5 U DNAse I, 17 U alkaline phosphatase, 0.1 U phosphodiesterase I.
764 An additional 3 μL of stable isotope-labeled internal standards were then added as needed,
765 including 5 nM [$^{15}N_5$]-$N^2$-CMdG 1 μM [$D^3$]-5-HMdC and 500 pM [$^{15}N_5$]- 1, $N^6$-ε-dA. The
766 digestion was allowed to occur overnight at 37 °C. Reactions were then processed using 10
767 KDa exclusion filters and centrifuged at 12,800×g for 10 min to remove proteins, followed by
768 analysis by LC-MS/MS.
769
770 ***DNA adductome analysis: Adduct discovery in pooled DNA samples.*** Adductome analysis
771 was performed on an HPLC Agilent 1290 Infinity II with an added Diode Array Detector

772 (DAD) coupled with the Agilent 6490 (Group #1 rats, mice, humans) or 6495 (Group #2 rats)
773 triple quadrupole mass spectrometer with an electrspray ion source operated in positive ion
774 mode with the following source parameters: drying gas temperature 200 °C with a flow of 12
775 L/min, nebulizer gas pressure 30 psi, sheath gas temperature 300 °C with a flow of 12 L/min,
776 ESI capillary voltage 4000 V and nozzle voltage 0 V. A Kinetex EVO C18 column (2.1 × 150
777 mm, 2.6 µm, Phenomenex) was used for chromatographic separation (column temperature 30
778 °C). Binary mobile phase flow rate was 0.400 mL/min (A – 0.1% formic acid in water, B –
779 0.1% formic acid in acetonitrile; delivered at 0% of B for 2 min; increased to 16% of B for 15
780 min; increased to 80% of B in 1 min and hold at 80% of B for 5 min; decreased to 0% of B in
781 1 min and re-equilibrate for 5 min). For 5-HMdC quantification, binary mobile phase was the
782 following: A – water, B – 5 mM of Ammonium Acetate in water, pH 5.3; delivered at 0% of
783 B for 2 min; increased to 16% of B for 15 min; increased to 80% of B in 1 min and hold at
784 80% of B for 5 min; decreased to 0% of B in 1 min and re-equilibrate for 5 min). The effluent
785 from the first min from the LC system was diverted to waste to minimize the contamination of
786 the ESI source. The strategy was designed to detect the neutral loss of 2'-deoxyribose from
787 positively ionized 2'-deoxynucleoside adducts by monitoring the samples transmitting their
788 $[M+H]^+ \rightarrow [M+H-116]^+$ transitions. For both the mix of DNA and the blank, 300 MRM
789 transitions were monitored from transition *m/z* 225→109 to transition *m/z* 525→409 with a
790 collision energy of 10 eV. Ten microliters of both samples were injected six times to cover
791 the monitoring of 300 MRM transitions. Each of 8 methods covering the 300 MRM
792 transitions has a final cycle length/ scan time of 790.4 ms (1.27 cycles/s). Transitions of
793 canonical 2'-deoxynucleosides, including *m/z* 228→112 ($[dC+H]^+$), *m/z* 243→127 ($[dT+H]^+$),
794 *m/z* 252→136 ($[dA+H]^+$) and *m/z* 268→152 ($[dG+H]^+$), were not monitored in this
795 adductome analysis. Finally, the 300 MRM signals were compared between the mix of DNA
796 and the blank. Each signal that was present in the mix of DNA sample but not in the blank

797 was collected for further investigation. These stepped MRM signals were normalized to
798 account for different amounts of injected DNA with each LC-MS run by dividing the mass
799 spectrometer signal intensity by the sum of the absorbance values from an in-line diode array
800 UV detector for the 4 canonical 2'-deoxyribonucleosides. The resulting normalized signal
801 intensities, detailed in **Supplementary Table 1a** for both groups of rats, were curated to
802 remove artifacts, noise, and other spurious signals using the decision tree detailed in **Figure**
803 **1a** to remove common salts, isotopomers, known artifacts, and potential noise (signal
804 intensity <2). The resulting relative peak intensity of these signals are detailed in
805 **Supplementary Table 1e** and plotted as a bubble chart in which the x-axis was the *m/z* and
806 the y-axis was retention time (**Fig. 1**).
807
808 ***DNA adductome analysis: Adduct quantification in individual DNA samples.*** The final set
809 of 114 putative DNA adducts and modifications detailed in **Supplementary Table 1e** was
810 used for targeted analysis of individual tissue samples from rats. Portions of isolated DNA
811 used for the adduct discovery described earlier were analyzed again with an MRM table for
812 the 114 putative DNA adducts and modifications and the same LC-MS/MS conditions. The
813 resulting signals were normalized by dividing by the sum of the UV signals for the four
814 canonical 2'-deoxyribonucleosides, as described earlier. The normalized signal intensities for
815 both groups of rats are detailed in **Supplementary Table 4**.
816
817 ***Identification of DNA damage products with standards.*** To tentatively identify the structures
818 of adductomic signals, a cocktail of 2'-deoxyribonucleoside standards (10 nM each in 30 μL
819 of water) was analyzed by LC-MS/MS using the same conditions noted earlier for adductomic
820 analysis. The goal here was to correlate retention times and MS/MS transitions for the
821 standards with the adductomic data. The following transitions were monitored: *m/z* 326 →

822 210 $N^2$-carboxymethyl-2'-deoxyguanosine, $N^2$-CMdG), *m/z* 258 → 142 (5-hydroxymethyl-2'-
823 deoxycytidine, 5-HMdC), *m/z* 276 → 160 (1,$N^6$-etheno-2'-deoxyadenosine, 1,$N^6$ ε-dA), *m/z*
824 242 → 126 (5-methyl-2'-deoxycytidine, 5-MdC), *m/z* 245 → 129 (4-thio-2'-deoxyuridine, 4-
825 thio-dU), *m/z* 252 → 136 (3,$N^4$-etheno-dC, 3, $N^4$-ε-dC), *m/z* 256 → 140 (5-formyl-2'-
826 deoxycytidine, 5-formyl-dC), *m/z* 262 → 146 (5'-chloro-2'-deoxycytidine, 5-CldC), *m/z* 268
827 → 152 (7-deaza-2'-deoxyxanthosine, 7-deaza-dX), *m/z* 268 → 152 (8-hydroxy-2'-
828 deoxyadenosine, 8-Oxo-dA), *m/z* 269 → 153 (2'-deoxyxanthosine, dX), *m/z* 274 → 158
829 (guanidinohydantoin, Gh), *m/z* 282 → 166 ($O^6$-methyl-2'-deoxyguanosine, $O^6$-MdG), *m/z* 282
830 → 166 ($N^2$-methyl-2'-deoxyguanosine, $N^2$-MdG), *m/z* 284 → 168 (8-hydroxy-2'-
831 deoxyguanosine, 8-Oxo-dG), *m/z* 284 → 168 (6-thio-2'-deoxyguanosine, S6-dG), *m/z* 294 →
832 178 (2'-deoxycytidine-5'-carboxylic acid sodium salt, dC carboxylic acid), *m/z* 300 → 184
833 (spiroiminodihydantoin, Sp), *m/z* 304 → 188 (pyrimidopurinone adduct of 2'-
834 deoxyguanosine, M1-dG), *m/z* 286 → 170 (8-chloro-2'deoxyadenosine, 8-Cl-dA), *m/z* 302 →
835 186 (8-chloro-2'-deoxyguanosine, 8-Cl-dG), *m/z* 286 → 170 (2-chloro-2'-deoxyadenosine, 2-
836 Cl-dA) and *m/z* 246 → 130 (5-fluoro-2'-deoxycytidine, 5-F-dC).
837
838 ***Identification of the cyclobutene dimer of dC.*** The identity of the putative DNA adduct with
839 *m/z* 455 → 339 was achieved by multi-stage mass spectrometry ($MS^n$) using an HPLC-
840 coupled Thermo Orbitrap ID-X Tribrid mass spectrometer coupled with a Vanquish LC
841 (Thermo Fisher). Five microliters of sample were injected on a Kinetex C18 column (150 ×
842 2.1 mm, Phenomenex) eluted with mobile phase A (0.1% formic acid) and mobile phase B
843 (0.1% formic acid in acetonitrile) using the following gradient: 2 min at 0% then to 18% B in
844 13 min and to 60% B in 2 min. The original conditions were re-established in 1 min and the
845 column re-equilibrated for 5 min. $MS^1$ data were acquired at 120,000 resolution in the
846 orbitrap with a 70% RF lens. $MS^2$ data were acquired on the targeted *m/z* identified

847 previously, using assisted HCD spanning 20, 35, and 60 normalized collision energy, and
848 assisted CID at 15,30,45 %. $MS^2$ fragments were measured in the orbitrap at 60,000
849 resolution. $MS^2$ fragments with >5% relative intensities were selected for $MS^3$, with a
850 fragmentation by CID at 30%, and analysis in the orbitrap at 30,000 resolution. All data were
851 analyzed in xcalibur qual browser, with some of the fragmentation additionally analyzed in
852 Massfrontier (V8.0, Thermo fisher).
853
854 ***DNA adduct quantification by isotope dilution LC-MS/MS.*** For absolute quantification of
855 specific DNA modifications and damage products, isotope dilution LC-MS/MS was
856 performed on isolated DNA used for adductomic analyses. Ten microliters of each sample
857 were injected into the LC-MS/MS instrument using the same LC-MS/MS parameters noted in
858 "DNA adductome analysis". Characteristic reactions and collision energies for DNA adducts
859 of interest are as follows (*m/z* precursor ion → *m/z* product ion, collision energy (eV),
860 retention time): $N^2$-CMdG (*m/z* 326→*m/z* 210, 10, 3.45min), [$^{15}N_5$]-$N^2$-CMdG (*m/z* 331→*m/z*
861 215, 10, 3.45min), 5-HMdC (*m/z* 258→*m/z* 242, 10, 2.72min), [$D^3$]-5-HMdC (*m/z* 261→*m/z*
862 245, 10, 2.71min), 1, $N^6$-ε-dA (*m/z* 276→*m/z* 160, 10, 2.93min), [$^{15}N_5$]- 1, $N^6$ ε-dA (*m/z*
863 281→*m/z* 165, 10, 2.93min), 8-Oxo-dG (*m/z* 284→*m/z* 170, 10, 6.65min), and [$^{13}C$,$^{15}N_2$]-8-
864 Oxo-dG (*m/z* 286→*m/z* 170, 10, 6.65min). Calibration curves for the labeled and unlabeled
865 forms were constructed by plotting the MRM signal ratios between the labeled and unlabeled
866 forms against their corresponding concentration ratios, as illustrated in **Extended Data**
867 **Figure 1**. Quantification for $N^2$-CMdG, 5-HMdC, 1, $N^6$-ε-dA and 8-Oxo-dG were achieved
868 using the ratio of peak areas of the unlabeled and labeled DNA adducts. Raw mass
869 spectrometric data were analyzed by MassHunter Qualitative Analysis software (Agilent
870 Technologies, USA).
871

872 ***Immunostaining and imaging.*** To prepare samples for immunostaining, samples were
873 dissected from patients and stored at -80 °C freezer prior to transferring to Tissue-Tek optimal
874 cutting temperature (OCT) compound for cryosection. Slides with sectioned samples were
875 then fixed with ice-cold 100% methanol for 15 min at 4 °C, washed 3 times with PBS, and
876 blocked for 1 h with a 5% goat serum, 0.3% Triton X-100 and PBS solution. Primary
877 antibodies were then applied to samples in antibody buffer (1% BSA and 0.3% Triton X-100
878 in PBS), overnight at 4 °C. For cardiomyocyte staining, anti-sarcomeric protein α-actinin 2
879 was applied at 1:250 or 2 μg/mL, and 1,$N^6$-εdA antibody was applied at 1:100, in antibody
880 buffer. Following five washes in PBS 0.3% Triton X-100, secondary antibodies were then
881 applied for 1hr in antibody buffer, at 1:1000. α-Actinin 2 was detected with Goat anti-Rabbit
882 IgG conjugated to Alexa 488, and 1,$N^6$-ε-dA was detected with Anti-Mouse IgG conjugated
883 to Alexa 555. To detect nuclei, Hoechst 33342 was applied for 10 min following secondary
884 antibody removal, at 1:10,000 in a 0.3% Triton-PBS. Slides were then washed 5 times in
885 0.3% Triton-PBS, washed two times in PBS, and TrueBlack Plus Lipofuscin
886 Autofluorescence Quencher was applied for 10 min to reduce autofluorescence. Finally,
887 TrueBlack was washed 3 times with PBS, and slides were mounted in ProLong Gold
888 Antifade. Images were captured with an LSM 980 Confocal Microscope (Zeiss), using the
889 Airyscan mode and a 63x oil objective. Z-stacks were captured across nine tiles, to sample
890 larger areas for image analysis.
891

892 ***Image analysis.*** To detect 1,$N^6$-ε-dA in cardiac cell nuclei within sectioned samples, we
893 performed image analysis on captured confocal images using arivas Vision 4D software
894 (arivas). We used the machine learning module to detect cardiac cell networks identified
895 by α-actinin 2 staining, the thresholding segmentation module to detect nuclei by
896 Hoechst 33342 staining, and the blob finder module to detect 1,$N^6$-εdA spots. Then we
897 used the compartments module to detect nuclei within cardiac cells networks and 1,$N^6$-
898 εdA within these nuclei. Selected nuclei were encompassed by greater than 70% of the
899 cardiac signal, and 1,$N^6$-εdA puncta were selected when they were encompassed by
900 greater than 50% nuclei signal in selected nuclei.
901
902 ***Statistical analyses.*** All statistical analyses were performed using either GraphPad Prism
903 (version 8.0) or R (http://www.r-project.org). Data are presented as means ± SEM (standard
904 error of the mean). Each biological sample was prepared at least two times to confirm
905 observation. Statistical significance was determined using one-way ANOVA with
906 Bonferroni's multiple comparisons test for rat study and a two-sided unpaired Student's *t*-test
907 to compare wild-type (WT) to KD/OE mice samples. Finally, we used Pearson's correlation
908 in human non-failing heart samples. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p <$
909 $0.0001$ was considered to indicate statistically significant differences. All statistical analyses
910 were performed using GraphPad Prism version 8.0 (GraphPad Software).
911
912 We traced differences of DNA adducts in rat sex and age using two-sided Wilcoxon rank sum
913 test (equivalent to the Mann-Whitney U test; wilcox.test function of stats package in R). The
914 false discovery rate (FDR), the rate of type I error associated with rejecting the null
915 hypothesis due to multiple comparisons, was estimated using qvalue function in the qvalue
916 package[71] in R.

917

918 To identify the most strongly covarying DNA adducts, we performed principal component
919 analysis (PCA) on shared adducts in rat samples (**Figure 3a, Extended Data Figure 3c**) in
920 both group, that is the most commonly used method for reduction of dimensionality[72], using
921 the prcomp function (stats package in R). To find underlying patterns, we used mixture model
922 clustering that is a probability-based approach in which we assume the data set is best
923 described as a mixture of probability models. We employed the mclust package[73] for
924 Gaussian mixture modeling (GMM), the most used mixture model-based clustering method,
925 to determine the underlying Gaussian mixture distributions using first two PC scores covering
926 62,71% variation of the data (**Extended Data Fig. 4c,f**).

927

928 To assess association between 162 DNA adducts, we estimated Spearman's rank correlation
929 coefficient ($\rho$) after zero filling the empty cells (**Supplementary Table 2**). Results are show
930 as a heatmap reordered by hierarchical clustering analysis (HCA) obtained by the R pvclust
931 package[74] after 3000 iterations of multiscale bootstrap resampling.

932

933 ## Methods References


934 9 Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from
935 deamination, oxidation and reaction with products of lipid peroxidation by liquid
936 chromatography isotope dilution tandem mass spectrometry. *Nat Protoc* **3**, 1287-1298,
937 (2008).

938 68 Yingxian Chen, C. *et al.* Suppression of detyrosinated microtubules improves
939 cardiomyocyte function in human heart failure. *Nat Med* **24**, 1225-1233, (2018).

940 69 Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J*
941 *Alz Dis* **64**, S161-S189, (2018).

70 Bennett, D. A. *et al.* Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837-1844, (2006).

71 Storey, J. D. A direct approach to false discovery rates. *J Royal Stat Soc Series B* **64**, 479-498, (2002).

72 Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat Meth* **14**, 849-863, (2017).

73 Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289-317, (2016).

74 Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542, (2006).

## Acknowledgments

The authors gratefully acknowledge support from grant R01-AG063341 (PIs: Lee and Dedon) from the National Institute on Aging of the National Institutes of Health and from Center grant P30-ES002109 from the National Institute of Environmental Health Sciences of the National Institutes of Health. ROSMAP is supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, U01AG61356 (to D.A.B.). ROSMPA resources can be requested at http://www.radc.rush.edu.

## Conflict of interest

The authors have no conflicts of interest to declare.

## Author contributions

P. Dedon and R. Lee conceived of the study, supervised all aspects of the study, and interpreted data. A. Guilbaud, F. Ghanegolmohammadi, Y. Wang, A. Kreymerman, J. Gamboa-Varela, J. Leng, J. Garbern, E. Ricci-Blair, H. Elwell, C. Vidoudez, F. Cao, S. Balamkundu, M. DeMott and Cui L. conducted the experiments, analyzed, and interpreted the data. K. Bedi, K. Margulies, D. Bennett, A. Palmer and A Barkley-Levenson participated in study design and provided key biological specimens. A. Guilbaud, F. Ghanegolmohammadi, Y. Wang, and A. Kreymerman drafted the manuscript and all authors contributed to reviewing and revising the manuscript.

## Data availability statement

Mass spectrometry data that support the findings of this study have been deposited in the Chorus Project (https://chorusproject.org/pages/index.html) with the accession code 1767.

## Competing Interests Statement

A. Palmer has a patent for the methods and use of Glo1 inhibitors (https://patents.google.com/patent/US11235020B2/en).

## Corresponding Author

Correspondence and requests for materials should be addressed to Peter Dedon at pcdedon@mit.edu.

## Supplementary Information

Supplementary Information is available for this paper:

991 **Supplementary Table 1.** DNA adduct discovery: Normalized adductomic data for age-
992 pooled samples in rat tissues from two independent experiments (rat groups #1 and #2).
993 Separate sheets for each step of data filtering.
994 **Supplementary Table 2.** Analysis of adductomic data in individual DNA samples in rat
995 group #1 as a function of sex and age.
996 **Supplementary Table 3.** DNA adduct levels in mouse, rat, and human tissues measured by
997 isotope dilution LC-MS/MS.
998 **Supplementary Table 4**. Normalized mass spectrometric signals for individual samples of rat
999 tissues from rat groups #1 and #2.
1000 **Supplementary Table 5**. DNA adduct discovery and normalized signal data for human brain
1001 and heart samples.
1002

1003 ## Extended Data Figure Legends

1004

1005 **Extended Data Fig. 1.** Parameters and data for isotope-dilution chromatography-coupled
1006 triple quadrupole mass spectrometric quantification of DNA adducts in rat and human tissues.
1007 (**a**) Representative extracted ion chromatograms for analysis of $N^2$-CMdG, 5-HMdC,8-oxo-
1008 dG, and 1,$N^6$-εdA and their isotope-labeled counterparts in rat and human tissues. (**b**)
1009 Illustrative calibration curves of the canonical 2'-deoxyribonucleosides and several DNA
1010 adducts. (**c**) LC-MS/MS parameters for known DNA adducts and modifications analyzed
1011 here. These data pertain to **Figures 4** and **5**.
1012

1013 **Extended Data Figure 2.** High–resolution $MS^4$ Orbitrap analysis reveals the structure of the
1014 putative DNA adduct at *m/z* 455. *Left to right*: The HPLC chromatogram shows retention time
1015 of 0.89 min. The mass spectrum shows an *m/z* value of 455.18794 for the protonated species,

1016 which equates to an exact mass of 454.1801 for the adduct. The $MS^4$ fragmentation series
1017 shows progressive loss of two cytosine-sized fragments and is consistent with a cyclobutane
1018 dimer of dC. These data pertain to **Figure 1** and **Supplementary Table 1**.
1019
1020 **Extended Data Figure 3.** DNA damage is strongly tissue-specific in rats. Comparison of
1021 normalized LC-MS signal intensities for 92 putative DNA adducts in 4 tissues in group #1
1022 rats (blue numbers in upper circles and lower labels) and 94 in group #2 rats (yellow
1023 numbers). Rank order was established relative to brain from group #1 (left-most plot) with
1024 comparisons to heart, kidney, and liver for group #1 and group #2 rats. For each tissue,
1025 numbers in overlapping circles correspond to unique adducts in each group of rats (blue for
1026 group #1, yellow for group #2) and shared adducts (black numbers), with matching colors of
1027 upper circles and lower labels. These data pertain to **Figures 1** and **3.**
1028
1029 **Extended Data Figure 4.** Multivariate statistical analysis of 52 (**a-c**) and 64 (**d-f**) shared
1030 DNA adducts in group #1 and group #2 rat tissues reveals 5 co-varying clusters based on age,
1031 sex, and tissue. (**a, d**) Principal components analysis identifies 2 components that account for
1032 63% of the variance. Black bars (left axis) indicating the contribution ratio (proportion of
1033 variance), red circles (right axis) indicating the cumulative proportion of variance (cumulative
1034 contribution ratio [CCR]), and horizontal dashed lines (right axis) indicating CCRs of 60%
1035 and 90%. (b, e) Bayesian analysis to define the number of components for Gaussian mixture
1036 model (GMM) clustering. The number of components, describing the underlying Gaussian
1037 distributions, was defined based on Bayesian information criterion values of models with
1038 differing parametrizations. The best fitting model (Ellipsoidal distributions with equal volume
1039 and variable shape and orientation axes, EVV) had 5 components that were used for GMM
1040 clustering. (**c, f**) Membership probability matrix based on GMM clustering. Each value shows

1041 the posterior probability based on the model in panels **b** and **e**. The posterior probability
1042 describes the likelihood of each sample belonging to each cluster. Logarithmic transformation
1043 of conditional probabilities from expectation maximization was used to generate the heatmap.
1044 The dendrogram illustrates model-based hierarchical agglomerative clustering based on the
1045 Gaussian probability model for maximizing the resulting likelihood. This figure is associated
1046 with **Figure 3b, h.**
1047
1048 **Extended Data Figure 5.** Age-biased DNA adducts in four rat tissues in two different sets of
1049 rats. **(a, d)** Heatmaps show pairwise comparisons of all ages in each tissue. Significant
1050 changes (FDR = 10%) are presented as red (increase) and green (decrease). Bar plot on the
1051 top shows the frequency in each compared age. Bar plot (black) on the right shows frequency
1052 in each adduct. (**b**, **e**) Bar plots of overall frequency of significant changes (FDR = 10%) in
1053 each tissue. (**c**, **f**) Bar plots of overall frequency of significant changes (FDR = 10%) in each
1054 age-pairwise comparison. These data pertain to **Figures 1** and **4,** and **Supplementary Table**
1055 **5.**
1056
1057 **Extended Data Figure 6. Analysis of DNA damage products in rat, mouse, and human**
1058 **tissues.** (**a-j**) Isotope dilution chromatography-coupled triple quadrupole mass spectrometric
1059 analysis of DNA adducts in rat and mouse tissues. 1,$N^6$-εdA levels in rat tissues as a function
1060 of age: liver (**a**), kidney (**b**), brain (**c**), and heart (**d**). Data are mean ± SEM for $n$ = 8 rats.
1061 One-way ANOVA with Bonferroni's multiple comparisons test was used to evaluate
1062 difference among ages., *$p$<0.05, **$p$<0.01, ***$p$<0.001, ****$p$<0.0001. (**e**) 8-Oxo-dG levels
1063 as a function of age in rat liver. Box and whisker plots show mean (heavy bar) and SEM
1064 (whiskers). One-way ANOVA with Bonferroni's multiple comparisons test was used to
1065 evaluate differences among ages, *$p$<0.05, **$p$<0.01 and **** $p$<0.0001. Associated with

1066 **Figure 3**. (**f-j**) Glyoxalase I levels do not affect the level of $N^2$-CMdG in transgenic mice. $N^2$-
1067 CMdG levels were measured in liver (**f, i**), brain (**g, j**), and heart (**h**) from glyoxalase I
1068 knockdown (KD) mice (**f-h**) and from glyoxalase I over-expressing (OE) mice (**i, j**). Box and
1069 whisker plots show mean (heavy bar) and SEM (whiskers). A two-tailed unpaired Student's *t*-
1070 test was used to evaluate the difference between wild-type (WT) and transgenic mice (KD,
1071 OE). ** $p < 0.01$, *ns*: not-significant (**f-h**). Plot of normalized MS signal intensities for
1072 putative $N^2$-CE-dG (*m/z* 340-224) and validated $N^2$-CMdG for rat and human tissues. Data
1073 represent mean for $n$=8 measurements of male and female tissues. This figure is associated
1074 with **Figure 3**.
1075

1076 ## Extended Data Tables

1077

1078 **Extended Data Table 1.** Characteristics of the human tissue donors**.** ***Upper: heart tissue.***
1079 LV: Left Ventricle, HMI: Heart Mass Index, LVMI: Left Ventricular Mass Index, BSA: Body
1080 Surface Area, BMI: Body Mass Index, LVEDd: Left Ventricle End-Diastolic dimension,
1081 LVEDs: Left Ventricle End-Systolic dimension, PW: Posterior Wall, LVEF: Left Ventricle
1082 Ejection Fraction. ***Lower: brain tissue.*** Donors are represented according to their ID number
1083 (#), category (control, CT), age at death and gender. $n$= 10 CT. These data pertain to **Figure**
1084 **5.**

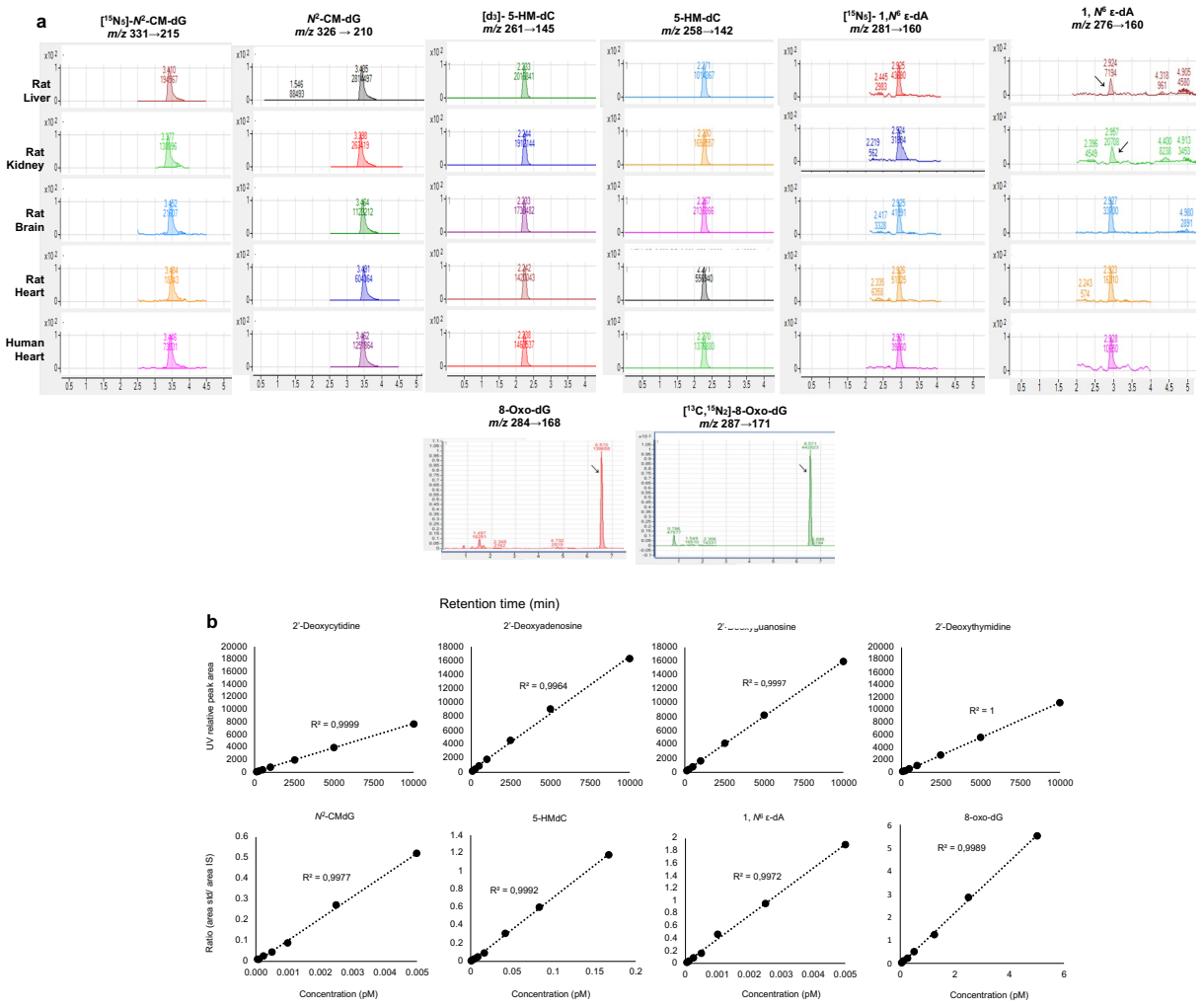
# Extended Data

*Tissue- and sex-specific DNA damage tracks aging in rats and humans*

Axel Guilbaud, Farzan Ghanegolmohammadi, Yijun Wang, Jiapeng Leng, Alexander Kreymerman, Jacqueline Gamboa Varela, Jessica Garbern, Hannah Elwell, Fang Cao, Elisabeth M. Ricci-Blair, Cui Liang, Seetharamsingh Balamkundu, Charles Vidoudez, Michael S. DeMott, Kenneth Bedi, Kenneth B. Margulies, David A. Bennett, Abraham A. Palmer, Amanda Barkley-Levenson, Richard T. Lee, Peter C. Dedon

***Contents***

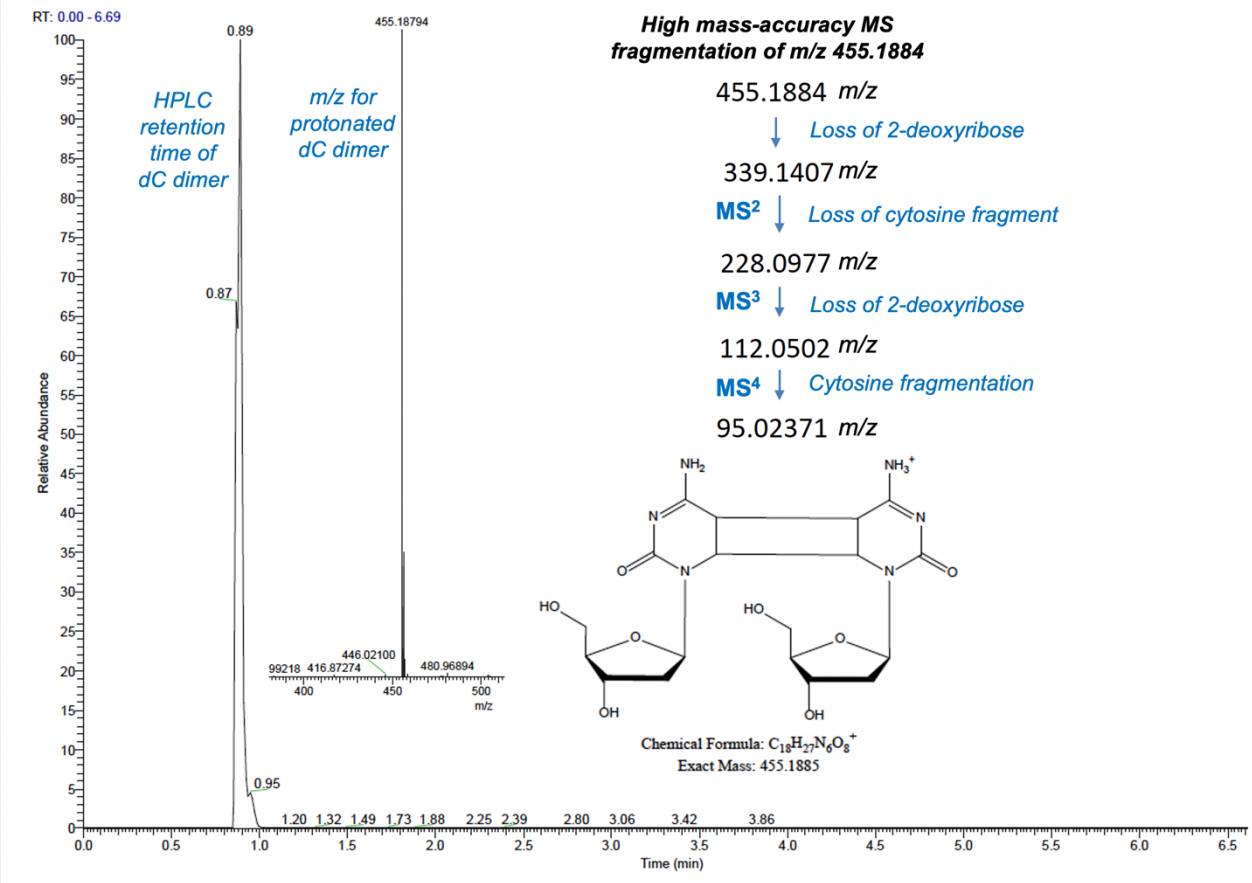
Extended Data Figures 1-6
Extended Data Table 1



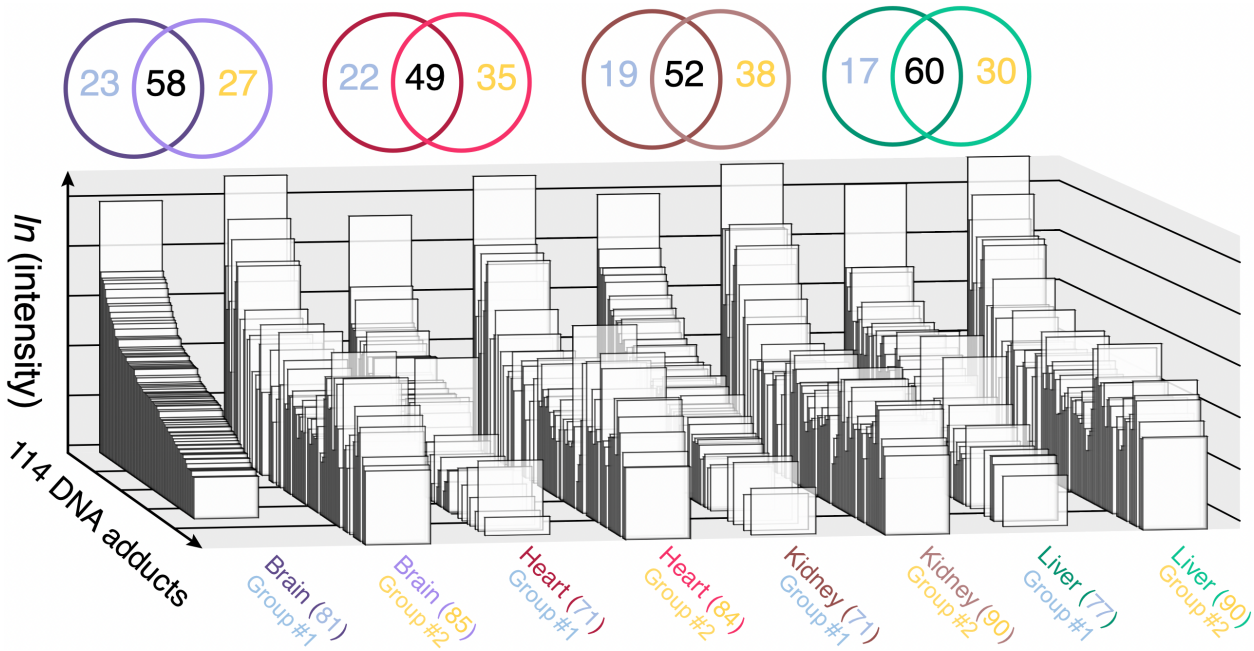

c

| | DNA modification name | Abbreviation | Mass, Da | [H]+, m/z | [M+H-116]+, m/z | RT (min) | Corresponds to Adduct? |
|---|---|---|---|---|---|---|---|
| Cytidines | 5-Methyl-2'-deoxycytidine | 5-MdC | 241 | 242 | 126 | 1.3 | YES |
| | 3,$N^4$-Etheno-2'-deoxycytidine | 3,$N^4$εdC | 251 | 252 | 136 | 2.2 | YES |
| | 5-Formyl-2'-deoxycytidine | 5-FdC | 255 | 256 | 140 | 4.6 | YES |
| | 5-Chloro-2'-deoxycytidine | 5-CldC | 261 | 262 | 146 | 4.2 | NO |
| | 5-Carboxy-2'-deoxycytidine sodium salt | 5-CadC $Na^+$ | 293 | 294 | 178 | 2.1 | NO |
| | 5-Hydroxymethyl-2'-deoxycytidine | 5-HMdG | 257 | 258 | 142 | 2.2 | YES |
| Guanosines | $N^2$-Carboxymethyl-2'-deoxyguanosine | $N^2$ CMdG | 325 | 326 | 210 | 3.4 | YES |
| | Spiroiminodihydantoin | Sp | 299 | 300 | 184 | 0.9 | NO |
| | Guanidinohydantoin | Gh | 273 | 274 | 158 | 4.2 | NO |
| | 2'-Deoxyxanthosine | dX | 268 | 269 | 153 | 5.7 | NO |
| | Purimidopurinone of 2'-deoxyguanosine | M1dG | 303 | 304 | 188 | 8.8 | YES |
| | $O^6$-Methyl-2'-deoxyguanosine | $O^6$MdG | 281 | 282 | 166 | 9.1 | NO |
| | $N^6$-Methyl-2'-deoxyguanosine | $N^6$ MdG | 281 | 282 | 166 | 7.3/ 8.1 | YES |
| | 8-Oxo-2'-deoxyguanosine | 8-oxodG | 283 | 284 | 168 | 6.6 | YES |
| | 8-Chloro-2'-deoxyguanosine | 8-CldG | 301 | 302 | 186 | 9.0 | NO |
| Adenosines | 2-Chloro-2'-deoxyadenosine | 2-CldA | 285 | 286 | 170 | 8.3 | NO |
| | 1,$N^6$Etheno-2'-deoxyadenosine | 1,$N^6$εdA | 275 | 276 | 160 | 2.9 | YES |
| | 8-Oxo-2'-deoxyadenosine | 8-oxodA | 267 | 268 | 152 | 6.9 | YES |
| | 8-Chloro-2'deoxyadenosine | 8-CldA | 285 | 286 | 170 | 10.4 | NO |

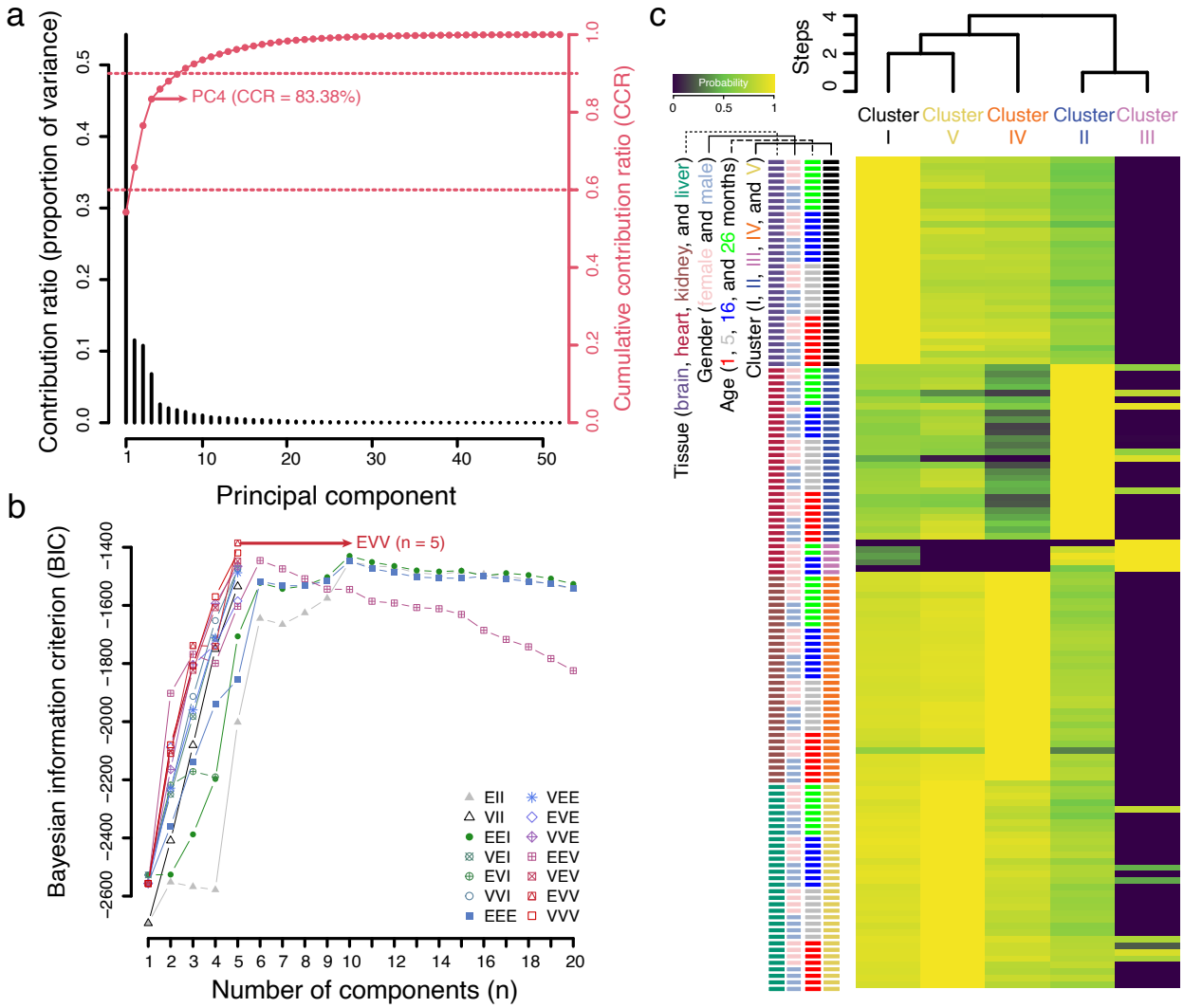
**Extended Data Fig. 1. Parameters and data for isotope-dilution chromatography-coupled triple quadrupole mass spectrometric quantification of DNA adducts in rat and human tissues.** (**a**) Representative extracted ion chromatograms for analysis of $N^2$-CMdG, 5-HMdC,8-oxo-dG, and 1,$N^6$-εdA and their isotope-labeled counterparts in rat and human tissues. (**b**) Illustrative calibration curves of the canonical 2'-deoxyribonucleosides and several DNA adducts. (**c**) LC-MS/MS parameters for known DNA adducts and modifications analyzed here. These data pertain to **Figures 4** and **5**.




**Extended Data Figure 2. High–resolution MS$^4$ Orbitrap analysis reveals the structure of the putative DNA adduct at *m/z* 455.** *Left to right*: The HPLC chromatogram shows retention time of 0.89 min. The mass spectrum shows an *m/z* value of 455.18794 for the protonated species, which equates to an exact mass of 454.1801 for the adduct. The MS$^4$ fragmentation series shows progressive loss of two cytosine-sized fragments and is consistent with a cyclobutane dimer of dC. These data pertain to **Figure 1** and **Supplementary Table 1**.

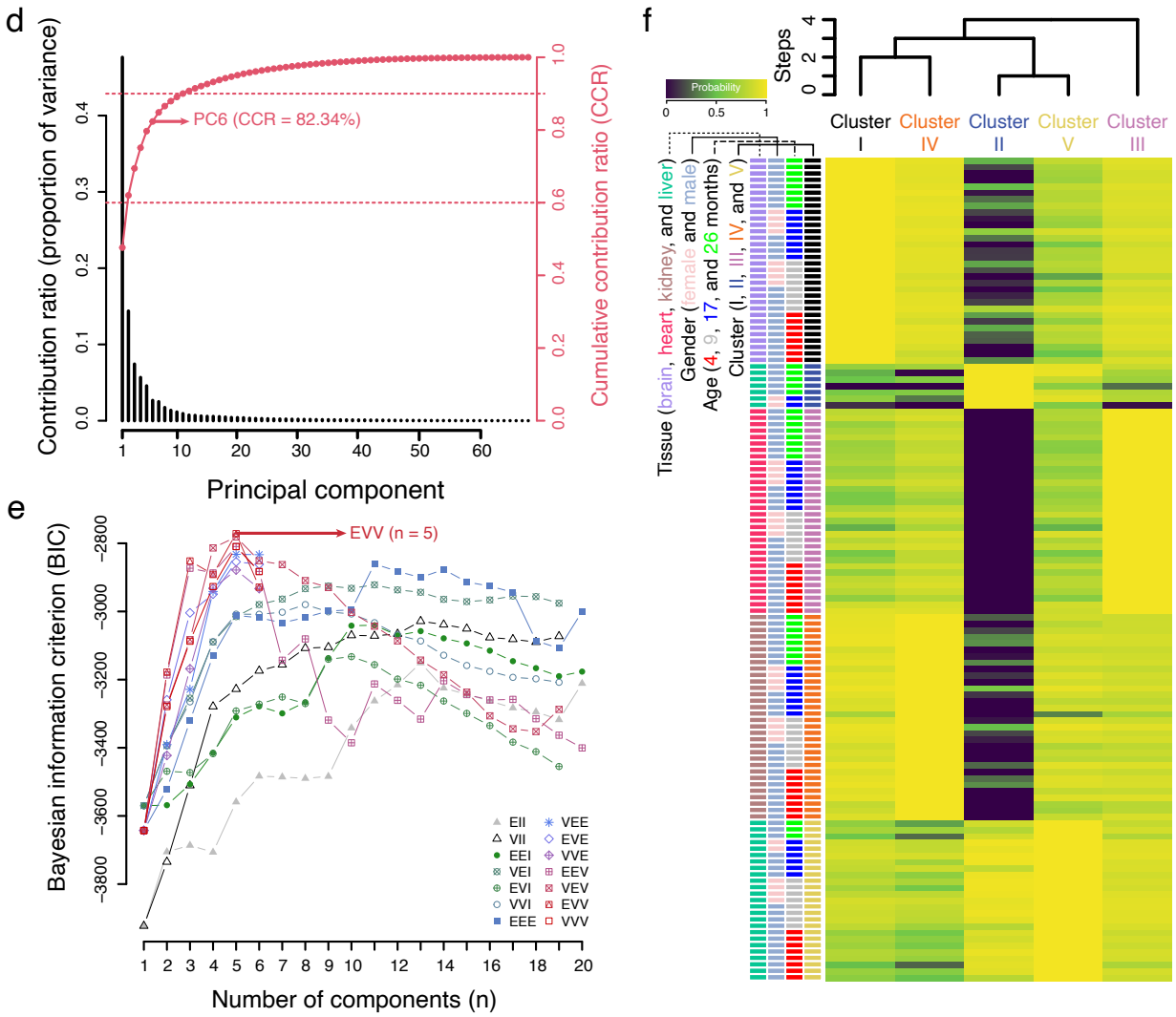



**Extended Data Figure 3.** DNA damage is strongly tissue-specific in rats. Comparison of normalized LC-MS signal intensities for 92 putative DNA adducts in 4 tissues in group #1 rats (blue numbers in upper circles and lower labels) and 94 in group #2 rats (yellow numbers). Rank order was established relative to brain from group #1 (left-most plot) with comparisons to heart, kidney, and liver for group #1 and group #2 rats. For each tissue, numbers in overlapping circles correspond to unique adducts in each group of rats (blue for group #1, yellow for group #2) and shared adducts (black numbers), with matching colors of upper circles and lower labels. These data pertain to **Figures 1** and **3**.

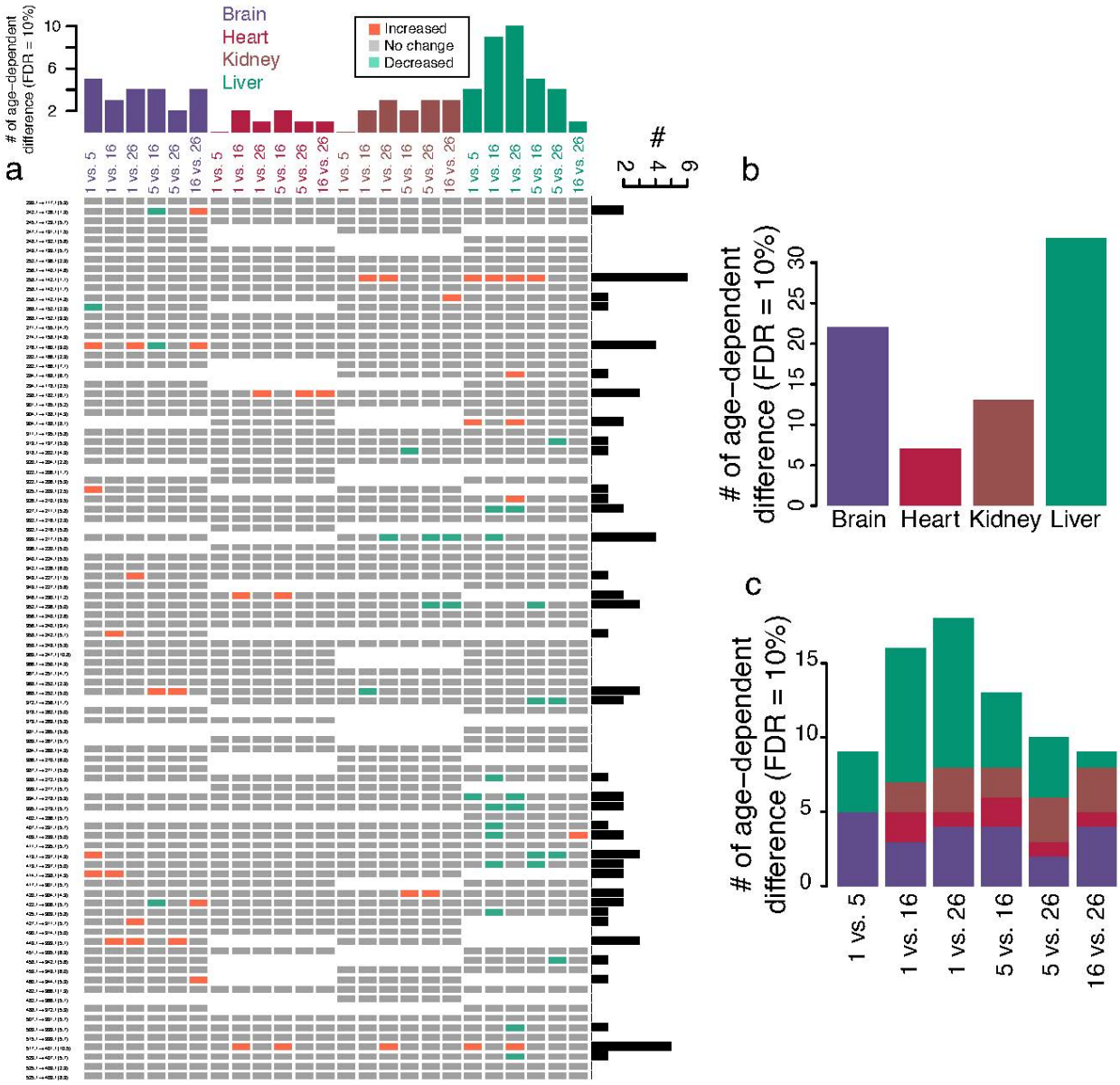



**Extended Data Figure 4.** Multivariate statistical analysis of 52 (**a-c**) and 64 (**d-f**) shared DNA adducts in group #1 and group #2 rat tissues reveals 5 co-varying clusters based on age, sex, and tissue. (**a, d**) Principal components analysis identifies 2 components that account for 63% of the variance. Black bars (left axis) indicating the contribution ratio (proportion of variance), red circles (right axis) indicating the cumulative proportion of variance (cumulative contribution ratio [CCR]), and horizontal dashed lines (right axis) indicating CCRs of 60% and 90%. (b, e) Bayesian analysis to define the number of components for Gaussian mixture model (GMM) clustering. The number of components, describing the underlying Gaussian distributions, was defined based on Bayesian information criterion values of models with differing parametrizations. The best fitting model (Ellipsoidal distributions with equal volume and variable shape and orientation axes, EVV) had 5 components that were used for GMM clustering. (**c, f**) Membership probability matrix based on GMM clustering. Each value shows the posterior probability based on the model in panels **b** and **e**. The posterior probability describes the likelihood of each sample belonging to each cluster. Logarithmic transformation of conditional probabilities from expectation maximization was used to generate the heatmap. The

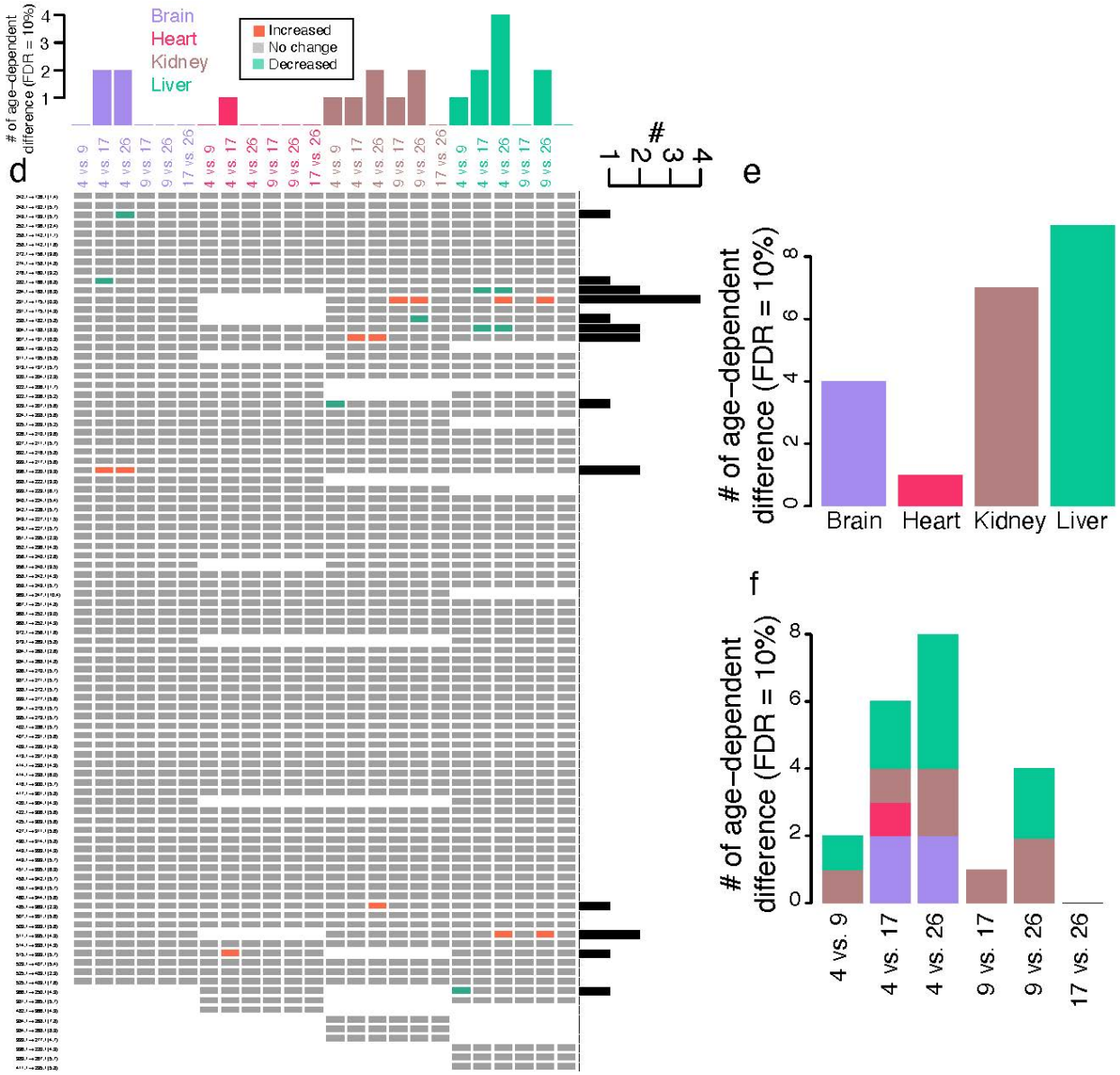
dendrogram illustrates model-based hierarchical agglomerative clustering based on the Gaussian probability model for maximizing the resulting likelihood. This figure is associated with **Figure 3b, h.**



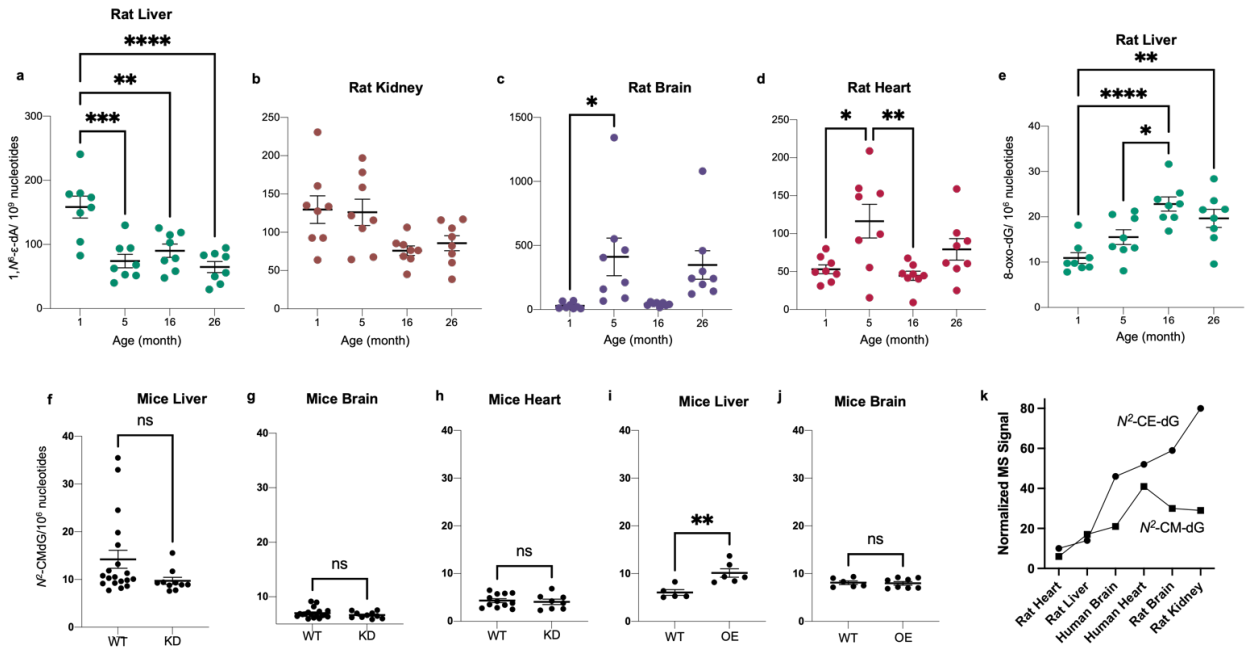

**Extended Data Figure 4,** continued. See legend above.




**Extended Data Figure 5. Age-biased DNA adducts in four rat tissues in two different sets of rats. (a, d)** Heatmaps show pairwise comparisons of all ages in each tissue. Significant changes (FDR = 10%) are presented as red (increase) and green (decrease). Bar plot on the top shows the frequency in each compared age. Bar plot (black) on the right shows frequency in each adduct. (**b**, **e**) Bar plots of overall frequency of significant changes (FDR = 10%) in each tissue. (**c**, **f**) Bar plots of overall frequency of significant changes (FDR = 10%) in each age-pairwise comparison. These data pertain to **Figures 1** and **4,** and **Supplementary Table 5.**




**Extended Data Figure 5. Age-biased DNA adducts in four rat tissues in two different sets of rats.** Continued from previous page.




**Extended Data Figure 6. Analysis of DNA damage products in rat, mouse, and human tissues.** (**a-j**) Isotope dilution chromatography-coupled triple quadrupole mass spectrometric analysis of DNA adducts in rat and mouse tissues. 1,$N^6$-ε-dA levels in rat tissues as a function of age: liver (**a**), kidney (**b**), brain (**c**), and heart (**d**). Data are mean ± SEM for $n$ = 8 rats. One-way ANOVA with Bonferroni's multiple comparisons test was used to evaluate difference among ages., *$p$<0.05, **$p$<0.01, ***$p$<0.001, ****$p$<0.0001. (**e**) 8-Oxo-dG levels as a function of age in rat liver. Box and whisker plots show mean (heavy bar) and SEM (whiskers). One-way ANOVA with Bonferroni's multiple comparisons test was used to evaluate differences among ages, *$p$<0.05, **$p$<0.01 and **** $p$<0.0001. Associated with **Figure 3**. (**f-j**) Glyoxalase I levels do not affect the level of $N^2$-CMdG in transgenic mice. $N^2$-CMdG levels were measured in liver (**f, i**), brain (**g, j**), and heart (**h**) from glyoxalase I knockdown (KD) mice (**f-h**) and from glyoxalase I over-expressing (OE) mice (**i, j**). Box and whisker plots show mean (heavy bar) and SEM (whiskers). A two-tailed unpaired Student's $t$-test was used to evaluate the difference between wild-type (WT) and transgenic mice (KD, OE). ** $p$< 0.01, *ns*: not-significant (**f-h**). Plot of normalized MS signal intensities for putative $N^2$-CE-dG (*m/z* 340-224) and validated $N^2$-CM-dG for rat and human tissues. Data represent mean for $n$=8 measurements of male and female tissues. This figure is associated with **Figure 3**.

**Extended Data Table 1. Characteristics of the human tissue donors.** ***Upper: heart tissue.*** LV: Left Ventricle, HMI: Heart Mass Index, LVMI: Left Ventricular Mass Index, BSA: Body Surface Area, BMI: Body Mass Index, LVEDd: Left Ventricle End-Diastolic dimension, LVEDs: Left Ventricle End-Systolic dimension, PW: Posterior Wall, LVEF: Left Ventricle Ejection Fraction. ***Lower: brain tissue.*** Donors are represented according to their ID number (#), category (control, CT), age at death and gender. *n*= 10 CT. These data pertain to **Figure 5**.

| Patient ID # | Age (year) | Gender | LV Mass | Heart Weight | % of LV/ Ht weight | HMI | LVMI | Weight (Kg) | Height (cm) | BSA | BMI | LVEDd | LVEDs | PW Thick | LVEF | Creatinine |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1584 | 22 | Male | 182 | 299 | 61 | 164 | 100 | 67 | 180 | 1.82 | 20 | | | | 55 | 1.0 |
| 1400 | 22 | Female | 168 | 278 | 60 | 145 | 88 | 88 | 150 | 1.91 | 39 | 4.1 | 2.2 | | 65 | 0.8 |
| 1761 | 26 | Female | 170 | 254 | 67 | 137 | 92 | 69 | 180 | 1.86 | 21 | 4.0 | 3.2 | 0.9 | 50 | 2.8 |
| 1664 | 27 | Female | 165 | 260 | 63 | 137 | 87 | 79 | 165 | 1.90 | 29 | | | | 58 | 1.9 |
| 1718 | 33 | Male | 189 | 278 | 68 | 157 | 107 | 65 | 173 | 1.77 | 22 | 4.9 | 3.7 | 0.7 | 55 | 0.7 |
| 1727 | 40 | Male | 198 | 300 | 66 | 145 | 95 | 90 | 172 | 2.07 | 30 | 4.9 | 3.0 | 1.3 | 60 | 3.7 |
| 1801 | 42 | Male | 208 | 377 | 55 | 206 | 114 | 69 | 175 | 1.83 | 23 | 4.4 | 2.6 | 1.2 | 62 | 0.9 |
| 1549 | 49 | Male | 213 | 383 | 56 | 168 | 93 | 104 | 180 | 2.28 | 32 | | | | 55 | 2.1 |
| 1750 | 51 | Female | 150 | 227 | 66 | 148 | 98 | 51 | 167 | 1.54 | 18 | 3.8 | 3.0 | 0.9 | 60 | 1.0 |
| 1600 | 51 | Female | 134 | 213 | 63 | 122 | 76 | 68 | 163 | 1.75 | 26 | 4.2 | 2.8 | 0.7 | 50 | 0.8 |
| 1739 | 52 | Male | 213 | 352 | 61 | 168 | 102 | 90 | 175 | 2.09 | 29 | 4.1 | 2.6 | 0.8 | 65 | 0.7 |
| 1666 | 54 | Male | 159 | 262 | 61 | 152 | 92 | 62 | 173 | 1.73 | 21 | | | 1.2 | 65 | 0.8 |
| 1622 | 56 | Male | 156 | 259 | 60 | 154 | 93 | 66 | 155 | 1.69 | 28 | 4.2 | 2.4 | 0.9 | 65 | 1.3 |
| 1690 | 63 | Female | 137 | 247 | 55 | 153 | 85 | 56 | 167 | 1.61 | 20 | 3.6 | 2.6 | 0.8 | 60 | 0.9 |
| 1716 | 64 | Female | 158 | 269 | 59 | 139 | 82 | 81 | 166 | 1.93 | 29 | 3.9 | 2.3 | 0.9 | 70 | 0.9 |
| 1490 | 71 | Female | 104 | 206 | 50 | 142 | 72 | 48 | 158 | 1.45 | 19 | | | | | 0.5 |
| 1732 | 72 | Female | 156 | 271 | 58 | 159 | 91 | 67 | 157 | 1.71 | 27 | 4.0 | 2.9 | 0.8 | 65 | 0.6 |
| 1485 | 78 | Male | 234 | 448 | 52 | 212 | 111 | 92 | 175 | 2.12 | 30 | | | | | 3.4 |
| 1488 | 81 | Male | 218 | 402 | 54 | 210 | 114 | 78 | 170 | 1.91 | 27 | | | | | 1.6 |

| Patient ID # | Category | Age (year) | Gender |
|---|---|---|---|
| 20998065 | CT | 81.6 | Female |
| 10315029 | CT | 87.7 | Male |
| 20875195 | CT | 74.1 | Female |
| 21000504 | CT | 85.9 | Female |
| 21001933 | CT | 80.8 | Female |
| 20240514 | CT | 89.7 | Female |
| 11331231 | CT | 81.1 | Male |
| 11615242 | CT | 71.6 | Male |
| 10589255 | CT | 74.3 | Male |
| 15196262 | CT | 81.1 | Male |