# DNA adductomics: Toward applications in biological effect assessment and environmental monitoring

Hitesh V. Motwani[1] and Elena Gorokhova[1]

[1]Stockholm University

December 11, 2023

## Background

**DNA adducts are well-established biomarkers.** They are chemical modifications occurring when certain chemicals bind covalently to DNA molecules. Unrepaired DNA adducts can disrupt DNA structure and function, potentially leading to mutations and adverse biological effects [1]. These adducts are associated with health issues, reproductive toxicity, genotoxicity, and epigenetic alterations in humans and wildlife. For the last 50 years, DNA adducts have been used as biomarkers of exposure in human health diagnostics and environmental toxicology, where the focus has mainly been on PAH adducts in fish and mussels as exposure biomarkers [2,3]. In environmental toxicology, polycyclic aromatic hydrocarbons (PAHs) have been most commonly linked to DNA adduct formation [4,5]; however, other contaminants have also been found to induce such modifications [6]. Also, in amphipods, epigenetic DNA modifications had been found to be particularly abundant in contaminated environments [7], especially in females that had carried embryos with developmental disorders [8]. Thus, ample evidence supports the informative value of detecting and quantifying DNA adducts in biological samples for assessing contaminant exposure and genomic effects.

**Traditional analytical methods for DNA adducts.** The detection of DNA adducts has commonly been conducted using immunochemical [9] and $^{32}$P-postlabelling [10] assays. Although useful, these assays do not provide chemical or structural information about the adducts to trace back to specific exposures. Moreover, these techniques have significant limitations: the immunochemical method is not sensitive enough to detect DNA adducts at trace levels, and $^{32}$P-postlabelling, albeit highly sensitive, involves radioactive phosphorus usage, implying safety concerns and regulatory challenges [11]. Indeed, researchers need to adhere to strict safety protocols for handling radioactive materials and waste disposal, which has led to a gradual disappearance of laboratories willing to conduct these measurements.

**Novel approach.** Nowadays, one of the most powerful techniques for detecting and quantifying DNA adducts is liquid chromatography-mass spectrometry (LC-MS). Using high-resolution mass spectrometry (HRMS), sensitive and selective analytical methods for detecting and identifying DNA adducts in the genome have been developed [12] and successfully applied in human health research [11,13] and ecotoxicological [8,7] diagnostics of adverse effects due to chemical exposure. Instead of analyzing a few adducts from a specific chemical exposure, HRMS gives the possibility to screen for DNA adducts from multiple classes of exposure, an "adductomics" approach. Thus, this is a new omics approach, with both target and non-targeted analytical methods available to comprehensively investigate the adductome via screening for known and unknown adducts in the genome. Various modification types, e.g., bulky PAH-adducts, methylation and oxidation, can be analyzed in a single sample by HRMS, including the determination of their chemical structures, which is useful for exposure diagnostics [14,15]. Thus, the current capacity for DNA adduct characterization is superior to the classical assays, advocating this omics approach to monitor the biological effects of contaminants [7].

# Methodology

**Test organisms**. Any plant or animal tissue is suitable for the analysis, including the whole body (for small-sized planktonic and benthic animals as well as embryos), blood, liver and muscle tissues. If fertilized eggs or embryos are present in small specimens, they should be dissected out and either analyzed separately or not included in the female DNA samples to avoid the natural ontogenetic variability in the epigenetic DNA modifications related to embryogenesis [8].

**Brief methodological description.** The workflow consists of three primary steps (Figure 1):

1. *DNA extraction and enzymatic digestion*, which yield the 2'-deoxyribonucleoside adducts and unmodified 2'-deoxyribonucleosides, follow the established protocol [8]. For extraction, we recommend a fast and easy method using Chelex 100, an ion exchange resin [16], that yields a high amount of DNA [17]. If short, Chelex 100 suspension is added to the sample, heated to release nucleic acids, and the supernatant is used for digestion. The enzymatic digestion by nuclease P1, snake venom phosphodiesterase I, and alkaline phosphatase yields the 2'-deoxyribonucleoside adducts and unmodified 2'-deoxyribonucleosides [8]. The samples can be stored at -20 degC until analysis by LC-HRMS/MS.
2. *Liquid chromatography HRMS analysis and data processing*, including low- and high-mass adduct detection [7], quantification and identification using open-source databases [18,19]. Typically, reversed phase chromatography coupled to an Orbitrap HRMS instrumentation can be used for this purpose, employing data-dependent acquisition with an inclusion list or data-independent acquisition. The resulting high-resolution accurate mass (HRAM) data allows for increased selectivity due to the ability to differentiate the adduct ion signals from isobaric background ions. The accurate mass measurements further provide sufficient information to determine the molecular formula of the adduct analytes, which are useful in the structural characterization of the adducts. Open-source software for peak screening of the DNA adducts is available [20] for non-targeted detection of DNA adducts based on characteristic neutral loss (deoxyribose moiety, 116.0474 Da) and without any prior knowledge of the adducts. To reduce false positives, data filtering is performed using retention time alignment, accurate mass window, MS/MS fragments and any associated metadata available. The output data are normalized peak areas for specific adducts that can be used for the downstream statistical evaluations (Figure 2). The data format is typically a matrix, with adduct relative abundance and samples (i.e., individuals) given in columns and rows or *vice-versa*.
3. *Statistical analysis* for the DNA adductome of each sentinel species used for monitoring purposes should include the following steps: (1) screening the output data to exclude non-varying adducts; (2) identification of the background variability; and (3) evaluation of the test samples against the background variability (i.e., reference state). However, currently, there are no established methods for deriving environmental quality standards from the omics data [21,(missing citation),23]; therefore, different approaches should be evaluated and compared using data for different species and environmental settings.

.

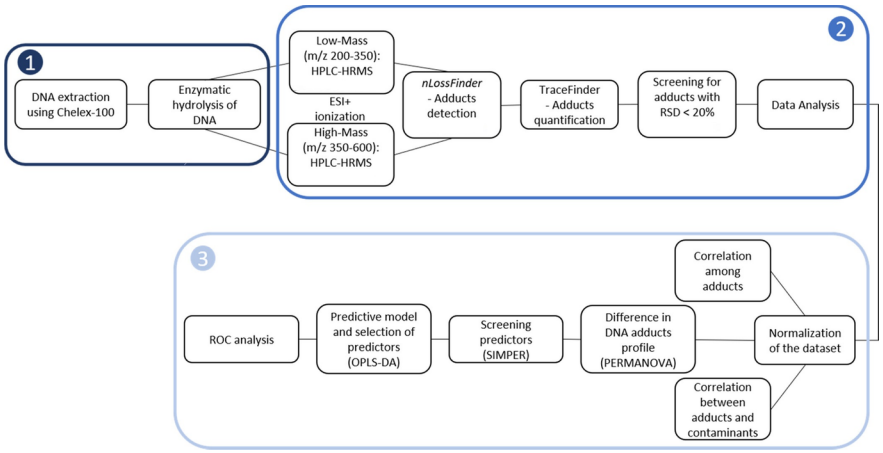


Figure 1: Workflow of the DNA adductome analysis, including sample preparation (1), HRMS analysis and data processing (2), and an example of a statistical evaluation (3) performing permutational multivariate analysis of variance (PERMANOVA), similarity percentage analysis (SIMPER), orthogonal partial least squares discriminant analysis (OPLS-DA), and receiver operating characteristic (ROC) analysis to link specific adducts to the chemical exposure. Modified from [7].

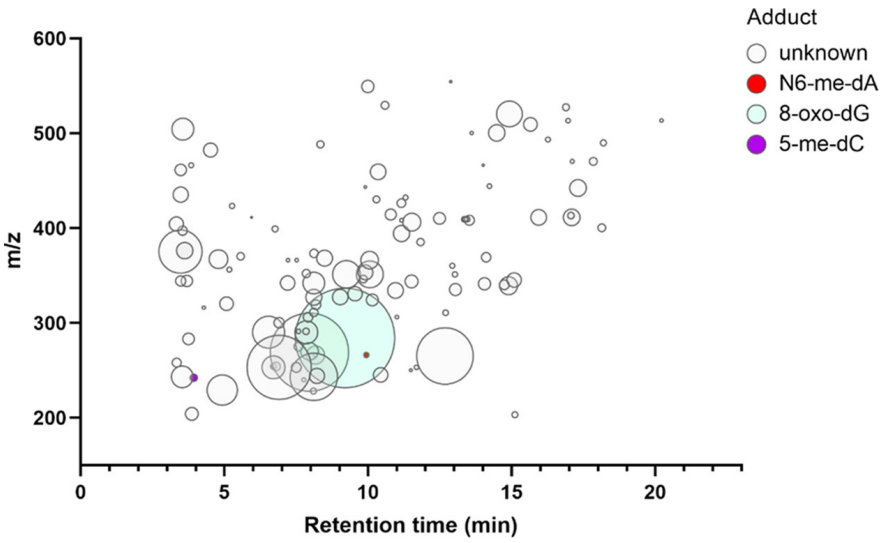


Figure 2: An example of the output data for the relative peak intensity for characterised and uncharacterised DNA adducts identified in the adductome of the amphipod *Monoporeia affinis*, a sentinel species in the Swedish National Marine Monitoring Programme. The data represent the adductome profile of a single individual. Modified from [24]

**Quality assurance (QA) and quality control (QC) of adductomics analysis.** QA and QC protocols should be followed to control and minimize fluctuations and biases that can affect the LC-HRMS adductomics workflow and data quality. The QA and QC guidelines developed for metabolomics [25,26,27] are applicable in assessing system performance and reporting the adductomics data due to the conceptual similarities in the techniques and workflows. Availability of reference standards corresponding to various 2'-deoxyribonucleoside adducts spanning over a wide range of chromatographic retention times and mass-to-charge ratios (m/z) is

crucial to evaluate the performance of the analytical system in terms of mass error, shift in chromatographic retention time, peak shape and MS-fragmentation efficiency. Other important parameters to ensure the reliability and accuracy of analytical results are instrument calibration, method validation, evaluation of quality control samples, data quality assessment, batch-to-batch consistency, as well as personnel training and documentation. By implementing a comprehensive QC/QA program, laboratories can enhance the reliability, reproducibility, and validity of the adductome data, which are essential for the consistency of the environmental assessment and comparability of the results in longitudinal data sets.

**Statistical considerations, targets and background variability for exposure assessment**. To establish the baseline variability of the DNA adductome for a target tissue/species (and, possibly, a population), replicate samples of single specimens for each assessment site are needed. Based on our experience with amphipods, about 15-20 individuals per site are sufficient when only females in the reproductive stage are considered [8]. However, species with a larger or lower inter-individual variability in unimpacted sites may require a larger sample size.

Before evaluating data in relation to the contaminants and other health parameters, some QC checks are conducted, using, e.g., Principal Component Analysis (PCA) as a checkpoint to screen for outlier data points and obtain a global perspective of the data [28]. All single adducts are evaluated for variability and those showing less than 1 % variation across the samples are omitted from the further analyses. Also, adducts that are present in less than 5 % of the samples are omitted.

The data are evaluated using two-tier diagnostics with standard multivariate approaches for omics data [29].

- **First**, the primary adductome data output is used to (1) define the variability of the data originating from unimpacted sites and/or healthy specimens (i.e., the target space in the multivariate ordination; Figure 3); (2) assess the predicted class membership or score to evaluate if the test sample aligns with the background samples; and (3) estimate the proportion of the test samples that do not align with the background samples. For that, a Partial Least Squares-Discriminant Analysis (PLS-DA), a supervised method that combines aspects of PCA and discriminant analysis, and, in particular, its upgraded version called OPLS-DA (orthogonal PLS-DA) is recommended (Figure 3B). For the environmental status assessment, we suggest applying the following principle: if more than 50% of the test samples from a site/area are classified as not belonging to the reference group, the site is considered as deviating significantly from the unimpacted state in terms of the DNA adduct composition and relative abundance.
- **Second**, the test samples that do not align with the reference adductome for the species/population in question are subjected to the analysis of individual variability of the influential adducts; the latter are identified by the PLS-DA model and ROC (Receiver Operating Curve) analysis (Figure 4). Once the PLS/OPLS-DA model is built, the VIP (variable influence of projection) measure can be obtained for the adducts based on their association with the identified predictive components. Each of these adducts may permit the identification of exposures to certain hazardous chemicals in the environment with a unique diagnostic value. For each individual adduct, the same principle as for many other biomarkers [30] can be used by defining the background assessment criteria (BAC) as the $90^{\text{th}}$ percentile of the relative abundance of this adduct in the areas regarded as less polluted reference areas. Bootstrapping (100 000 runs) can be used to derive mean, median and $90^{\text{th}}$ percentile values. The significant deviations of the influential adducts from the corresponding BAC values should be reported to facilitate the interpretation of the overall DNA adductome response.

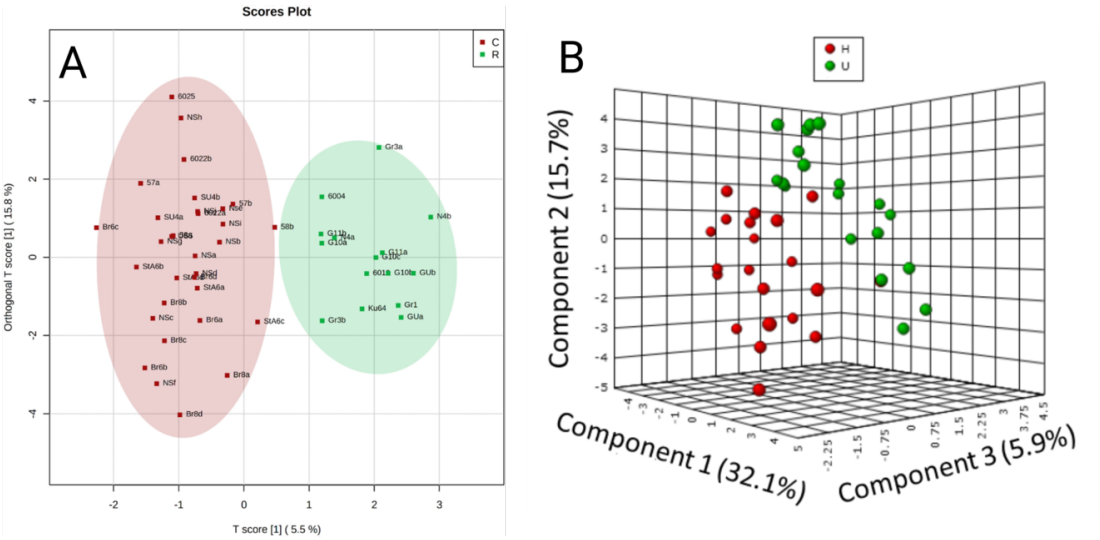


Figure 3: An example of PLS-DA for DNA adductome profile in the amphipod *Monoporeia affinis* from different sites in the Baltic Sea (A); modified from [7]. The data points represent the samples collected in the stations classified as contaminated (C; red) or reference (R; green); the labels indicate the station codes. In this example, none of the samples collected from the contaminated stations aligned with the ordination space of the reference samples. The B-panel shows a separation based on the DNA adductome between field-collected females carrying different proportions of embryos with developmental aberrations and classified as Healthy (H: aberration frequency is less than 5 %) and Unhealthy (U: up to 60 %); modified from [8].

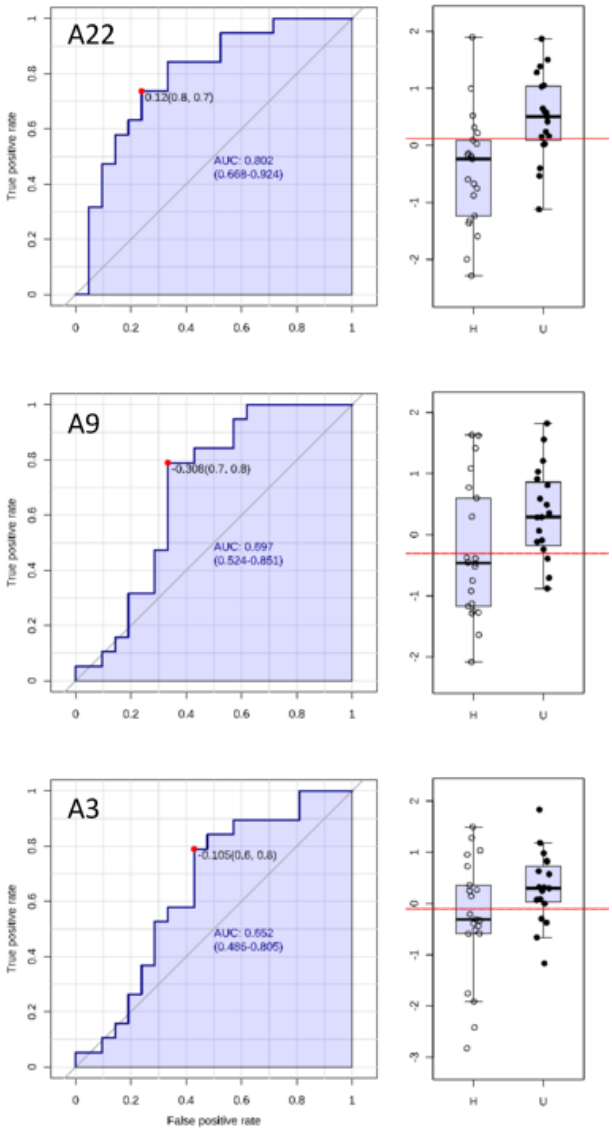


Figure 4: An example of the single adduct evaluation using the area under the receiver operating characteristic (ROC) curve (AUC) logistic regression approach for discrimination between the amphipod females with low (H: Healthy) and high (U: unhealthy) embryo aberration frequency. For each of the three discriminating adducts, the left panel shows the AUC confidence interval, true positive and false positive rates, and confidence interval (CI), the right panel shows the normalized values for the adducts in healthy and unhealthy individuals. AUC logistic regression approach identified adducts A22 ( $N^6$-Methyl-2'-deoxyadenosine, $N^6$-me-dA; AUC = 0.802), A9 (uncharacterized adduct; AUC = 0.697), and A3 (5-Methyl-2'-deoxycytidine, 5-me-dC; AUC = 0.652) to have the greatest specificity and sensitivity in relation to the developmental pathologies in these amphipods. The logistic regression based on these adducts had 80 and 85 % classification accuracy for predicting unimpacted and impacted animals, respectively; modified from [8].

## Conclusions

Environmental omics, including nucleic acid adductomics, can provide valuable biomarkers for assessing exposure to environmental contaminants and effects. Incorporating DNA adductomics into the existing biomarker batteries would be a much-needed tool for the early detection of genome-level alterations in wildlife and represent a significant advancement in environmental health assessment.

The DNA adductomics workflow described here demonstrates that technology is ready for the integration of DNA adductome screening in large-scale environmental monitoring, offering several advantages. It can screen and process data for both low-mass adducts, such as epigenetic marks or oxidative adducts, and bulky adducts, thereby providing a broad range of detectable adducts for environmental assessment. Open-source databases and software are available, allowing the screening of individual samples without a library with a reasonably short data processing time and at reasonable costs. Additionally, the raw files are routinely archived and can be re-processed at any time to obtain more information from a single sample.

Next, it is crucial to focus on the structural identification of these modifications, link the candidate DNA adducts with potential metabolites, and identify possible exposure agents to explain the responses in different species. In parallel, experimental studies are needed to validate relationships between specific adducts, exposure agents and health outcomes using controlled exposure systems and establish critical levels of the adduct biomarkers using the toxicity data.

## References

1.Phillips, D. H. & Arlt, V. M. Genotoxicity: damage to DNA and its consequences. in *Molecular, Clinical and Environmental Toxicology* 87–110 (Birkhäuser Basel, 2009). doi:10.1007/978-3-7643-8336-$7_4$.

2.Pampanin, D. M. *et al.*. DNA adducts in marine fish as biological marker of genotoxicity in environmental monitoring: The way forward. *Mar Environ Res* **125**, 49–62 (2017).

3.Dolcetti, L., Zuanna, L. D. & Venier, P. DNA adducts in mussels and fish exposed to bulky genotoxic compounds. *Mar Environ Res* **54**, 481–6 (2002).

4.Skarphédinsdóttir, H., Ericson, G., Svavarsson, J. & Naes, K. DNA adducts and polycyclic aromatic hydrocarbon (PAH) tissue levels in blue mussels (Mytilus spp.) from Nordic coastal sites.. *Mar Environ Res* **64**, 479–91 (2007).

5.Amat, A., Pfohl-Leszkowicz, A., Burgeot, T. & Castegnaro, M. DNA Adducts as a biomarker of pollution: Field study on the genotoxic impact. Evolution of the Erika oil spill on mussels (Mytilus edulus) over a period of eleven months. *Polycyclic Aromatic Compounds* **24**, (2004).

6.Guilherme, S., Gaivão, I., Santos, M. A. & Pacheco, M. DNA damage in fish (Anguilla anguilla) exposed to a glyphosate-based herbicide – Elucidation of organ-specificity and the role of oxidative stress. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **743**, (2012).

7.Martella, G. *et al.*. DNA Adductomics for the Biological Effect Assessment of Contaminant Exposure in Marine Sediments. *Environ Sci Technol* **57**, 10591–10603 (2023).

8.Gorokhova, E. *et al.*. DNA epigenetic marks are linked to embryo aberrations in amphipods. *Sci Rep* **10**, 655 (2020).

9.Müller, R., Adamkiewicz, J. & Rajewsky, M. F. Immunological detection and quantification of carcinogen-modified DNA components.. *IARC Sci Publ* 463–79 (1982).

10.Poirier, M. C., Santella, R. M. & Weston, A. Carcinogen macromolecular adducts and their measurement.. *Carcinogenesis* **21**, 353–9 (2000).

11.BH, Y., Guo, J., Bellamri, M. & Turesky, R. J. DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. *Mass Spectrom Rev* **39**, 55–82 (2020).

12.Hemeryck, L. Y., Moore, S. A. & Vanhaecke, L. Mass Spectrometric Mapping of the DNA Adductome as a Means to Study Genotoxin Exposure, Metabolism, and Effect. *Analytical Chemistry* **88**, (2016).

13.Totsuka, Y., Watanabe, M. & Lin, Y. New horizons of DNA adductome for exploring environmental causes of cancer. *Cancer Sci* **112**, 7–15 (2021).

14.Balbo, S., Turesky, R. J. & Villalta, P. W. DNA adductomics. *Chem Res Toxicol* **27**, 356–66 (2014).

15.Villalta, P. W. & Balbo, S. The Future of DNA Adductomic Analysis. *Int J Mol Sci* **18**, (2017).

16.Walsh, P. S., Metzger, D. A. & Higushi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4): 506-13 (April 1991). *Biotechniques* **54**, 134–9 (2013).

17.Phillips, K., McCallum, N. & Welch, L. A comparison of methods for forensic DNA extraction: Chelex-100® and the QIAGEN DNA Investigator Kit (manual and automated). *Forensic Sci Int Genet* **6**, 282–5 (2012).

18.La, B. G., Nommesen, K. D., Cuparencu, C., Stanstrup, J. & Dragsted, L. O. A Comprehensive Database for DNA Adductomics.. *Front Chem* **10**, 908572 (2022).

19.Guo, J. *et al.*. Development of a DNA Adductome Mass Spectral Database. *Chemical Research in Toxicology* **33**, (2020).

20.Sousa, P. F. M., Martella, G., Åberg, K. M., Esfahani, B. & Motwani, H. V. nLossFinder-A Graphical User Interface Program for the Nontargeted Detection of DNA Adducts. *Toxics* **9**, (2021).

21.Henke, A. N., Chilukuri, S., Langan, L. M. & Brooks, B. W. Reporting and reproducibility: Proteomics of fish models in environmental toxicology and ecotoxicology. *Sci Total Environ* **912**, 168455 (2023).

23.Ebner, J. N. Trends in the Application of Omics to Ecotoxicology and Stress Ecology. *Genes (Basel)* **12**, (2021).

24.Martella, G. DNA adductomics: Method development and integration in biological effect monitoring. (http://su.diva-portal.org/smash/record.jsf?pid=diva2%3A1752375&dswid=-6434, 2023).

25.Dudzik, D., Barbas-Bernardos, C., García, A. & Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review.. *J Pharm Biomed Anal* **147**, 149–173 (2018).

26.Broadhurst, D. *et al.*. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72 (2018).

27.Beger, R. D. *et al.*. Towards quality assurance and quality control in untargeted metabolomics studies. *Metabolomics* **15**, 4 (2019).

28.Anwardeen, N. R., Diboun, I., Mokrab, Y., Althani, A. A. & Elrayess, M. A. Statistical methods and resources for biomarker discovery using metabolomics. *BMC Bioinformatics* **24**, 250 (2023).

29.Brereton, R. G. Multivariate classification models. *Journal of Chemometrics* **35**, (2021).

30.Hylland, K. *et al.*. Integrated chemical and biological assessment of contaminant impacts in selected European coastal and offshore marine areas. *Mar Environ Res* **124**, 130–138 (2017).