ENVIRONMENTAL EPIGENETICS

Environmental Epigenetics, 2024, **10(1)**, dvae020

DOI: https://doi.org/10.1093/eep/dvae020
Advance Access Publication Date: 29 November 2024
Research Article

A new approach to study stochastic epigenetic mutations in sperm methylome of Vietnam war veterans directly exposed to Agent Orange

Luigi Corsaro[1,2,†], **Davide Sacco**[1,2,†], **Carlo Corbetta**[3], **Davide Gentilini**[2], **Alice Faversani**[1], **Fulvio Ferrara**[3], **Lucy Costantino**[1,*]

[1]Medical Genetics Laboratory, Centro Diagnostico Italiano, Milan, MI 20147, Italy
[2]Department of Brain and Behavioural Sciences, Università di Pavia, Pavia, PV 27100, Italy
[3]Integrated Laboratory Medicine Services, Centro Diagnostico Italiano, Milan, MI 20147, Italy

[†]These authors contributed equally to this work.
*Corresponding author. Lucy Costantino, Medical Genetics Laboratory, Centro Diagnostico Italiano, Milan, MI 20147, Italy. E-mail: lucy.costantino@cdi.it.

Abstract

Among the various environmental pollutants, dioxin, a highly toxic and widely used compound, is associated with numerous adverse health effects, including a potentially toxic multigenerational effect. Understanding the mechanisms by which dioxin exposure can affect sperm epigenetics is critical to comprehending the potential consequences for offspring health and development. This study investigates the possible association between weighted epimutations, hypothesized as markers of epigenetic drift, and dioxin exposure in sperm tissues. We used a public online methylation dataset consisting of 37 participants: 26 Vietnam veterans exposed to Agent Orange, an herbicide contaminated with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and 11 individuals not directly exposed to TCDD but whose serum dioxin levels are equivalent to the background. In our study, conducted at the gene level, 437 epimutated genes were identified as significantly associated with each single-digit increase in serum dioxin levels. We found no significant association between the rise in total epimutation load and serum dioxin levels. The pathway analysis performed on the genes reveals biological processes mainly related to changes in embryonic morphology, development, and reproduction. Results from our current study suggest the importance of further investigations on the consequences of dioxin exposure in humans with specific reference to germinal tissue and related heredity.

Keywords: TCDD; 2,3,7,8-tetrachlorodibenzodioxin; methylation; epistochastic mutation

Introduction

Agent Orange is a mixture of tactical herbicides that the US Army sprayed in different locations (Vietnam and other southeast locations) from 1962 to 1971 during the Vietnam War. The US soldiers who may have been exposed to Agent Orange include soldiers who served in these locations and who flew on or worked on C-123 Aircraft. The two active ingredients in the Agent Orange herbicide combination were equal amounts of 2,4-dichlorophenoxyacetic acid (2,4-D) and 2,4,5-trichlorophenoxyacetic acid (2,4,5-T), which contained as primary contaminant 2, 3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), an unwanted byproduct of herbicide production [1, 2]. TCDD is a highly toxic chemical belonging to the dioxin family and is considered the most toxic of this group by the World Health Organization [3]. Dioxins are well known to be potent endocrine disruptors. TCDD demonstrates this property by interacting with the aryl hydrocarbon receptor (Ahr), which activates various signalling pathways and alters hormonal signals. This substance leads to metabolic disorders, immune disorders, an increased risk of cancer, and reproductive and developmental problems [4–7].

Furthermore, this chemical poses a danger exacerbated by its remarkable ability to persist in various environmental matrices and mammalian adipose tissue [8; 9].

How long TCDD remains accumulated depends on the organism; in directly exposed humans, it generally has an observed half-life of between 7 and 11 years. The persistence in the organism hinders the elimination of the pollutant from the body, leading to an accumulation over time that increases the risk of developing health problems mentioned earlier [10–13].

Due to its persistence, the damage caused by exposure to TCDD is not limited to only those directly exposed but can also affect their offspring [14]. Studies over several generations of rats have shown that TCDD exposure of the ancestors impairs the reproductive health of their offspring [15]. Specifically, male offspring show problems with sperm quality and production, while female

Received 28 May 2024; **revised** 16 September 2024; **accepted** 28 October 2024

offspring experience issues like ovarian cysts and pregnancy complications [16, 17].

In humans, given the rarity of intense exposure to TCDD, our knowledge of the multi-generational effects of this pollutant comes primarily from studies conducted on the newborns from mothers exposed to large amounts of TCDD almost 7 years before pregnancy during the Seveso incident in 1976 [18]. Specifically, researchers have linked the identified multigenerational consequences to an altered sex ratio of offspring, higher mean neonatal blood thyroid-stimulating hormone levels, and various impacts on different seminal parameters, including concentration, count, and motility [6, 19]. The exact mechanisms that lead to the development of certain diseases caused by environmental toxic substances are not yet fully understood. Furthermore, many studies suggest that the exposome, which refers to an individual's total exposure to various chemical, biological, and radiation agents from birth to death, plays a crucial role in shaping human epigenetic mechanisms [20, 21]. Unlike genomic mutations, which result in permanent changes to the DNA, epigenetic modifications can be transient and reversible [22, 23]. This characteristic allows the epigenome to reflect an individual's environmental history and respond to changes in exposure, thus influencing gene expression without altering the genomic sequence itself [21,24–26]. Those types of changes at the gene expression level lead to health outcomes [27, 28]. The primary epigenetic mechanisms impacted by the exposome are DNA methylation, histone modifications, and the role of noncoding RNAs. These epigenetic changes, mainly induced by environmental substances, are observed in animal models and are present not only in those directly exposed but also in their offspring, impacting the phenotypes of future generations [11].

DNA methylation is the most easily shaped and is the optimal candidate for use as a marker due to its dynamic nature [20]. DNA methylation studies often utilize differential analysis to detect variations in methylation patterns associated with specific traits or environmental factors [29, 30]. This technique identifies genomic regions with differential methylation that may influence gene expression. Additionally, stochasticity—random fluctuations in methylation levels— is crucial and can be influenced by genetic and environmental factors [31]. These stochastic epigenetic mutations (SEMs), which occur during embryogenesis and cell division, are influenced by environmental factors and developmental processes and contribute to phenotypic diversity [31]. Combining differential analysis with understanding stochasticity offers a complete perspective on methylation variation and its significance in adaptation and evolution [31].

Studies have found that exposure to endocrine-disrupting chemicals can lead to changes in DNA methylation [32–34]. The exposure of animal models to these chemicals during crucial developmental stages can alter the methylation of somatic cells, leading to changes that can affect the proper development of the exposed individual [35]. Furthermore, if the exposure occurs in germinal cells, these changes can be inherited and affect the health of the offspring [36].

TCDD can alter the methylation values of imprinting genes, which typically should not undergo methylation changes, based on research conducted in cell culture, animal models, and humans [35,37,38].

Methylation is a tissue-specific epigenetic marker involved in cellular differentiation, and therefore, we studied the possible implications at a multigenerational level, focusing on the germinal tissue [39].

Research focused on sperm methylation and direct exposure to TCDD is limited. The primary subjects of these investigations have been a small group of young Russian adults who were exposed to TCDD during their prepubertal age [40] and Vietnam veterans who were accidentally sprayed with varying quantities of Agent Orange, an herbicide contaminated with TCDD [1]. However, in the 'Russian Boys' cohort, exposed to a dioxin-polluted environment, investigators linked the intensity of TCDD exposure to differences in methylation in various gene regions involved in cellular maintenance and assembly functions. In the second case, research on the sperm methylome of Vietnam veterans has shown that serum TCDD levels, measured at their exposure, are associated with altered methylation in two imprinting genes [41]. Additionally, Nwanaji *et al.* [42] observed a possible connection between TCDD exposure and accelerated ageing in sperm associated with higher serum TCDD levels.

The biostatistics methodologies for investigating the human methylome are increasingly expanding [43; 44]. Calculating SEMs offers an alternative method to quantify widespread aberrations in the methylome [45]. This type of epigenetic alteration is viewed as stochastic aberrations scattered across the methylome and differs in detection from commonly identified methylation markers through differential analysis. The latter targets specific sites or regions with different average methylation values between two groups. In contrast, SEMs are considered values significantly higher or lower than those of a reference population [45]. This type of analysis thus allows for the exploration of a less usual and rarely investigated aspect in the context of environmental pollutants, except in a few cases, such as to people directly exposed to polybrominated biphenyl, where an increase in the epimutational load, understood as the total sum of the detected SEMs, has been observed in directly exposed subjects [46]. Various studies over the years have shown that the aggregation of SEMs is significantly associated with multiple pathological conditions [47, 48]. Furthermore, SEMs are also associated with various physiological processes, such as development and ageing [49, 50]. Previous studies have confirmed an altered methylome as measured by the signal mean difference between the TCDD-exposed and control groups [40,41,51].

However, the relationship between SEMs and exposure to TCDD is entirely unknown.

Our study investigates how TCDD exposure affects the germline methylome by modelling the effect using stochastic epigenetic mutational load (EML), intended as the burden or accumulation of SEMs [52].

We used the publicly available dataset GSE139307, which consists of data on the methylation signals of spermatic tissue from Vietnam veterans. The study will investigate whether SEMs connect with direct exposure at the individual level, analysing all SEMs dispersed throughout the genome or at the level of specific genes, and then evaluate the biological processes associated with SEMs that can impact the health of offspring.

Results

Differential analysis of the methylation signal revealed 90 significant genes associated with TCDD concentration (Supplementary Data 1). The regression coefficient linking methylation level to TCDD concentration never exceeds 1/1000 of the increase due to a TCDD increase of 1 ppt, indicating that the median level of methylation signal increases insignificantly for all affected genes.

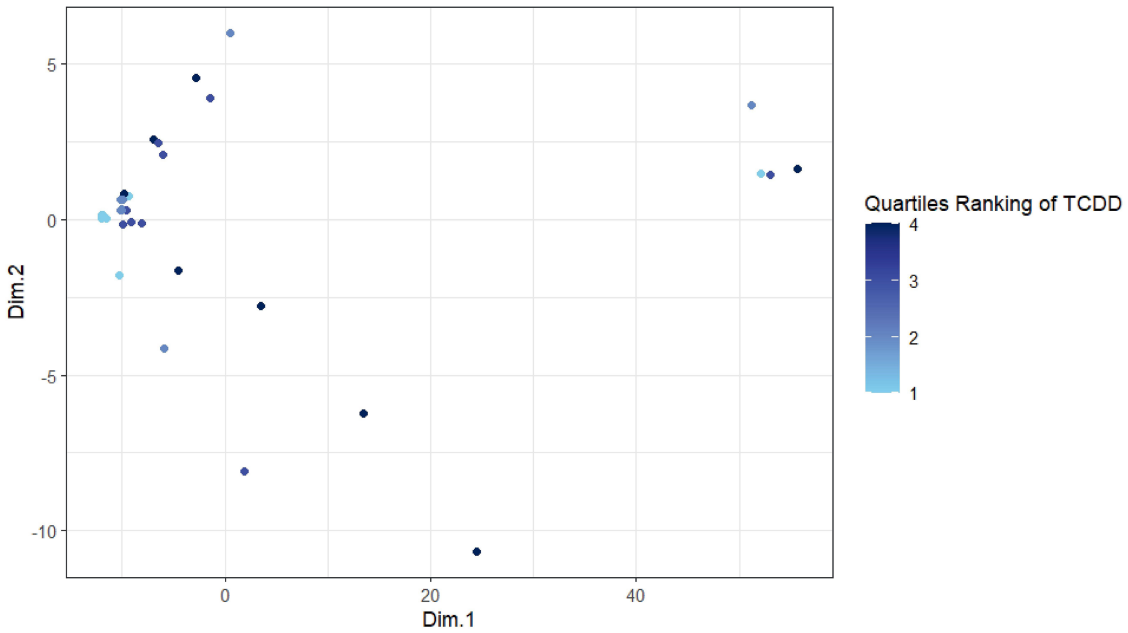

Fig 1. Scatterplot based on the principal components analysis results of gene epimutational burden, with individuals shaded from lighter to darker tones to represent their assigned dioxin exposure quartile, from lowest to highest

Pathway analysis also shows no effects on biological processes or molecular functions.

During our study on Weighted Stochastic Epimutation analysis, we examined the impact of TCDD exposure on the SEM load per studied subject. After a thorough analysis, we found no significant correlation between the two. Subsequently, we conducted our analysis with the weighted stochastic epimutation calculation at the gene level. After adjusting the *P*-value using the Benjamini–Hochberg method, 437 genes showed significant associations with dioxin exposure. In Fig. 1, it is possible to graphically observe the synthesis of the SEM burden for genes that were significant for each individual related to serum TCDD levels reported in quartiles.

For the complete list of significant genes, including the model's *P*-value and the beta coefficient, refer to Supplementary Data 2.

In addition to the P_{value}, which indicates the significance of the association with dioxin, we also considered the beta coefficient to interpret the percentage increase in weighted SEMs with each unit increase in dioxin levels. Unlike conventional methods such as differential methylation analysis, where the intensity of the effect on methylation levels is considered in terms of log fold change (logFC), here we consider the beta of the regression, and there are no specific cutoff values to understand when the change is biologically significant. We found that the percentage increase in serum dioxin (part per trillion) ranged from 1% to 43%. Table 1 shows the genes that exhibited a more significant percentage increase than the unit increase in dioxin.

By analysing the biological processes summarizing all 437 significant genes, we mainly identified 10 significantly enriched pathways with an adjusted $P_{value} < 0.05$, which belong to two main pathway families:

1 Embryonic and general organismal development
2 Reproduction.

Table 2 lists the various biological pathways, the number of identified genes by our analysis involved in each pathway (nGenes), the total count of the pathway's gene (PathwayGenes), the identification code, and the significance of the enrichment (Enrichment False Detection Rate (FDR)).

Table 1. The top five genes with the highest percentage increase in beta coefficients and their corresponding methylation alterations.

Gene	Epimutation cause	Beta coefficient (%)
HOXA5	Hypermethylated	43.26
HOXA3	Hypermethylated	42.38
WT1	Hypermethylated	35.64
HOXA4	Hypermethylated	28.81
ASCL2	Hypermethylated	27.94

For further information regarding which genes belong to which pathway, see Supplementary Data 3. Furthermore, in Fig. 2, they are graphically displayed in order of significance along with their respective enrichment levels. The boxplots in Fig. 3 show how the relationship between the percentage increase in weighted SEM exposure and the rise in serum dioxin levels correlates with the biological pathways. We note that the pathways belonging to the first family, which concern the organism's development, are enriched with more disrupted genes than those related to reproductive processes. Furthermore, we observe that the top five genes prioritized for biological impact as determined by the beta coefficient, whose names are shown in Fig. 3, are significantly different from the others in terms of the beta coefficient. Subsequently, we wanted to graphically represent the relationship from the subjects' perspective between the amount of dioxin exposure and the accumulation of weighted SEMs burden for each subject in Fig. 4. Despite the wide distribution of dioxin, for which we resorted to quartile ranking to illustrate this relationship, it is noticeable that this trend is similar across all genes.

Discussion

Exposure to dioxins and dioxin-like substances is associated in humans with different harmful effects, ranging from acute

Table 2. Pathways impacted by altered epimutation load genes, sorted by fold enrichment.

Enrichment FDR	nGenes	Pathway genes	Fold enrichment	Identification code—pathway
4.51E-07	20	229	6.16	GO:0009952 anterior/posterior pattern specification
1.30E-03	17	312	3.84	GO:0048562 embryonic organ morphogenesis
3.42E-04	20	377	3.74	GO:0003002 regionalization
3.62E-04	23	491	3.30	GO:0007389 pattern specification proc.
8.78E-05	29	641	3.19	GO:0048598 embryonic morphogenesis
1.33E-03	27	702	2.71	GO:0009792 embryo development ending in birth or egg hatching
2.05E-03	26	680	2.70	GO:0043009 chordate embryonic development
1.82E-04	41	1189	2.43	GO:0009790 embryo development
1.97E-04	50	1635	2.16	GO:0022414 reproductive proc.
1.97E-04	50	1644	2.14	GO:0000003 reproduction

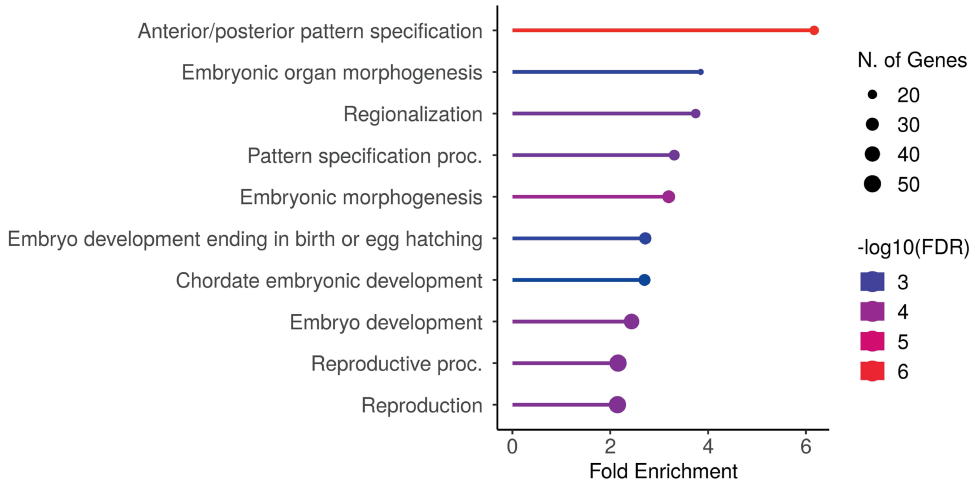

Fig 2. Lollipop plot depicting significantly enriched biological pathways. The x-axis represents fold enrichment values, while the y-axis lists the pathway names. The circle size indicates the number of genes within each pathway at the end of each line, and the colour of the circles denotes the significance level from multiple testing (FDR)

toxic responses (chloracne) to the development of long-term chronic diseases. These include reproductive, developmental, and neurodevelopmental effects, altered male-to-female birth ratio, immunotoxicity, alterations in thyroid hormones, liver, and tooth development, with particular sensitivity in fetuses and infants [6,7,18,53–55]. Zhang *et al.* attributed TCDD's binding ability to the AhR as the primary mechanism to explain the toxicity of dioxin and dioxin-like substances [56]. This transcription factor regulates a variety of genes involved in cellular processes. A cascade of downstream events occurs upon AhR activation, leading to the expression of genes implicated in xenobiotic metabolism, cell proliferation, and apoptosis. Given the role of the AhR in mediating dioxin toxicity and its known involvement in regulating a variety of genes, it is plausible that dioxin exposure could lead to epigenetic alterations in these genes. Habano *et al.* showed that AhR activation influences DNA methylation patterns, suggesting that dioxin exposure could alter the epigenetic landscape of genes involved in development, metabolism, and immune function [57]. The AhR, the primary cellular receptor for dioxins, is a crucial mediator of their toxic effects. Upon binding to dioxins, the AhR undergoes a conformational change, translocates to the nucleus, and forms a heterodimer with the aryl hydrocarbon nuclear translocator. This heterodimer binds to specific DNA sequences, known as dioxin response elements, activating target gene transcription [57]. Dioxin's ability to perturb cellular processes, particularly development-related, has been extensively studied. The potential for dioxin to induce epigenetic alterations involves changes in gene expression without altering the underlying DNA sequence, leading to long-lasting consequences for exposed individuals and their offspring. Epigenetic mutations, such as DNA methylation and histone modifications, have been implicated in mediating multigenerational effects and can be influenced by environmental factors, including exposure to toxic substances like dioxin [35, 36]. By altering gene expression patterns, epigenetic changes can affect development, health, and disease susceptibility. DNA methylation, in particular, has been implicated in silencing Hox genes following dioxin exposure. Multiple studies have shown that dioxin exposure significantly alters the expression profiles of Hox genes [58–60]. These genes serve as master regulators of developmental patterning and are particularly vulnerable to environmental disruptions [61, 62]. Disruptions in the expression of Hox genes can result in various congenital malformations and diseases. The consequences of Hox gene dysregulation due to dioxin exposure are extensive and can include a range of developmental abnormalities: skeletal malformations, cardiovascular defects, neural tube defects, and cancer [63, 64]. The mechanisms underlying dioxin-induced dysregulation of Hox genes are complex and involve several potential contributing factors: direct binding of AhR to Hox gene regulatory regions, indirect effects through other transcription factors, and epigenetic modifications. Various studies in the literature, both *in vitro* and *in vivo* in mice exposed in utero, have demonstrated how exposure to TCDD can alter the expression of the Hox gene family. In mice exposed to TCDD *in utero*, the Hox genes

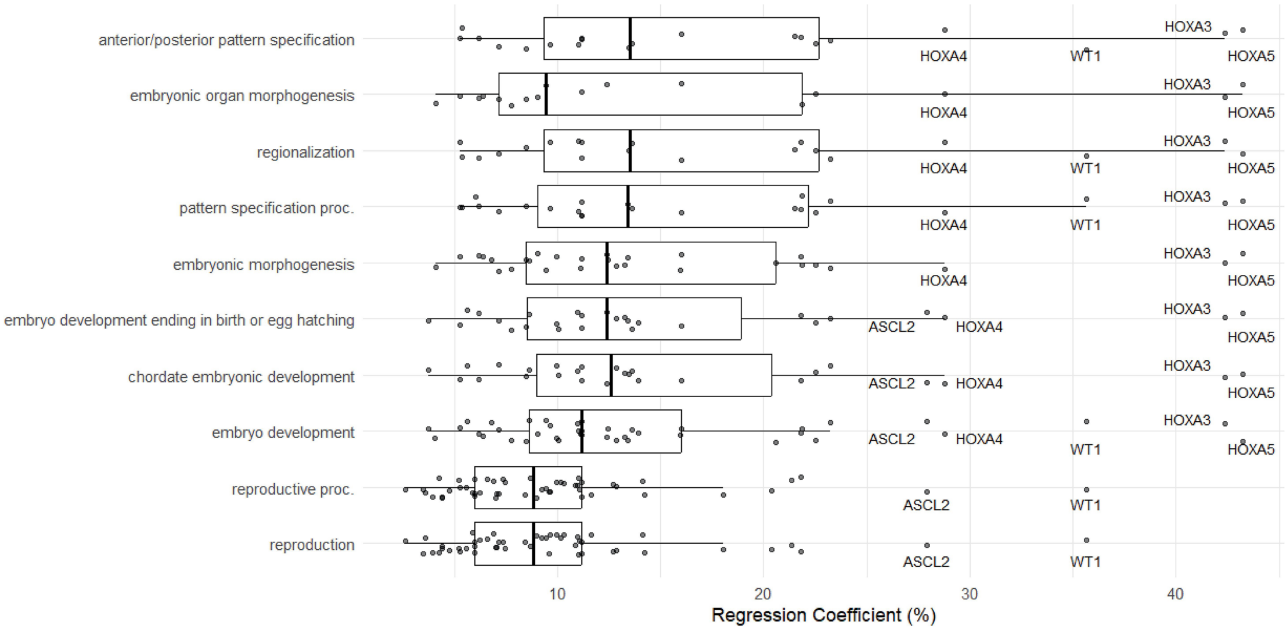

Fig 3. Boxplot showing the relationship between the regression coefficient, expressed in percentage on the *x*-axis, of epimutated genes depicted as dots, and the names of the pathways on the *y*-axis. Additionally, the names of the top five most impacted genes are displayed.

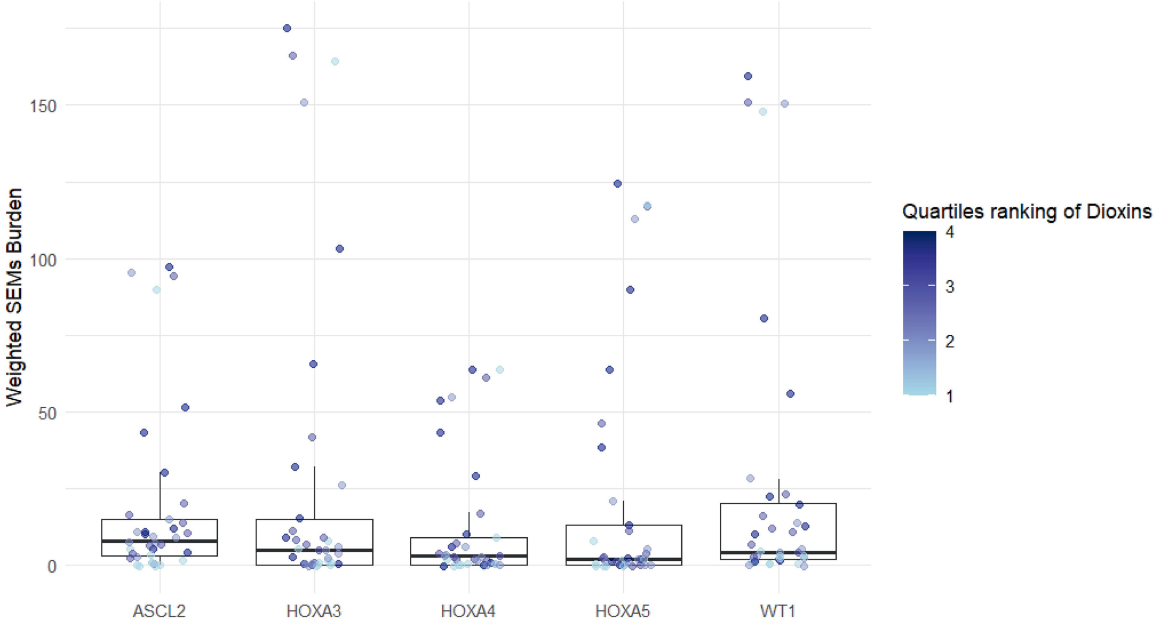

Fig 4. Boxplot showing the number of Weighted SEMs per subject(represented in dots), coloured by TCDD quartile rank, divided for the top five genes most associated. Light blue indicates the lowest TCDD exposure levels, while dark blue represents the highest dioxin values.

were predominantly repressed, leading to developmental issues in organs such as the kidneys and craniofacial abnormalities [58]. *In vitro*, researchers found that treating mouse cardiomyocytes with TCDD alters their differentiation, which manifests as cardiac dysfunction due to inadequate heart formation [59, 60]. In the latter case, the alteration of Hox gene expression caused by TCDD exposure varies, with some genes being overexpressed and others underexpressed, depending on the amount of TCDD, which influences the gene's expression status. The literature does not provide information regarding a direct relationship between TCDD and its ability to alter the methylation status of these genes, particularly in germline tissue. However, the role of methylation status in Hox genes is crucial in various developmental processes, including neural tube formation and muscle cell differentiation [65, 66].

Our study investigated the link between direct exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and SEM accumulation in germline tissue. The limited number and rarity of samples present a challenge in finding inferential statistical associations. The small sample size can make it difficult to detect potentially significant findings due to low statistical power in commonly used approaches, such as differential methylation analysis. Moreover, the scarcity of human TCDD exposure makes obtaining additional datasets for aggregate results in methodologies such as

meta-analysis difficult. However, in their study, Kelsey *et al.* were able to identify methylated regions of the H19 and TEAD3 genes in the same dataset, with nominally significant *P*-value and great beta-value differences when comparing high-exposed subjects with lower-exposed ones [41]. Based on these results and findings from the literature that shows how dioxin alters the methylome in animal models, we wondered if there might be other epigenetic changes that are difficult to detect using methods that heavily rely on statistical power. Additionally, we questioned whether exposure to TCDD is associated with epigenetic drift, which is indicated by the accumulation of scattered epigenetic alterations throughout the entire genome. To assess this aspect, we used stochastic epimutations as markers through the method introduced by Gentilini *et al.*, and we applied a new approach to the method defined by Gentilini, assigning intensity weights to the epimutations. Our approach allowed us to move from a binary 'epimutated or not' to a more nuanced five-level scale, ranging from 'not epimutated' to 'epimutated level 4'. This new methodology approach streamlines the concept of epimutation and provides more statistical stability for our association models. As a primary result, from a statistical point of view, our analysis identified 90 genes that showed evidence of significant association with TCDD levels but with a small beta coefficient figure. Furthermore, we proceeded to do an association analysis with each gene's EML (epimutation load). Given the result obtained, we corrected the *P*-value for multiple tests using the Benjamini–Hochberg methodology and identified 437 genes that might be affected by dioxin exposure. Therefore, to assess the biological processes affected by these altered genes, we conducted a pathway analysis of the biological processes. Figure 2 shows the significantly enriched processes mostly related to embryonic development and reproduction. The primary genes that enriched these identified pathways belong to the Hox family and are HOXA2, HOXA3, HOXA4, HOXA5, HOXA6, HOXD4, and HOXD9. In our case, the identified Hox genes are all commonly hypermethylated, a state typically associated with gene repression, which aligns with some of the abovementioned aspects. Specifically, the hypermethylation of Hox genes, HOXA3 and HOXD9, which we also identified, has been characterized by tissue ageing [67].

Previous studies on this dataset have highlighted how exposure to TCDD is associated with increased biological age, typically linked to a rise in SEMs characteristic of epigenetic drift [42,45,68]. These mutations have been considered valuable markers for quantifying this effect [45]. We, therefore, hypothesize that the increase in epigenetic drift observed in found genes might result from sperm ageing induced by TCDD exposure, potentially leading to harm in both the offspring and the directly exposed individuals. This study focuses on multigenerational effects, where direct exposure can affect the health of offspring even if they were not directly exposed. The proposed mechanism involves epigenetic alterations, specifically DNA methylation changes, in sperm cells. DNA methylation regulates gene expression and is susceptible to environmental influences. While the analysis identified statistically significant associations between TCDD exposure and methylation levels in 90 genes (Supplementary Data 1), the overall effect size was small. The median methylation level increase across these genes due to TCDD exposure was negligible, and pathway analysis revealed no significant impact on biological processes or molecular functions. However, after adjusting for multiple tests, a more detailed analysis focusing on individual gene regions identified a much larger number of genes (437) with significant associations with TCDD exposure. These genes displayed a wider range of beta coefficients, indicating a variable increase in SEM burden associated with TCDD exposure, ranging from 1% to 43% per unit increase in serum dioxin levels. The most compelling finding of this study is the identification of 10 significantly enriched biological pathways associated with the 437 genes linked to TCDD exposure. These pathways primarily fell into two categories: embryonic and general organismal development and reproduction. Notably, the pathways related to development appeared to be more affected by TCDD exposure, as evidenced by a higher percentage increase in SEM burden in the associated genes than those involved in reproduction as observable at Figure 2. This finding suggests that TCDD exposure might have a more pronounced effect on early development stages.

Since we do not have methylation data from the offspring, we cannot definitively state that the identified alterations in these genes are hereditary and causative of phenotypic alteration in the offspring. However, several studies on animal models demonstrate that alterations in various Hox genes in germline tissue can be inherited, manifesting as phenotypic abnormalities that influence offspring outcomes [69, 70].

Unfortunately, the literature on these aspects is lacking concerning humans due to the difficulty of studying DNA methylation multigenerational effects under conditions of exposure to TCDD, despite known serum levels. In conclusion, dioxin exposure poses a significant risk to human health by disrupting the regular expression of Hox genes. Understanding the mechanisms behind these effects is crucial for developing strategies to prevent and treat dioxin-related diseases.

Study limitations

We conducted the research based on a specific dataset (GSE139307) involving Vietnam veterans directly exposed to Agent Orange, which does not represent the broader population. Furthermore, while we used the Benjamini–Hochberg method to adjust for multiple tests, a larger sample size will increase the statistical power to confirm the observed associations and provide more robust estimates of effect sizes. Additionally, the functional consequences of the identified SEMs remain unclear.

Future perspectives

Future research should explore the functional consequences of the identified SEMs to determine whether these methylation changes translate into altered gene expression and ultimately impact developmental or reproductive health outcomes. Additionally, it should investigate the mechanisms by which TCDD exposure might lead to increased SEM burden, specifically in developmental pathways.

Conclusions

Our study provides valuable insights into identifying genetic variations associated with specific biological factors with initial evidence, suggesting that TCDD exposure might be related to alterations in the germline methylome, potentially affecting developmental processes. Despite sample size limitations, our approach offers a novel method to explore genes affected by specific biological factors. Although our study may not provide definitive answers regarding the impact of identified variations, it lays the groundwork for future research. By leveraging our approach, researchers can overcome sample size limitations and better understand the complex interplay between genetic variations and biological factors. In summary, while our study has significant limitations, it

Table 3. Summary of the phenotypic variables, divided by exposed and non-exposed subjects

	Exposed (*N* = 26)	Not exposed (*N* = 11)	Overall (*N* = 37)
Serum TCDD (ppt)			
Mean (SD)	32.5 (38.5)	4.27 (1.37)	24.1 (34.7)
Median [Min, Max]	17.1 [5.62, 168]	4.09 [2.21, 7.05]	12.1 [2.21, 168]
Age (years)			
Mean (SD)	74.0 (8.31)	70.7 (6.81)	73.0 (7.94)
Median [Min, Max]	69.0 [65.0, 94.0]	68.0 [65.0, 84.0]	69.0 [65.0, 94.0]
BMI			
Mean (SD)	29.6 (3.75)	32.4 (5.67)	30.4 (4.52)
Median [Min, Max]	28.5 [23.9, 38.0]	31.7 [22.9, 43.7]	29.7 [22.9, 43.7]
Smoking			
Current	3 (11.5%)	1 (9.1%)	4 (10.8%)
Former	15 (57.7%)	8 (72.7%)	23 (62.2%)
Never	8 (30.8%)	2 (18.2%)	10 (27.0%)

Reported are the mean values with their respective standard deviations and medians with minimum and maximum ranges for continuous variables such as serum dioxin, age, and Body Mass Index (BMI), as well as the percentage prevalence for categorical variables such as smoking status.

offers a promising avenue for further investigation. Although the clinical implications of these altered genes identified in the sperm remain speculative without further functional studies, our results can aid future investigations.

Materials and methods

Dataset description

In this study, we analysed the dataset derived from the Air Force Health Study, a long-term cohort study initiated in 1982 to assess health outcomes among American veterans, particularly those involved in Operation 'Ranch Hand' during the Vietnam War, which exposed them to the herbicide Agent Orange, whose primary contaminant was 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). The specific dataset used in this study is available on the Gene Expression Omnibus (GEO) platform under the accession code GSE139307. It comprises DNA methylation data obtained from sperm samples using the Infinium Methylation 450k technology from 26 veterans who were directly exposed to Agent Orange as part of Operation Ranch Hand. It also includes DNA methylation data from 11 veterans, part of the same cohort, who served in cargo transport operations far from Southeast Asia during the Vietnam war and were, therefore, not exposed to tactical herbicides [71]. In addition to DNA methylation data, the dataset contains various phenotypic variables, such as serum dioxin levels, smoking habits, Body Mass Index (BMI), and age. The 11 veterans who were not exposed to Agent Orange exhibit minimal serum dioxin levels, which are considered as the dioxin background. This information may be useful in reconstructing an increasing trend in serum dioxin levels. Table 3 lists phenotypic variables available for this dataset.

Statistical analysis

Quality control and preprocessing

We utilized the Champ package [72] to conduct a comprehensive quality control assessment to identify and mitigate technical and biological biases commonly encountered in DNA methylation datasets. Our stringent evaluation involved excluding probes with excessively low or high fluorescence values and those that did not display significant β values compared to the background. We also implemented a filtering step to eliminate probes associated with known single-nucleotide polymorphism. Quality control preserved all samples because all have less than 10% missing information at the Cytosine-Phosphate-Guanine dinucleotide (CpG) level.

Since all study participants were male, we decided not to eliminate the probes for the sex chromosome. This decision is consistent with the nature of the tissue under study, as excluding this region would undermine the aims of our research.

After we finished the quality control phase, we conducted a principal component analysis of the probe data, which confirmed the lack of batch effects by showing no significant clustering associated with Sentrix ID. To ensure complete genome coverage, we analysed the percentage of Illumina-designed probes present in our data for each genic region, including Body, 1st Exon, 3'UTR, 5'UTR, TSS200, and TSS1550, as well as for each CpG island region, including n-shelf, n-shore, s-shelf, and s-shore. We represented the coverage in Fig. 5.

Based on these observations, we selected Subset-quantile Within Array Normalization [73; 74], the most appropriate normalization method to reduce differences between samples caused by technical factors unrelated to batch effects.

Differential methylation analysis

We started our research by conducting a thorough analysis of methylation signals, which included making necessary adjustments for confounders such as age, BMI, and smoking habits to ensure accuracy. We used the quantile regression model at the median to perform the analysis. We aimed to identify the genomic regions where we could model the methylation levels with TCDD concentration.

Weighted stochastic epimutation analysis

We used the R package [75] Semseeker's [76] functions to identify the SEMs (semseeker::semseeker) and search genomic regions with a significant burden of epimutations associable with TCDD serum levels (semseeker::association).

Many are the methods to quantify epimutations, including those described by Teschendorff [77], Gentilini [49], and Irizarry [78]. The Gentilini method identifies extreme outliers in the methylation signals and helps apply regression models or statistical tests to the cumulative burden observed in specific probes.

Our study adopted a method based on the Gentilini definition of epimutations, which is compatible with regression modelling. This approach aims to identify signals that overflow significantly from a reference signal interval and assign binary values to probes based on their agreement within or overflowing from this interval. The method uses the range reference calculated using the equations (3) and (4). The method computes the reference range using Q1, the first quartile, Q3, the third quartile, and the interquartile range (IQR).

$$LMin_p = Q1 - (3 \times IQR), \tag{1}$$

$$LMax_p = Q3 + (3 \times IQR). \tag{2}$$

The Gentilini method considers the epimutation a fixed value of one, as in the following equations:

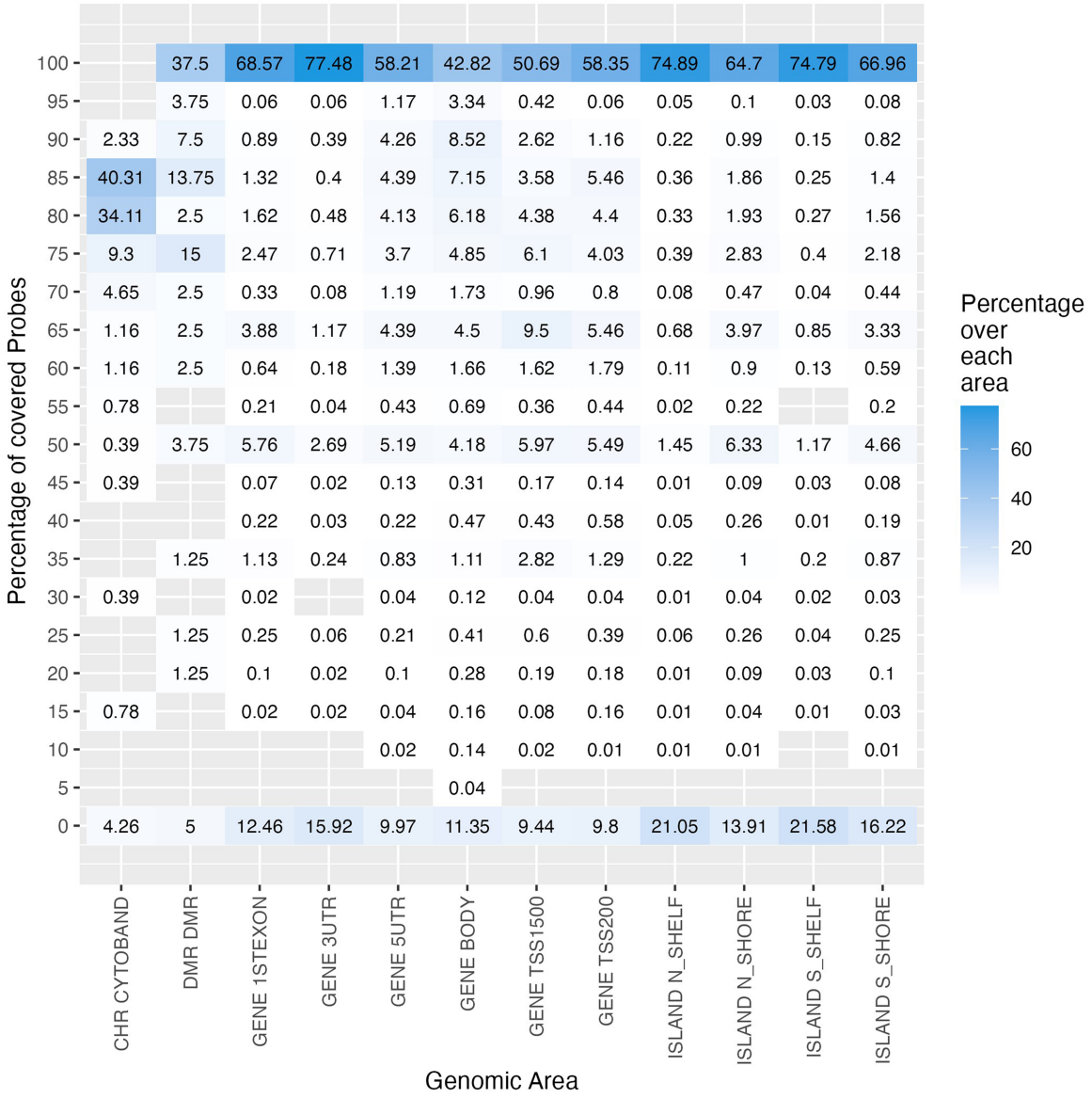

Fig 5. The chart displays the percentage of genomic probe coverage for each specific area (indicated by the columns). Each cell shows the percentage of genomic areas with coverage defined by the corresponding row. The shade of blue illustrates this percentage of the overall genome coverage: darker shades indicate a higher prevalence than the entire genome. The sum of the values per column corresponds to the total coverage of that area. For example, the first row of the first column indicates 68.57% of all the first exons are 100% covered, while any probe does not cover 12.46% of the exons.

$$\Delta_{\text{hyper}} = \begin{cases} M_{\text{value}} - LMax_p, & \text{if } M_{\text{value}} > LMax_p, \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$\Delta_{\text{hypo}} = \begin{cases} LMin_p - M_{\text{value}}, & \text{if } M_{\text{value}} < LMin_p. \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Furthermore, we calculated abs(Δ_{hyper}) and abs(Δ_{hyper}) associated quartiles across all samples and probes to weigh the overflowing signals, as done by Corsaro *et al.* [79]. This operation allows us to observe different weights, indicating their significance. Converting the signal to quartiles is necessary to avoid misleading it as an absolute value because the overflowing signal starts at zero. Still, this zero is, *de facto*, the boundary of the benchmark range.

We used the control group's methylation signal interval as the reference range, as pictured in Image A of Fig. 6. To identify the epimutation, we use the residual signal above (blue) and below (orange) the benchmark range as in Image C of Fig. 6. A positive residual value indicates hypermethylation (blue), while a negative value indicates hypomethylation (orange).

To quantify the importance of the epimutations, we calculated the signal strength ratio outside the reference interval to the interval itself as in Image D of Fig. 6, thus assessing the relative importance of these deviations. In addition, we used quartiles to assign integer values to these ratios as in Image E of Fig. 6. In this way, we assigned a weight to the epimutated probes to mimic the importance of the values based on their deviation from the reference interval, even though they have signals close to zero, defined as the edge of the reference interval.

The hypermethylated probes in Image A of Fig. 6, coloured in blue, exceed the range by nearly five times the average excess of the hypomethylated probes. The delta (DELTAS) depicted in Image C of Fig. 6 ranges between 0 and 0.4. At the same time,

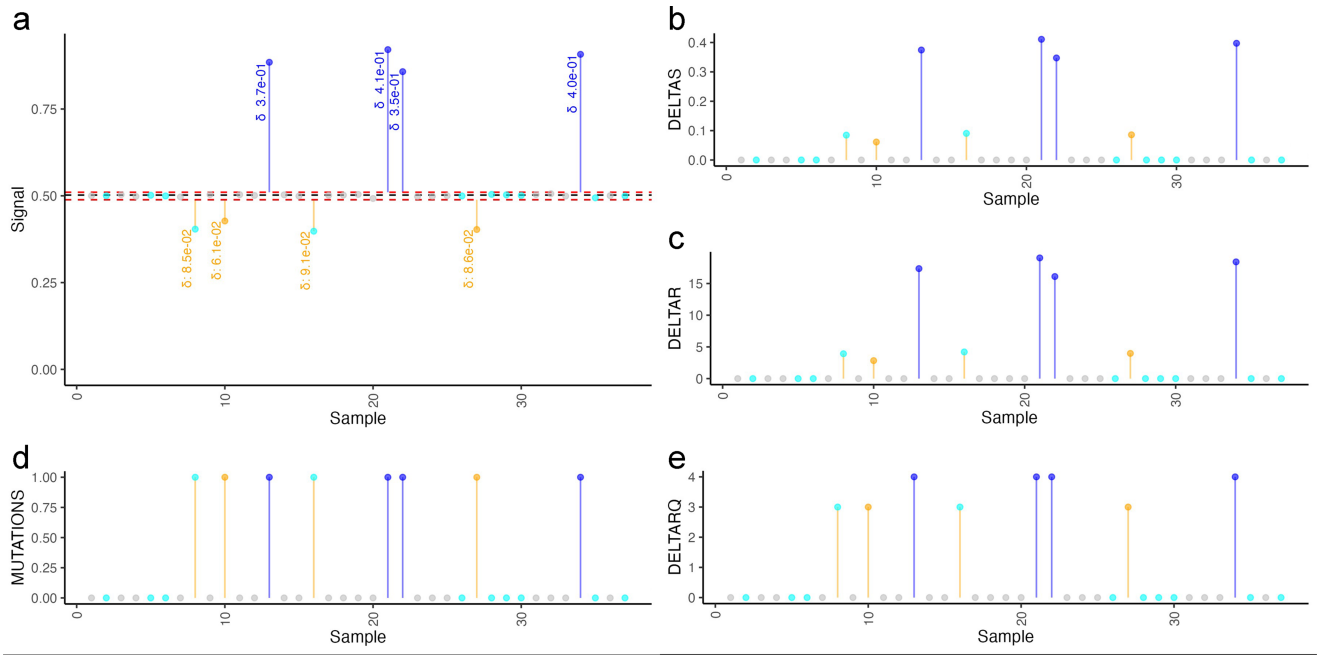

Fig 6. The signal analysis, as β values, of a specific probe across all study participants. The vertical axis represents the signals, while the horizontal axis lists the subjects, each marked distinctly. (A) demonstrates the calculation method for the reference range of the probe in question across all analysed subjects. Values determined by formulas (1) and (2) are highlighted in red, and dashed red lines denote the boundaries of the range. Instances where the probe/subject signals fall outside the range are coloured blue for hypermethylation and orange for hypomethylation, with subjects within the range in grey. Furthermore, the subjects of the reference population are coloured in cyan. (B) displays the presence of the probe in subjects according to the criteria defined by Gentilini, where a value of one indicates presence. The colour code remains unchanged: blue for hypermethylation, orange for hypomethylation, and grey for no alteration. (C) presents the distribution of probe signals among subjects as shown in A but referenced to zero. The colour code remains unchanged: blue for hypermethylation, orange for hypomethylation, and grey for no alteration. (D) illustrates the ratio of signals exceeding the reference range to the range's excursion for each subject examined with the same probe. Subjects not exhibiting out-of-range values for the probe in question are grey. (E) shows the quartiles assigned to each subject over the entire genome. The colour code remains unchanged: blue for hypermethylation, orange for hypomethylation, and grey for no alteration.

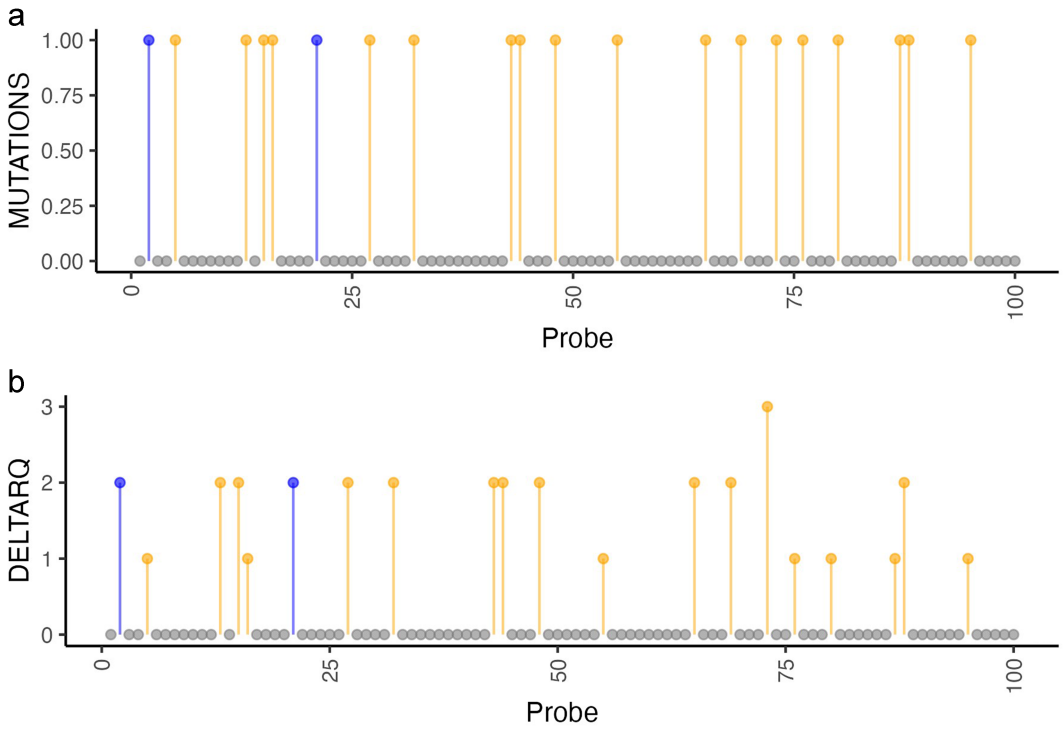

Fig 7. These figures display the burden of epimutations across 100 probes for the same subject, (A) as adopted in this study and (B) according to the Gentilini method.

the ratio to the comparison interval fluctuates between 0 and 10, with the hypermethylated probes standing out compared to the hypomethylated ones. Assigning the number of the associated quartile to the probe highlights how, over the entire genome, the first, fourth, and seventh probes have a numerically higher weight than the second probe. We could end up losing this distinction using only the delta (Image C of Fig. 6) or each probe's ratio to the reference range (Image D of Fig. 6) because both figures start from the absolute zero.

It is crucial to consider the probe's burden while aggregating data. We may combine the burden with different scales if we do not calculate the ratios. For instance, as shown in Fig. 7, if we do not perform ratio calculation and quantisation, we may lose crucial information about the overflowing signal, as illustrated in Image B of Fig. 7. Neglecting the overflowing signal magnitude implies that we could end up having the same total burden even when the importance of the epimutation varies greatly.

To investigate possible associations, we have set the null hypothesis H_0 as the concept that the connection between epimutation load and TCDD (for instance, the quantile at the median regression beta) results from the casualty. In contrast, our alternative hypothesis H_1 suggests that the observed association (the regression beta) is due to the exposure to TCDD. Our analysis used permutation tests to investigate potential associations between epimutation load (EML) and blood TCDD levels for each gene. We broke the relationship between the weighted epimutation load and the measured dioxin levels to create a null hypothesis and generated random associations.

We initially conducted 100 permutations, which produced 100 regression beta coefficients. We then compared the H_1 observed regression beta value within the H_0 (regression beta) confidence interval range. If it is significant, e.g. the alternative hypothesis's regression beta value was outside the null hypothesis's confidence interval, we carried out additional permutations, up to 10 000. We then verified if the regression beta falls out of the H_0 confidence interval accepting the alternative hypothesys H_1, we computed the P_{value} as the number of coefficients of H_0 falling out of the observed H_1 regression beta.

Using this rigorous permutation process, we could validate the alternative hypothesis if its regression beta consistently fell outside the confidential range defined (95%) for the null hypothesis. Additionally, we filtered perturbed genes for an exact $P_{value} <$ .05. We utilized the perturbed genes to identify the biological pathways that were significantly affected. To analyse these pathways, we used the shinyGo web platform [80]. This comprehensive approach uncovered the intricate relationships between epimutations, TCDD exposure, and their collective effects on epigenomic integrity and function.

Author contributions

Luigi Corsaro (Study design, Statistical conception, Preparation of the results, Data visualisation, Writing–introduction, Examination of the genetics of the output, Interpretation of the results, Drafting and Wrote–manuscript, Approval of the final content for journal submission and publication), Davide Sacco (Study design, Data visualisation, Writing–introduction, Examination of the genetics of the output, Interpretation of the results, Drafting and Wrote–manuscript, Approval of the final content for journal submission and publication), Lucy Costantino (Study design, Drafting and Wrote–manuscript, Approval of the final content for journal submission and publication), Carlo Corbetta (Writing–introduction, Examination of the genetics of the output, Interpretation of the results, Drafting and Wrote–manuscript, Approval of the final content for journal submission and publication), Alice Faversani (Editing–manuscript, Approval of the final content for journal submission and publication), Fulvio Ferrara (Editing–manuscript, Approval of the final content for journal submission and publication), and Davide Gentilini (Editing–manuscript, Approval of the final content for journal submission and publication).

Supplementary data

Supplementary data is available at *EnvEpig* online.

Conflict of interest: None declared.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data supporting this study's findings are available in the GEO repository, hosted by the National Center for Biotechnology Information. The dataset(s) can be accessed using the accession number GSE139307. These datasets include raw data files and associated metadata. Researchers can access the data directly from the GEO website (https://www.ncbi.nlm.nih.gov/geo/) by searching for the specified accession numbers. The data are freely available for reuse under the terms provided by the GEO repository. If you need any additional information or assistance accessing the data, do not hesitate to contact the corresponding author.

References

1. Engineering National Academies of Sciences and Medicine. *Veterans and Agent Orange: Update 11 (2018)*. Washington: The National Academies Press, 2018.
2. *U.S. Department of Veterans Affairs. Health and Medicine Division (HMD) of the National Academies of Sciences, Engineering, and Medicine*. https://publichealth.va.gov/exposures/agentorange/ (20 October 2024, date last accessed), 2024.
3. Birnbaum LS. The mechanism of dioxin toxicity: relationship to risk assessment. *Environ Health Perspect* 1994;**102 Suppl 9**: 157–67.
4. Safe S. Polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *Crit Rev Toxicol* 1990;**21**:51–88.
5. Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol* 2003;**43**:309–34.
6. Baccarelli A, Giacomini SM, Corbetta C. *et al.* Neonatal thyroid function in Seveso 25 years after maternal exposure to dioxin. *PLoS Med* 2008;**5**:161.
7. Pesatori AC, Consonni D, Rubagotti M. *et al.* Cancer incidence in the population exposed to dioxin after the 'Seveso accident': twenty years of follow-up. *Environ Health* 2009;**8**:39.
8. Cerlesi S, Domenico AD, Ratti S. 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) persistence in the Seveso (Milan, Italy) soil. *Ecotoxicol Environ Saf* 1989;**18**:149–64.
9. Gao J, Xu Y, Zhong T. *et al.* A review of food contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin and its toxicity associated with metabolic disorders. *Curr Res Food Sci* 2023;**7**: 100617.

10. Michalek JE, Pirkle JL, Needham LL. *et al.* Pharmacokinetics of 2,3,7,8-tetrachlorodibenzo-p-dioxin in Seveso adults and veterans of operation Ranch Hand. *J Expo Anal Environ Epidemiol* 2002;**12**:44–53.
11. Gaspari L, Paris F, Kalfa N. *et al.* Experimental evidence of 2,3,7,8-tetrachlordibenzo-p-dioxin (TCDD) transgenerational effects on reproductive health. *Int J Mol Sci* 2021;**22**:9091.
12. Joffin N, Noirez P, Antignac J-P. *et al.* Release and toxicity of adipose tissue-stored TCDD: direct evidence from a xenografted fat model. *Environ Int* 2018;**121**:1113–20.
13. Orlowska K, Swigonska S, Sadowska A. *et al.* Proteomic changes of aryl hydrocarbon receptor (AhR)-silenced porcine granulosa cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *PLoS One* 2019;**14**:0223420.
14. Wang C, Yuan X-G, Liu C-P. *et al.* Preliminary research on DNA methylation changes during murine palatogenesis induced by TCDD. *J Craniomaxillofac Surg* 2017;**45**:678–84.
15. Viluksela M, Pohjanvirta R. Multigenerational and transgenerational effects of dioxins. *Int J Mol Sci* 2019;**20**:2947.
16. Faiad W, Soukkarieh C, Murphy DJ. *et al.* Effects of dioxins on animal spermatogenesis: a state-of-the-art review. *Front Reprod Health* 2022;**4**:1009090.
17. Gaspari L, Paris F, Kalfa N. *et al.* Experimental evidence of 2,3,7,8-tetrachlordibenzo-p-dioxin (TCDD) transgenerational effects on reproductive health. *Int J Mol Sci* 2021;**22**:9091.
18. Eskenazi B, Warner M, Brambilla P. *et al.* The Seveso accident: a look at 40 years of health research and beyond. *Environ Int* 2018;**121**:71–84.
19. Mocarelli P, Gerthoux PM, Needham LL. *et al.* Perinatal exposure to low doses of dioxin can permanently impair human semen quality. *Environ Health Perspect* 2011;**119**:713–8.
20. Colwell ML, Townsel C, Petroff RL. *et al.* Epigenetics and the exposome: DNA methylation as a proxy for health impacts of prenatal environmental exposures. *Exposome* 2023;**3**:osad001.
21. Toraño EG, García MG, Fernández-Morera JL. *et al.* The impact of external factors on the epigenome: in utero and over lifetime. *Biomed Res Int* 2016;**2016**:2568635.
22. Blomen VA, Boonstra J. Stable transmission of reversible modifications: maintenance of epigenetic information through the cell cycle. *Cell Mol Life Sci* 2011;**68**:27–44.
23. Wright J. Epigenetics: reversible tags. *Nature* 2013;**498**:10–1.
24. Holland N. Future of environmental research in the age of epigenomics and exposomics. *Rev Environ Health* 2017;**32**:45–54.
25. Rozek LS, Dolinoy DC, Sartor MA. *et al.* Epigenetics: relevance and implications for public health. *Annu Rev Public Health* 2014;**35**:105–22.
26. Vineis P, Robinson O, Chadeau-Hyam M. *et al.* What is new in the exposome? *Environ Int* 2020;**143**:105887.
27. Olmedo-Suárez MÁ, Ramírez-Díaz I, Pérez-González A. *et al.* Epigenetic regulation in exposome-induced tumorigenesis: emerging roles of ncRNAs. *Biomolecules* 2022;**12**:513.
28. Bollati V, Biganzoli EM, Carugno M. What if …? A new hypothesis to approach the relationship between environmental stimuli, biological features, and health. *Heliyon* 2023;**9**:22985.
29. Akalin A, Kormaksson M, Li S. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 2012;**13**:87.
30. Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 2016;**32**:1446–53.
31. Biwer C, Kawam B, Chapelle V. *et al.* The role of stochasticity in the origin of epigenetic variation in animal populations. *Integr Comp Biol* 2020;**60**:1544–57.
32. Lu X, Fraszczyk E, Meer TP. *et al.* An epigenome-wide association study identifies multiple DNA methylation markers of exposure to endocrine disruptors. *Environ Int* 2020;**144**:106016.
33. Montjean D, Neyroud A-S, Yefimova MG. *et al.* Impact of endocrine disruptors upon non-genetic inheritance. *Int J Mol Sci* 2022;**23**:3350.
34. Mattonet K, Nowack-Weyers N, Vogel V. *et al.* Prenatal exposure to endocrine disrupting chemicals is associated with altered DNA methylation in cord blood. *Epigenetics* 2022;**17**:935–52.
35. Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends Endocrinol Metab* 2010;**21**:214–22.
36. Manikkam M, Tracey R, Guerrero-Bosagna C. *et al.* Dioxin (TCDD) induces epigenetic transgenerational inheritance of adult onset disease and sperm epimutations. *PLoS One* 2012;**7**:46249.
37. Wu Q, Ohsako S, Ishimura R. *et al.* Exposure of mouse preimplantation embryos to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) alters the methylation status of imprinted genes H19 and Igf2. *Biol Reprod* 2004;**70**:1790–7.
38. Giuliani C, Biggs D, Nguyen TT. *et al.* First evidence of association between past environmental exposure to dioxin and DNA methylation of CYP1A1 and IGF2 genes in present day Vietnamese population. *Environ Pollut* 2018;**242**:976–85.
39. Ehrlich M, Lacey M. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. *Epigenomics* 2013;**5**:553–68.
40. Korrick SA, Lee MM, Williams PL. *et al.* Dioxin exposure and age of pubertal onset among Russian boys. *Environ Health Perspect* 2011;**119**:1339–44.
41. Kelsey KT, Rytel M, Dere E. *et al.* Serum dioxin and DNA methylation in the sperm of operation ranch hand veterans exposed to agent orange. *Environ Health* 2019;**18**:91.
42. Nwanaji-Enwerem JC, Jenkins TG, Colicino E. *et al.* Serum dioxin levels and sperm DNA methylation age: findings in Vietnam war veterans exposed to agent orange. *Reprod Toxicol* 2020;**96**:27–35.
43. Campagna MP, Xavier A, Lechner-Scott J. *et al.* Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin Epigenetics* 2021;**13**:214.
44. Merkel A, Esteller M. Experimental and bioinformatic approaches to studying DNA methylation in cancer. *Cancers* 2022;**14**:349.
45. Gentilini D, Garagnani P, Pisoni S. *et al.* Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with X chromosome inactivation skewing in females. *Aging* 2015;**7**:568–78.
46. Curtis SW, Cobb DO, Kilaru V. *et al.* Exposure to polybrominated biphenyl and stochastic epigenetic mutations: application of a novel epigenetic approach to environmental exposure in the Michigan polybrominated biphenyl registry. *Epigenetics* 2019;**14**:1003–18.
47. Chen GK, Yan Q, Paul KC. *et al.* Stochastic epigenetic mutations influence Parkinson's disease risk, progression, and mortality. *J Parkinsons Dis* 2022;**12**:545–56.
48. Gagliardi A, Dugué P-A, Nøst TH. *et al.* Stochastic epigenetic mutations are associated with risk of breast cancer, lung cancer, and mature B-cell neoplasms. *Cancer Epidemiol Biomarkers Prev* 2020;**29**:2026–37.
49. Gentilini D, Garagnani P, Pisoni S. *et al.* Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with X chromosome inactivation skewing in females. *Aging* 2015;**7**:568–78.
50. Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of

development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA* 2010;**107 Suppl 1**:1757–64.
51. Pilsner JR, Parker M, Sergeyev O. *et al.* Spermatogenesis disruption by dioxins: epigenetic reprograming and windows of susceptibility. *Reprod Toxicol* 2017;**69**:221–9.
52. Yan Q, Paul KC, Lu AT. *et al.* Epigenetic mutation load is weakly correlated with epigenetic age acceleration. *Aging* 2020;**12**:17863–94.
53. Mocarelli P, Gerthoux PM, Needham LL. *et al.* Perinatal exposure to low doses of dioxin can permanently impair human semen quality. *Environ Health Perspect* 2011;**119**:713–18.
54. Mocarelli P, Gerthoux PM, Ferrari E. *et al.* Paternal concentrations of dioxin and sex ratio of offspring. *Lancet* 2000;**355**:1858–63.
55. Chemical Safety and Health Unit (CHE), Nutrition and Food Safety (NFS). *Exposure to Dioxins and Dioxin-Like Substances: A Major Public Health Concern*. 2019. https://www.who.int/publications/i/item/WHO-CED-PHE-EPE-19.4.4 (20 October 2024, date last accessed).
56. Zhang W, Xie HQ, Li Y. *et al.* The aryl hydrocarbon receptor: a predominant mediator for the toxicity of emerging dioxin-like compounds. *J Hazard Mater* 2022;**426**:128084.
57. Habano W, Miura T, Terashima J. *et al.* Aryl hydrocarbon receptor as a DNA methylation reader in the stress response pathway. *Toxicology* 2022;**470**:153154.
58. Kojima T, Takahashi N. Influence of teratogenic factors on mouse 39 hox gene expression. *Biosci Biotechnol Biochem* 2009;**73**:2416–21.
59. Behrens AN, Iacovino M, Lohr JL. *et al.* Nkx2-5 mediates differential cardiac differentiation through interaction with Hoxa10. *Stem Cells Dev* 2013;**22**:2211–20.
60. Wang Y, Fan Y, Puga A. Dioxin exposure disrupts the differentiation of mouse embryonic stem cells into cardiomyocytes. *Toxicol Sci* 2010;**115**:225–37.
61. Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. *Pediatr Res* 1997;**42**:421–9.
62. Whiting J. Craniofacial abnormalities induced by the ectopic expression of homeobox genes. *Mutat Res* 1997;**396**:97–112.
63. Gaunt SJ. Seeking sense in the hox gene cluster. *J Dev Biol* 2022;**10**:48.
64. Hubert KA, Wellik DM. Hox genes in development and beyond. *Development* 2023;**150**:dev192476.
65. Rochtus A, Izzi B, Vangeel E. *et al.* DNA methylation analysis of homeobox genes implicates HOXB7 hypomethylation as risk factor for neural tube defects. *Epigenetics* 2015;**10**:92–101.
66. Tsumagari K, Baribault C, Terragni J. *et al.* DNA methylation and differentiation: HOX genes in muscle cells. *Epigenetics Chromatin* 2013;**6**:25.
67. Turner DC, Gorski PP, Maasar MF. *et al.* DNA methylation across the genome in aged human skeletal muscle tissue and muscle-derived cells: the role of HOX genes and physical activity. *Sci Rep* 2020;**10**:15360.
68. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* 2013;**22**:7–15.
69. Schrott R, Modliszewski JL, Hawkey AB. *et al.* Sperm DNA methylation alterations from cannabis extract exposure are evident in offspring. *Epigenetics Chromatin* 2022;**15**:33.
70. Liu S, Fang L, Zhou Y. *et al.* Analyses of inter-individual variations of sperm DNA methylation and their potential implications in cattle. *BMC Genomics* 2019;**20**:888.
71. Styka AN, Butler DA. The air force health study data and specimens as a resource for researchers. *Mil Med* 2015;**180**:79–84.
72. Tian Y, Morris TJ, Webster AP. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 2017;**33**:3982–4.
73. Müller F, Scherer M, Assenov Y. *et al.* RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol* 2019;**20**:55.
74. Maksimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol* 2012;**13**:44.
75. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023.
76. Corsaro L. *SemSeeker*. 2024. https://github.com/drake69/semseeker (20 October 2024, date last accessed).
77. Teschendorff AE, Jones A, Widschwendter M. Stochastic epigenetic outliers can define field defects in cancer. *BMC Bioinformatics* 2016;**17**:178.
78. Irizarry RA, Ladd-Acosta C, Wen B. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009;**41**:178–86.
79. Corsaro L, Gentilini D, Calzari L. *et al.* Notch, SUMOylation, and ESR-mediated signalling are the main molecular pathways showing significantly different epimutation scores between expressing or not oestrogen receptor breast cancer in three public EWAS datasets. *Cancers* 2024;**15**:4109.
80. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 2019;**36**:2628–9.

Environmental Epigenetics, 2024, 10(1), dvae020 , DOI: https://doi.org/10.1093/eep/dvae020, Advance Access Publication Date: 29 November 2024, Research Article

Received 28 May 2024; **revised** 16 September 2024; **accepted** 28 October 2024