OPEN ACCESS Freely available online

PLOS ONE

Identification of Genomic Features in Environmentally Induced Epigenetic Transgenerational Inherited Sperm Epimutations

Carlos Guerrero-Bosagna[1,2], Shelby Weeks[1], Michael K. Skinner[1]*

1 Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, Washington, United States of America, **2** Department of Physics, Biology and Chemistry, Linköping University, Linköping, Sweden

Abstract

A variety of environmental toxicants have been shown to induce the epigenetic transgenerational inheritance of disease and phenotypic variation. The process involves exposure of a gestating female and the developing fetus to environmental factors that promote permanent alterations in the epigenetic programming of the germline. The molecular aspects of the phenomenon involve epigenetic modifications (epimutations) in the germline (e.g. sperm) that are transmitted to subsequent generations. The current study integrates previously described experimental epigenomic transgenerational data and web-based bioinformatic analyses to identify genomic features associated with these transgenerationally transmitted epimutations. A previously identified genomic feature associated with these epimutations is a low CpG density (<12/100bp). The current observations suggest the transgenerational differential DNA methylation regions (DMR) in sperm contain unique consensus DNA sequence motifs, zinc finger motifs and G-quadruplex sequences. Interaction of molecular factors with these sequences could alter chromatin structure and accessibility of proteins with DNA methyltransferases to alter de novo DNA methylation patterns. G-quadruplex regions can promote the opening of the chromatin that may influence the action of DNA methyltransferases, or factors interacting with them, for the establishment of epigenetic marks. Zinc finger binding factors can also promote this chromatin remodeling and influence the expression of non-coding RNA. The current study identified genomic features associated with sperm epimutations that may explain in part how these sites become susceptible for transgenerational programming.

Citation: Guerrero-Bosagna C, Weeks S, Skinner MK (2014) Identification of Genomic Features in Environmentally Induced Epigenetic Transgenerational Inherited Sperm Epimutations. PLoS ONE 9(6): e100194. doi:10.1371/journal.pone.0100194

Editor: W. Steven Ward, University of Hawaii at Manoa, John A. Burns School of Medicine, United States of America

Received January 2, 2014; **Accepted** May 22, 2014; **Published** June 17, 2014

Copyright: © 2014 Guerrero-Bosagna et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was supported by an NIH NIEHS grant to MKS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: skinner@wsu.edu

Introduction

A number of environmental factors have been shown to induce the epigenetic transgenerational inheritance of disease and phenotypic variation [1,2,3,4,5,6]. The initiation of this transgenerational inheritance process involves exposure of a gestating female and the developing fetus during gonadal sex determination to environmental factors (e.g. toxicants). The exposures promote alterations in the epigenetic programming of the germline that are transmitted to subsequent generations [3,6,7]. A variety of environmental toxicants have been shown to induce the epigenetic transgenerational inheritance of disease including the fungicide vinclozolin [1,3,4], dioxin [2,6], pesticides [5,6], jet fuel hydrocarbons [8] and platicizers (i.e. bisphenol A (BPA) and phthalates) [6]. Environmentally-induced epigenetic modifications in the germline have been shown to involve DNA methylation changes that are transmitted transgenerationally [6]. These germline epigenetic modifications also induce epigenetic alterations in somatic tissues which correlate with transgenerational transcriptome changes [9] and phenotypic abnormalities [10].

Germline epigenetic transgenerational inheritance has been described in several different organisms including plants, flies, worms, rodents, and humans [3,6,11,12,13,14,15]. The role of the germline in the transgenerational process is crucial since it is the only cell that transmits genetic material and stable epigenetic marks (e.g. imprinted genes) to subsequent generations. The initiation of germline development involves a major epigenetic reprogramming through alterations in DNA methylation [16,17,18]. DNA methylation erasure takes place during the migration of primordial germ cells to the genital ridge (before colonization of the gonads), while re-methylation is initiated during gonadal sex determination in a sex specific manner [19,20]. This reprogramming of DNA methylation and the occurrence of other major epigenetic events during primordial germ cell development [21] represents a critical window of exposure for environmental factors [22]. Environmental exposures [23,24] and epigenetic alterations [25] in this developmental window have been shown to promote the epigenetic transgenerational inheritance of disease and phenotypic variation.

Previous studies have shown that different exposures produce distinct sets of transgenerationally altered differential DNA methylation regions (DMR) in male germ cells, termed epimutations [6]. Interestingly, the transgenerationally altered sperm epimutations among these different exposure groups were found to have minimal overlap [6]. The methylation status of these DMR

appears to be transmitted transgenerationally in similar ways to DNA methylation transmission of imprinted genes (imprinted-like mechanism). The DMR identified in these previous studies were found to be exposure specific suggesting potential genomic features among these distinct DMR may exist. The current study was designed to elucidate the potential molecular mechanisms involved in the susceptibility of these epimutations to escape the DNA methylation erasure following fertilization and become transgenerationally programmed.

Recent studies have investigated a variety of DNA-protein interactions that are involved in the establishment of DNA methylation. For example, the presence of protein binding factors in CpG-poor regulatory regions is one feature that would influence DNA methylation [26]. The configuration of specific protein/DNA complexes, formed for example by CTCF or Sp1, can prevent local actions of *de novo* methyltransferases in the genome [27]. The presence of repeat elements has also been reported to influence the establishment of DNA methylation. For example, the presence of CTG/CAG repeats act as a DNA methylation sensitive insulator [28]. Features such as repeat element composition in genomic domains [29] or the composition of nucleotides flanking CpG sites [30] are shown to influence the susceptibility of methylation by Dnmt3 methytransferases. Chromatin marks of histone binding are also shown to correlate with genomic domains in which DNA methylation changes occur [31,32]. The current study used a bioinformatic analysis of patterns of DNA sequence (motifs) in previously identified exposure specific sets of transgenerational DMR. Observations provide insights into the DNA sequence motifs potentially involved in the establishment of these transgenerational sperm epimutations.

A variety of approaches have been used to identify DNA sequence patterns (motifs) that could be functionally relevant from the perspective of gene expression regulation [33,34]. These DNA motifs are known to serve as binding sites for transcription factors and other regulatory factors [33]. Methods to identify DNA motifs have evolved from the visual alignment of a few sequences to the use of complex algorithms and computer programs [34]. Identification of consensus sequences or position weight matrices in genomic regions characterize these DNA motifs [34]. Recently, several algorithms have been developed to identify DNA motifs in a given set of sequences and to determine if they are over-represented compared to that expected by chance [33]. The integration of these computational analyses with experimental techniques is becoming fundamental to identify genome-scale regulatory elements [35,36,37]. Examples of recent studies using motif analysis at a genomic scale include genome-wide identification of estrogen receptor binding sites [38], identification of CTCF-binding sites in the human genome [39] and identification of motifs associated with aberrant CpG island methylation [40]. The current study integrates previously described experimental epigenomic transgenerational data and web-based bioinformatic analyses to identify DNA motifs to help elucidate the molecular mechanisms involved in environmentally induced transgenerational inheritance of sperm epimutations.

Results

The main goal of the current study was to identify genomic features associated with the environmentally induced epigenetic transgenerational inherited sperm epimutations. Previously described transgenerational differential DNA methylation regions (DMR) in the rat sperm were investigated [3,6]. These DMR were identified using a methylated DNA immunoprecipitation (MeDIP) followed by genome wide promoter tiling array (Chip) for an MeDIP-Chip protocol previously described [3,6]. The web-based bioinformatics tool GLAM2 (Gapped Local Alignment of Motifs) [41], which is part of MEME suite [42], was used to identify DNA motifs associated with these transgenerational sperm epimutations. DNA motifs were built from different sets of germline transgenerational DMR derived from different environmental exposures and compared. These were then grouped based on similarities using a familial binding analysis available in the web-based tool for motif analysis termed STAMP [43]. The association tree produced showed two groups of motifs. One branch of the tree was represented only by the "environmentally induced DNA methylation motif 1" (EDM1) (vinclozolin) previously identified [3], while another branch included motifs from the other exposures (plastics, pesticides, dioxin and jet fuel [6]). A familial binding motif representing this branch was named "environmentally induced DNA methylation motif 2" (EDM2) (Figure 1). Interestingly, EDM1 is A/T rich while EDM2 is G rich.

The presence of EDM1 was tested in a variety of transgenerationally altered DMR from sperm and from somatic Sertoli and granulosa cells [44,45]. These transgenerational F3 generation DMR sets included vinclozolin (52 DMR) [3] dioxin (50 DMR) [2], hydrocarbons, jet fuel (33 DMR) [8], pesticide, permethrin and DEET (367 DMR) [5], plasticizers, BPA and phthalates (198 DMR) [46]. The somatic Sertoli cells and granulosa cells were obtained from F3 generation vinclozolin lineage animals. The DMR were identified with a comparative hybridization MeDIP-Chip analysis on F3 generation control versus exposures lineage cells. A subset of the vinclozolin F3 generation sperm DMR that were confirmed with bisulfite-mass spectrometry were also examined separately and termed "confirmed". A computer generated random set of DNA sequences using the same genetic features of size and promoter association was created to act as a control for the comparisons (random occurrence). The presence of EDM1 was found to be significantly increased in the vinclozolin DMR set (52 sequences) and in the "confirmed" subset of 16 sequences [3] when compared to a random occurrence set of computer-generated sequences (Figure 2). Interestingly, EDM1 incidence tended to be decreased in sperm DMR from non-vinclozolin exposures or in sets from somatic cells when compared with its occurrence in a random set of sequences. Significant decreases are observed for the sperm plastics and pesticides DMR groups and for the somatic group of Sertoli cells from the vinclozolin exposure lineage (Figure 2A). The presence of EDM2 motif was also tested against the distinct sets of DMR sequences (Figure 3). It was found that EDM2 was significantly increased in the promoter associated sperm DMR sets of dioxin, plastics and pesticides lineages, and in vinclozolin lineage Sertoli cells when compared with its occurrence in a random set of sequences. The most significant increases were in the plastics and pesticides DMR groups with an over two-fold increase in EDM2 incidence. Therefore, two different motifs were identified with EDM1 being primarily associated with vinclozolin lineage DMR and EDM2 being predominant in a number of the other exposures.

The two DNA motifs identified from the analysis of the DMR groups were then compared to a database of eukaryotic transcription factor binding sites. The top five similarities of known transcription factors binding sites for each motif are shown in Table 1. The presence of motifs of these transcription factor binding sites was then tested against the different exposure lineage sets of DMR sequences (Figure 4). Observations indicate the KROX, SP1, UF1H3-beta and ZNF219 were consistently increased in dioxin, plastics and pesticides groups. However, their incidence in the jet fuel group was variable, with significant increases of only KROX in this group. RREB1 was observed in

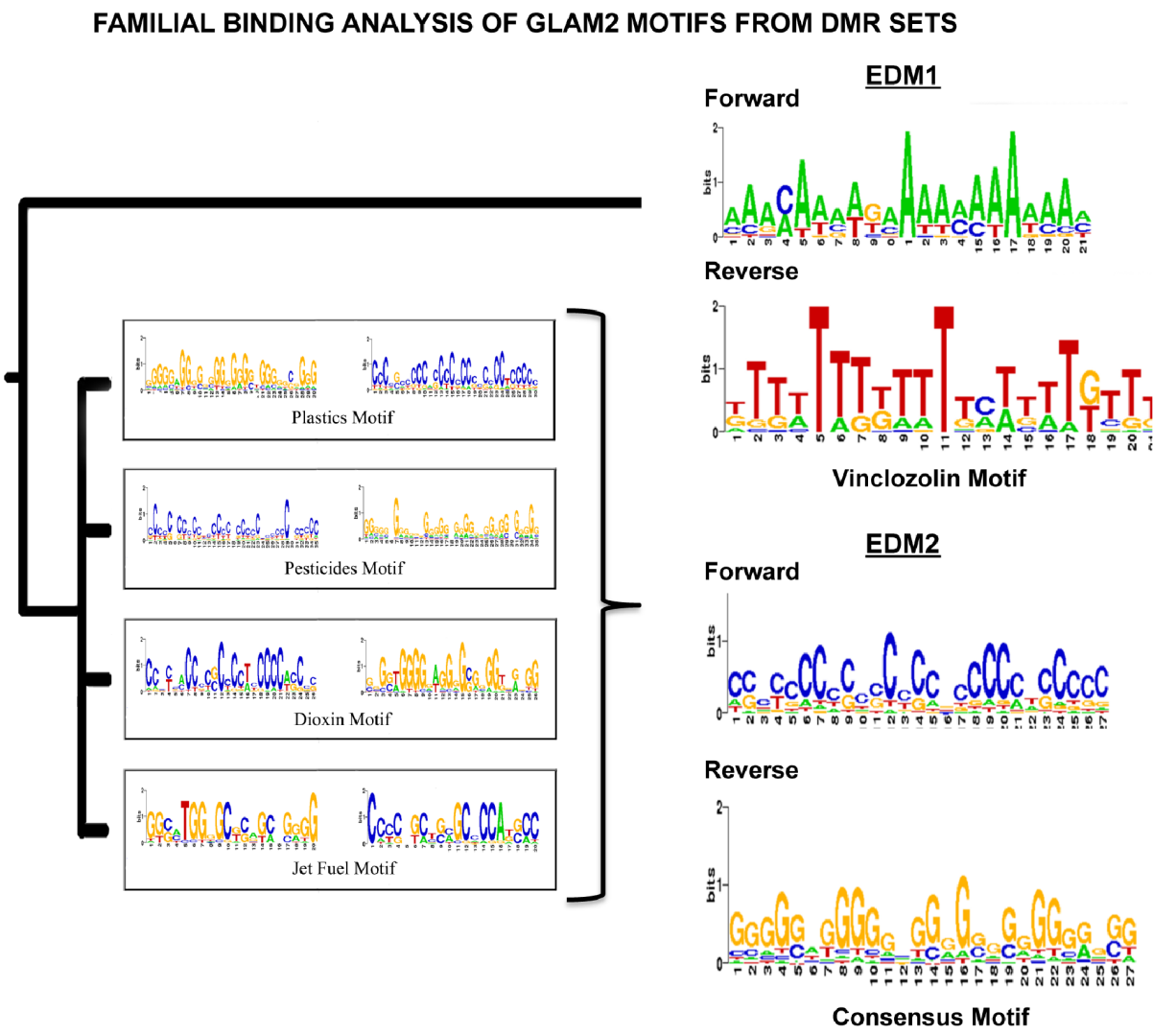

Figure 1. Exposure specific DMR set DNA sequence motifs using GLAM2 and the familial binding tree. The forward (left) and reverse (right) sequences for each motif are presented. The consensus EDM2 motif is presented.
doi:10.1371/journal.pone.0100194.g001

the jet fuel, plastics and pesticides groups. With the exception of UF1H3-beta, all these transcription factors are zinc fingers. FOXP1 was significantly decreased only in the pesticides group. Alfin1 was significantly increased in the plastics and pesticides group, and had a tendency to increase in the dioxin group. Therefore, zinc finger binding sites are apparently associated with the sperm transgenerational epimutations.

The possibility that an altered density of EDM1 or EDM2 motifs could be observed in DMR sets versus the random set was examined. No significant changes were observed for the density of EDM1 or EDM2 in DMR from the exposure groups in reference to the random set (Figure 5). Given the composition of EDM1 in terms of being an A/T rich sequence and the reported role of A/T rich sequences as a recognition site for de novo DNA methylation [47], the density distribution of this feature across the different sets of DMR was also analyzed. Differences in the density distribution of A/T strings were found between the plastics, pesticides, dioxin and jet fuel groups and equivalent random sets of sequences. There is an overall reduction in the density of A/T strings in DMRs from these groups in comparison to the random set (Figure 6 A–E; $p<0.01$). However, A/T string density in the vinclozolin group is similar to the random distribution (Figure 6F).

EDM2 was observed to be a G/C rich sequence. Interestingly, previous reports show that G quadruplexes associate with zinc finger binding sites [48] and have a role in restricting DNA methylation [49] to influence chromatin dependent epigenetic instability [50]. Therefore, the distribution of G-quadruplexes across the different sets of DMR was analyzed. Interesting differences were also found in the distribution of G-quadruplexes in the exposure lineage DMR sets versus the random set of sequences. In the plastics, pesticides, jet fuel and dioxin groups an overall increase in G-quadruplex density regarding the random group was observed (Figure 7 A–E; $p<0.01$). The vinclozolin group had a distribution comparable with the random set (Figure 7F).

Schematic visualization of the features analyzed in selected sets of transgenerational DMRs previously confirmed for the vinclozolin and other exposure DMR groups are shown in Figure 8A,B. The locations of EDM1 and EDM2 in selected sequences are shown in Figure 8. Detailed schematic representations of the locations of these features in selected DMR from the vinclozolin

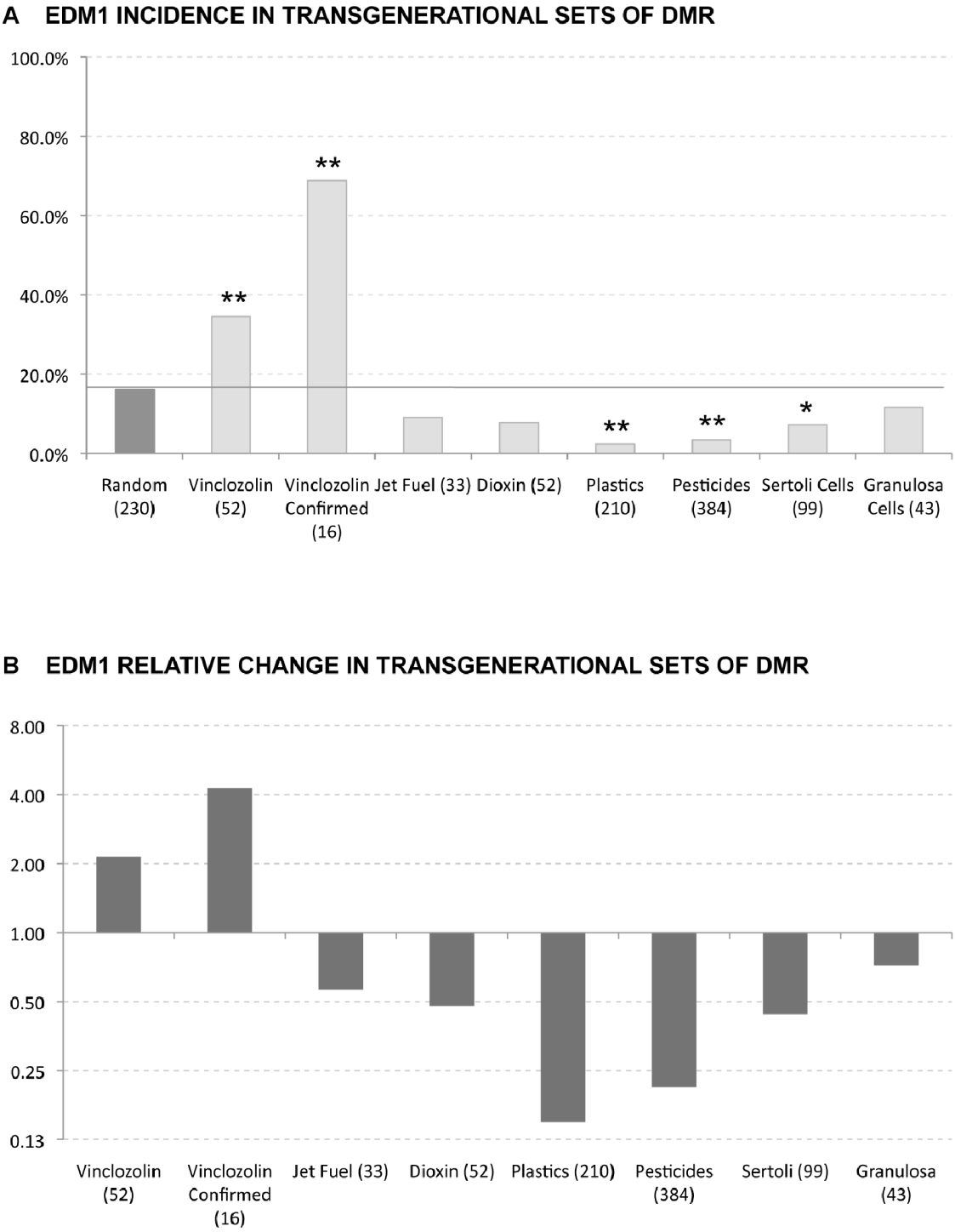

Figure 2. EDM1 incidence in exposure specific epimutation data sets. (A) Individual occurrence (percentage) of EDM1 in a variety of sets with transgenerational DMR and (B) relative change of EDM1 in these DMR. Columns with ** represent significant change with $p<0.01$, while columns with * represent significant change with $p<0.05$.
doi:10.1371/journal.pone.0100194.g002

lineage exposure are shown in Figure 9. Therefore, a number of genomic features were identified and appeared to be associated with the transgenerational sperm epimutations investigated.

A follow up experiment was done to help confirm the observations regarding the genomic features associated with the transgenerational DMR (i.e. epimutations). A more recently developed dichlorodiphenyltrichloroethane (DDT) induced transgenerational set of DMR in F3 generation sperm was investigated [51]. This DDT transgenerational DMR set was not used in the development of the EDM1 or EDM2 sequences, nor the other genomic feature identification. A comparison of the DDT DMR set with the random set of sequences demonstrated a 12.8%

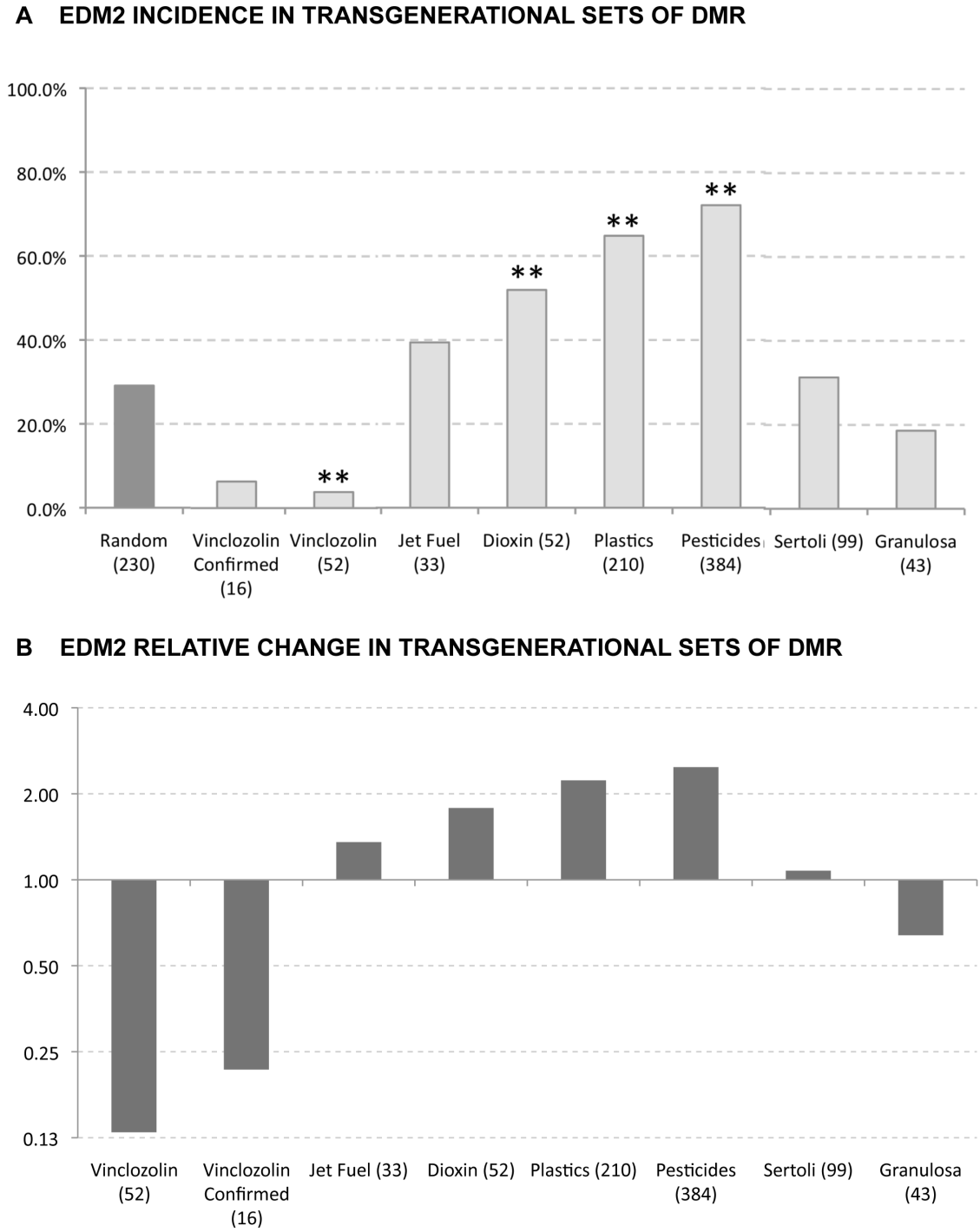

Figure 3. EDM2 incidence in exposure epimutation data sets. (A) Individual occurrence (percentage) of EDM2 in a variety of sets with transgenerational DMR and (B) relative change of EDM2 in these DMR. Columns with ** represent significant change with $p<0.01$, while columns with * represent significant change with $p<0.05$.
doi:10.1371/journal.pone.0100194.g003

incidence in the presence of EDM1, which is similar to the random sequence incidence. A 64.1% incidence of EDM2 is observed, which represents a statistically significant increase ($p<0.01$) compared to the random sequence occurrence. The density distribution of the A/T strings decreased significantly ($p<0.01$) and G-quadruplexes increased significantly ($p<0.01$) in the DDT DMR compared to random sequences (Figure S1). Therefore, the patterns of incidence of EDM1, EDM2, G-quadruplexes and A/T strings in the DDT set are similar to the plastic, pesticide, jet fuel and dioxin DMR sets. Therefore, many of the same genomic features were also present in the transgenerational DDT sperm DMR [51] compared to the random sequences. Although further experiments are required to address the functional importance of these genomic features, this analysis helps confirm the presence of

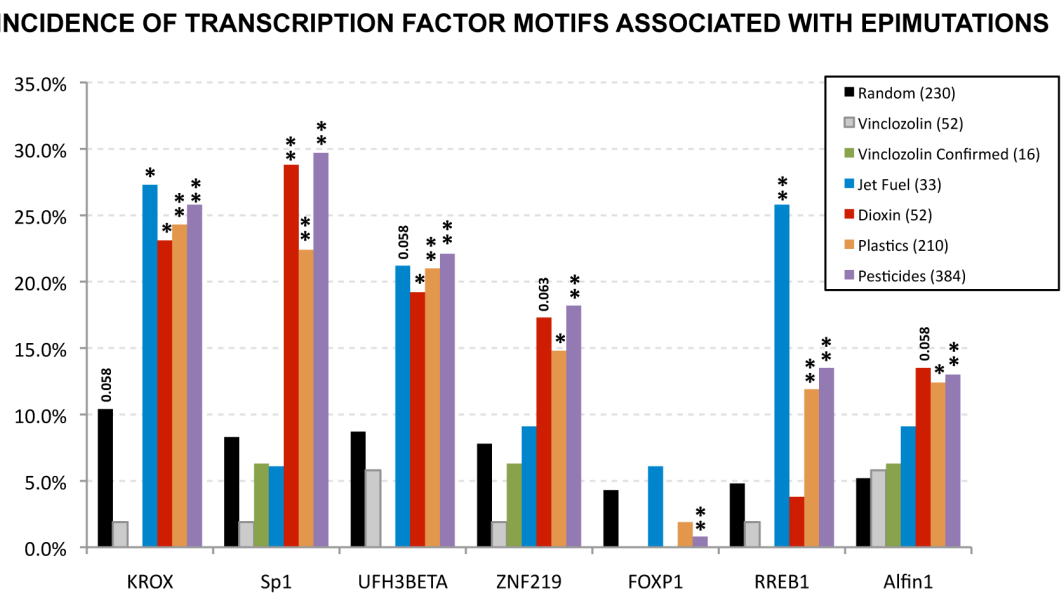

Figure 4. Incidence of consensus motifs of known transcription factor binding sites in DMR from a variety of F3 generation exposure lineage sperm DNA. Columns with (**) represent significant change with $p<0.01$, while columns with (*) represent significant change with $p<0.05$. For nearly significant changes p-values are shown in the respective column. Colored legend for specific exposure lineage DMR sets with total number indicated. The percentage incidence is indicated for the different transition factor sites.
doi:10.1371/journal.pone.0100194.g004

these genomic features in the environmentally induced epigenetic transgenerational inheritance of the sperm epimutations.

Discussion

The phenomenon of environmentally induced epigenetic transgenerational inheritance is a germline mediated process [1,2,3,4,5,6]. Germline epigenetic marks are altered during early fetal development and these environmentally induced epigenetic modifications (epimutations) can be transmitted to subsequent generations. Although the initial genomic feature associated with all the epimutations previously identified was a low CpG density (<12 GpG/100bp), other genomic features are anticipated. The current study is designed to perform a bioinformatic analysis and identify patterns of DNA sequences (motifs) in sperm and somatic cell DMR. Observations are anticipated to provide insights into the potential molecular mechanisms involved in establishing these epigenetic marks.

Identification of DNA sequence motif incidences was performed in different sets of previously identified sperm and somatic cell DMR sequences. Two motifs were identified that are associated with the DMR sets from different environmental exposures. Two different motifs were identified, EDM1, which is an A/T rich motif that is present in the transgenerational vinclozolin DMR set, and EDM2 that is a G/C rich motif present in the other germline transgenerational DMR sets investigated (jet fuel, pesticides, plastics and dioxin). The incidence of EDM1 is over-represented only in the vinclozolin lineage sperm DMRs, but is not over-

Table 1. Similarities of GLAM2 created motifs with known transcription factor for binding sites.

Sequence sets used to build the Glam2 Motif	Best Motif similarities in STAMP database with respective E-values				
	1st	2nd	3rd	4th	5th
Vinclozolin	AZF1	FOXP1	HMG-IY	STE11	BR-C
	1.44E-14	1.53E-10	5.09E-08	4.95E-07	2.11E-02
Dioxin	UF1H3BETA	KROX	RREB1	ZNF219	PAX
	3.41E-14	2.68E-08	2.01E-07	6.39E-07	49.56E-07
Jet Fuel	CDC5	KROX	Dde	IME1	HEN
	2.07E-05	4.26E-05	8.58E-05	9.53E-05	11.05E-04
Pesticides	ZNF219	PAX	KROX	UF1H3BETA	MAZ
	3.04E-10	41.50E-09	3.27E-08	1.74E-08	1.61E-07
Plastics	KROX	RREB1	PAX	ZNF219	UF1H3BETA
	1.25E-12	1.57E-11	41.39E-10	1.89E-09	8.94E-09

Similarity between motifs created from a variety of F3 generation exposure lineage sperm DNA and known transcription factor binding site matrices. The top five similarities are shown for each created motif with their respective statistically significant E-values.
doi:10.1371/journal.pone.0100194.t001

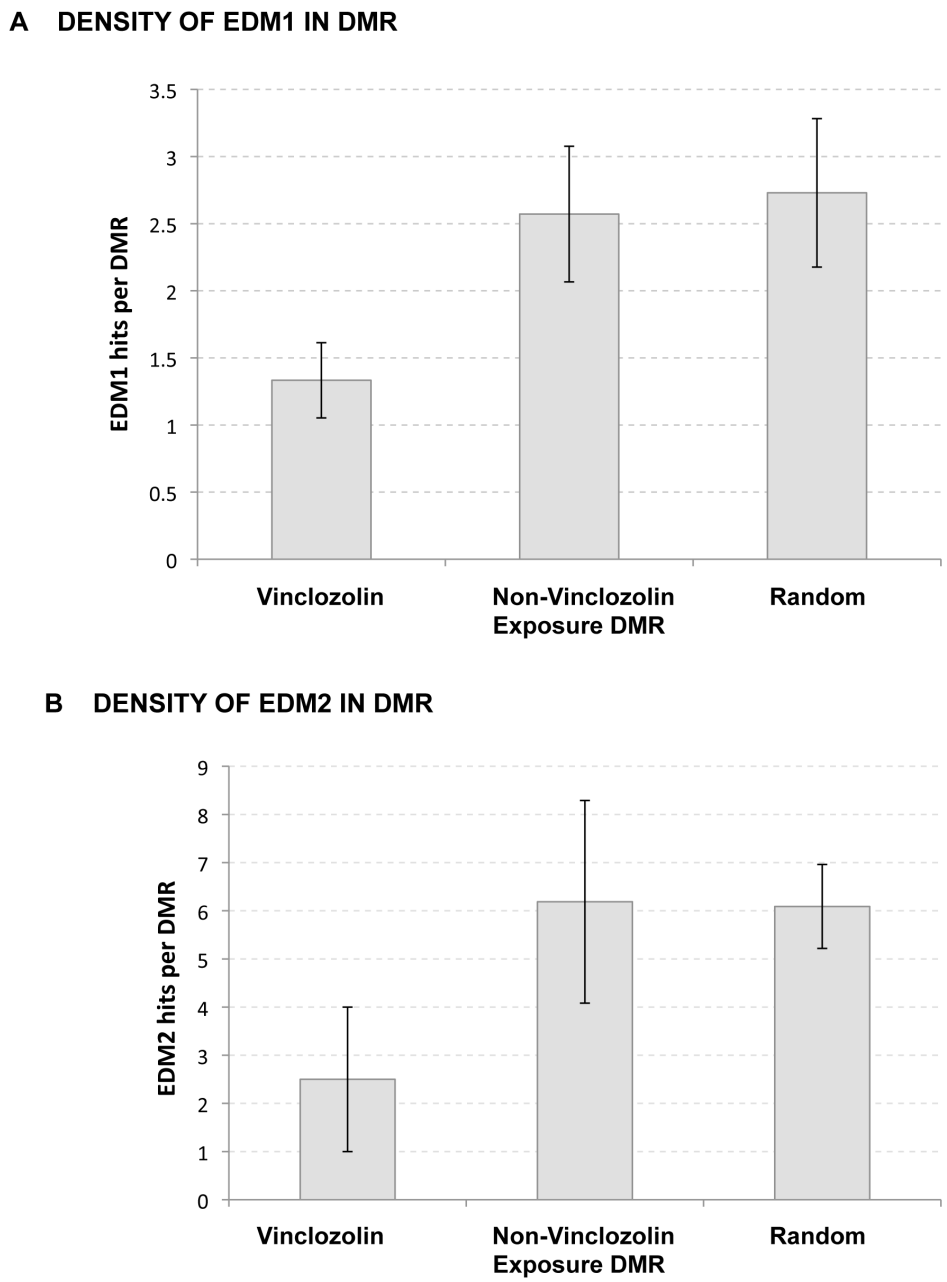

Figure 5. Density of incidences of (A) EDM1 and (B) EDM2 in sets of transgenerational DMR. The vinclozolin DMR, combination of plastics (BIP), pesticide (PIP), jet fuel (JIP) and dioxin (HIP) DMR, and a random set of genomic sites were investigated. The number of EDM1 or EDM2 sites per DMR is presented with the mean± SEM.
doi:10.1371/journal.pone.0100194.g005

represented in the somatic cell vinclozolin lineage DMR sets tested. These observations suggest the somatic cell epigenetic modifications are secondary to the germline epigenetic modifications and probably occur through alternate developmental mechanisms. EDM2 shows the opposite pattern of EDM1, being over-represented in the DMRs of all the exposure lineages, except for vinclozolin.

An analysis of known motifs with a database of transcription factor motifs shows that zinc finger motifs are associated with the sperm epimutations. Further analysis of the presence of these transcription factor binding sites in promoter DMR regions of the exposure sets shows that zinc finger transcriptome factor consensus DNA binding motifs for Krox, Sp1, Znf219 and Rreb1 are over-represented in the majority of the transgenerational sperm DMR sets investigated. These observations suggest that zinc finger containing DNA binding factors may have a role in the molecular mechanism of epigenetic transgenerational inheritance of sperm epimutations. UFH3Beta might also be relevant, since it follows the same patterns as these zinc finger binding factors. Interestingly, previous studies have highlighted the role of zinc fingers in epigenetic reprogramming. For example, the zinc-finger protein UHRF1 has recently been shown to have a role in maintaining DNA methylation in specific genomic regions in mammals [52]. Other studies have shown that the zinc-finger ZBTB4 preferen-

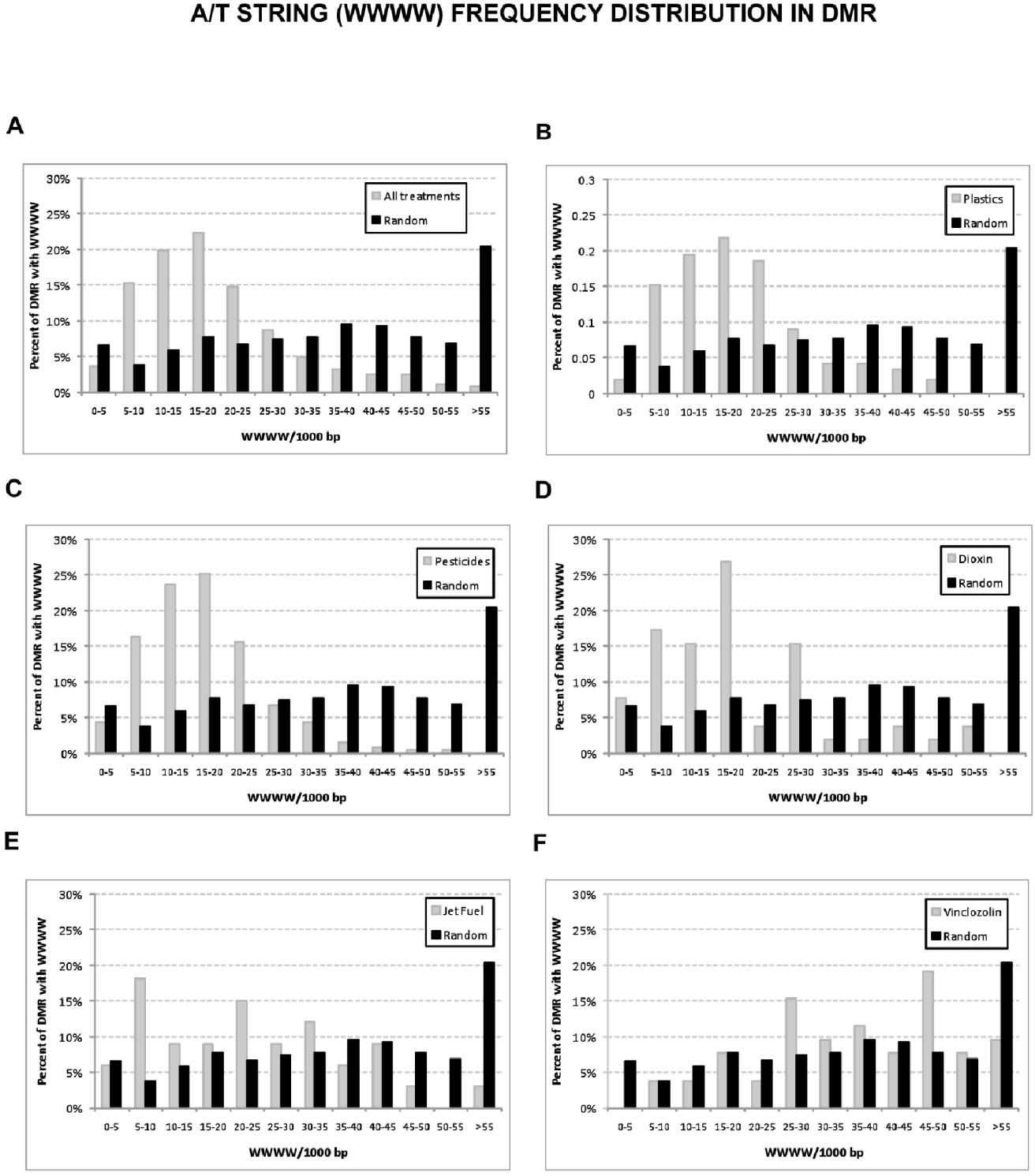

Figure 6. Distribution of A/T string (WWWW) incidence across the transgenerational DMR sets. The percent of DMR with A/T string sequences for all DMR (A), plastics (B), pesticides (C), dioxin (D), jet fuel (E), and vinclozolin (F) are presented compared to the random sequence data set.
doi:10.1371/journal.pone.0100194.g006

tially bind to methylated DNA [53]. The potential that zinc finger binding regions may be targets for DNA methylation changes that will maintain DNA methylation alterations transgenerationally needs to be further investigated. DNA methylation changes in zinc finger rich regions may also preferentially occur in the germline. Another correlation of interest is FoxP1, which is reduced in all but one treatment (jet fuel). Interestingly, FoxP1 expression has been shown to be altered by the hypomethylating agent 5-azacytidine and by micro RNA expression neighboring the FoxP1 gene in human hepatocellular carcinoma cell lines [54]. The

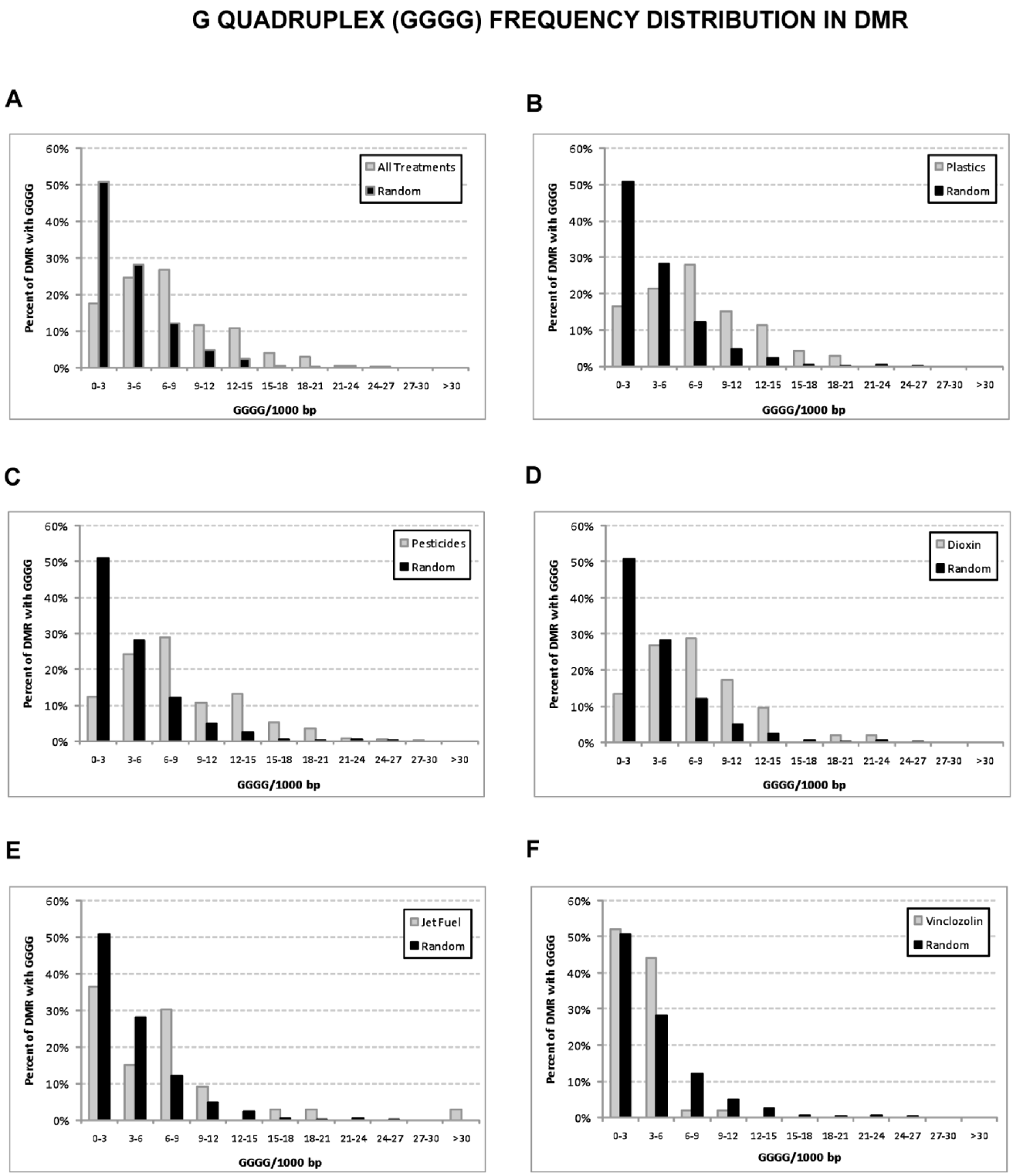

Figure 7. Distribution of G quadruplexes (GGGG) incidence across the transgenerational DMR sets. The percent of DMR with G-quadruplexes of sequences for all DMR (A), plastics (B), pesticides (C), dioxin (D), jet fuel (E), and vinclozolin (F) are presented compared to the random sequence data set.
doi:10.1371/journal.pone.0100194.g007

reduced incidence of FoxP1 sequences in the transgenerational DMR suggests the potential absence of epigenetic mechanisms that may correct the epigenetic defect induced. This may allow for these modifications in DNA methylation to be permanently transmitted to subsequent generations. These potential mechanisms need to be further investigated.

Another genomic feature analyzed was the density of the incidence of EDM1 or EDM2 within the DMR. Because of the

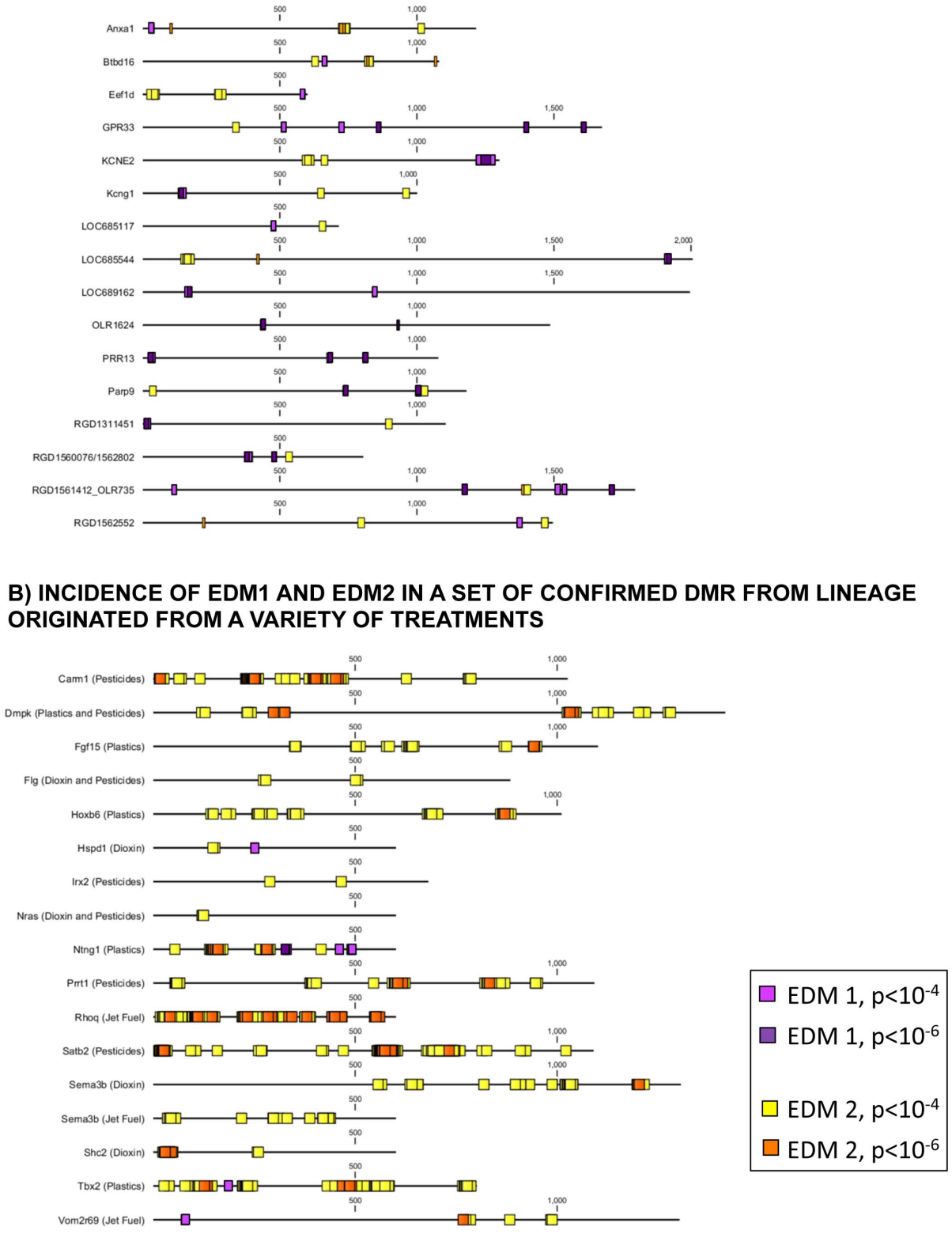

Figure 8. Visualization of DNA motifs associated with epigenetic transgenerational inheritance in selected sets of sequences: (A) incidence of EDM1 and EDM2 in a set of confirmed DMR from vinclozolin-lineage and (B) incidence of EDM1 and EDM2 in a set of confirmed DMR from lineage originated from a variety of exposures listed. The colored legend for EDM1 versus EDM2 motifs are presented.

doi:10.1371/journal.pone.0100194.g008

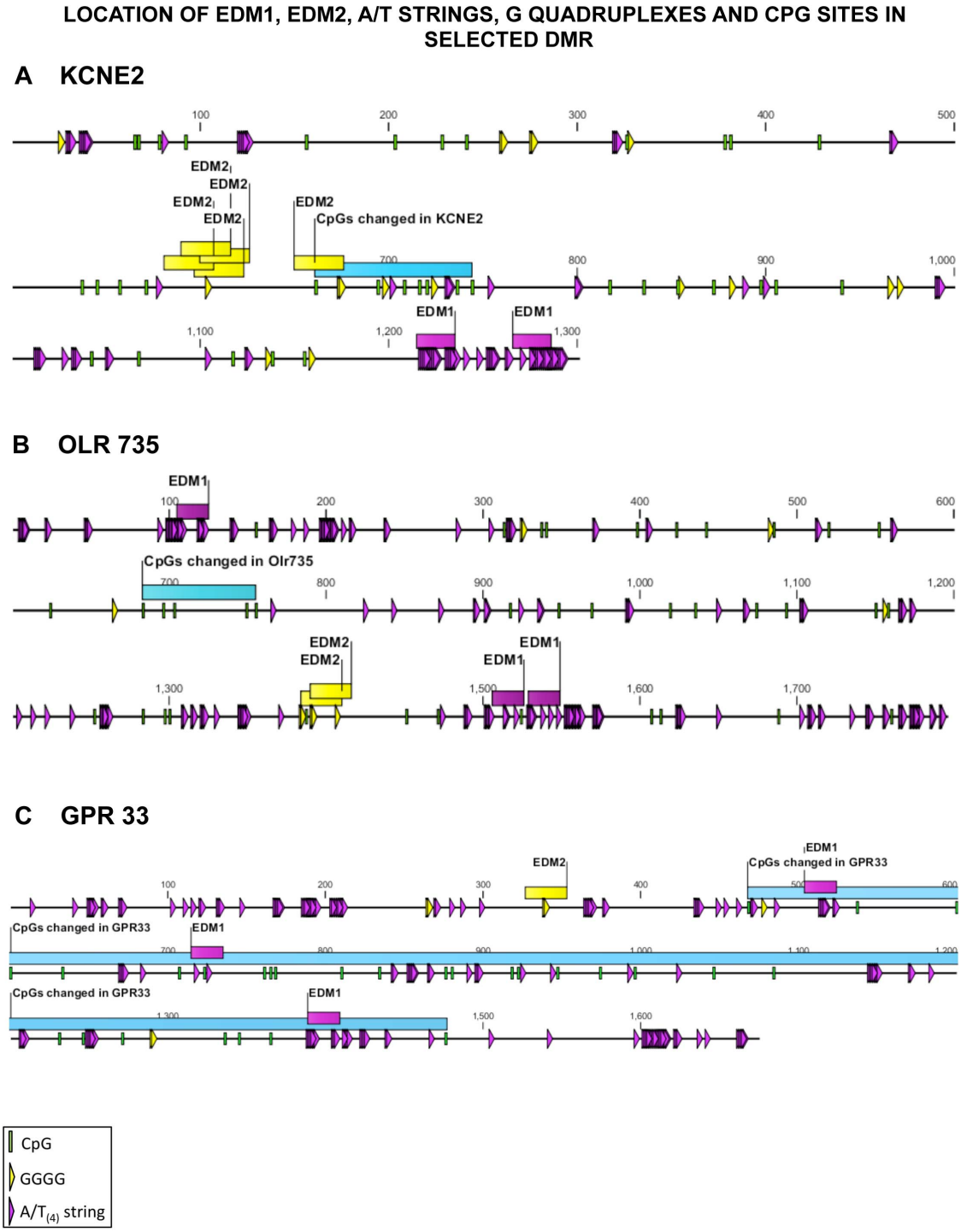

Figure 9. Location of EDM1, EDM2, G quadruplexes, A/T strings and CpG sites in selected DMR from the vinclozolin set: (A) KCNE2; (B) OLR 735; (C) GPR 33. The colored legend for CpG, GGGG sequence, and A/T string is presented, with the blue box being the DMR, yellow box EDM2, and purple box EDM1.
doi:10.1371/journal.pone.0100194.g009

possibility that even if over-representation of individual matches does not occur, a cluster incidence of these motifs might occur. The density of EDM1 or EDM2 in the DMR was determined and no difference was found between any of the exposure DMR groups and the random sequence set (Figure 5). Since EDM1 is A/T rich, the frequency of short A/T strings was also tested in the DMR sets. Interestingly, the incidence of A/T strings (WWWW) is less frequent than in the random sequence set for all the exposures but vinclozolin. In *Neurospora crassa* A/T-rich sequences are shown to be a fundamental recognition site for *de novo* DNA methylation [47]. A/T strings adjacent to CpGs seem to be a requirement for binding of some DNA binding proteins such as MeCP2 [55]. Therefore, the presence of the A/T string is a genomic feature that

contributes to the susceptibility of the epimutations to develop and/or be transmitted.

The other DNA motif obtained was called EDM2 and it was found to have guanine rich regions. Previous studies have shown that CpGs with high methylation are generally not present in G-quadruplexes (GGGG), which suggests that DNA methylation is restricted when G-quadruplex features exist [49]. G-quadruplex unwinding is a conserved mechanism which prevents G-quadruplex-induced damages such as genetic and epigenetic changes [56]. Interestingly, the observations show that the incidence of G-quadruplexes is more frequent in the random sequence group than in all the exposure DMR groups, but vinclozolin. This G-quadruplex conformation forms pockets of accessibility that could open during specific times during development, allowing for epigenetic modifications to be established. Indeed the formation of G-quadruplex structures depends of the RAV1 factor, which when absent alters the incorporation of histones [50]. As mentioned above, zinc finger binding sites are enriched in the transgenerational sperm DMR. Interestingly, previous reports show that G-quadruplexes associate with zinc fingers [48]. Therefore, the presence of G-quadruplexes, zinc fingers and/or chromatin remodeling proteins appear to be associated with the transgenerational sperm epimutations.

The motif associated with the vinclozolin DMR (EDM1) was found to be distinct from the motif associated to the other exposure DMR (EDM2). One speculative mechanism to explain the difference is the variable signaling mechanisms of the compounds generating the transgenerational germline epimutations. While vinclozolin is a known anti-androgenic compound [57], several of the other compounds investigated are associated with estrogenic effects. The estrogenic effects of bisphenol A (BPA) and the phthalates have been established [58]. The action of both permethrin [59] and dioxin [60] are also reported to have estrogenic effects. Jet fuel (JP8) has been reported to reduce LH levels in women [61] by disruption of testosterone conversion to estradiol by aromatase [62]. Although the actions of the toxicants are on the F1 generation fetus, the altered epigenetic programming may be in part different due to the distinct signaling. This potential differential signaling effects on the germline epimutations needs to be further investigated.

An initial experiment to help validate the presence of these genomic features in the transgenerational sperm epimutations used a recent DDT sperm DMR set for analysis [51]. This DDT DMR set was not used in the identification of the genomic features. A number of the genomic features were also found to be present in the DDT sperm epimutations. The distribution of the genomic features in the DDT DMR group has a similar pattern to the plastics, pesticides, jet fuel and dioxin DMR groups. The mechanism of action of DDT is primarily to act as an estrogenic compound [63], which is generally similar to the actions of the other compounds, and distinct from the anti-androgenic actions of vinclozolin. This initial validation helps confirm the presence of the features, however, the functional role of these features remains to be elucidated.

Observations lead to the speculation that the mechanism of the transgenerational epigenetic programming of germline epimutations may be in part based on the action of zinc finger motifs and G-quadruplex sequences that can alter chromatin structure and accessibility to proteins. This alteration may allow an opening of DNA that alters the action of DNA methyltransferases or interacting factors. G-quadruplex regions would be more prone for this opening to epigenetic marks to occur. The zinc finger factors may interact with other proteins to promote this chromatin remodeling and/or altered expression of non-coding RNA. The current study identified a number of motifs and genomic features potentially associated with the DMR involved in the environmentally induced epigenetic transgenerational inheritance of sperm epimutations. Future studies are now needed to further investigate the specific proteins involved and developmental aspects of these epimutations.

Methods

DMR Sequence Sets

Exposure sets of DMR sequences used to perform the bioinformatic analyses were obtained from previous studies from our group showing transgenerational epigenetic changes in the F3 generation sperm and somatic cells [3,6,10,45]. The DMR sequences used were from the published data sets, using a p-value cut-off of 10^{-7} instead of the p-value cut-off of 10^{-5} used in these publications. The DMR data sets were reduced in size, based on the above statistics, to allow for creation of the DNA motifs by the web-based tools used, which have size limitations. The GEO accession number (GSE57693) for these previous publications and additional information on the data access and bioinformatics can be found at www.skinner.wsu.edu/arrays.

Motif Analyses

Two main computational methods exist to identify shared motifs in sets of sequences: (i) application of *ab initio* motif discovery algorithms, which search for recurring patterns in a set of DNA sequences and (ii) assessment of statistical over-representation of previously characterized motifs (from transcription factor binding site databases) in sequences [64]. In the present study the GLAM2 algorithm (Gapped Local Alignment of Motifs, available online on MEME suite) was used for *ab initio* motif discovery. GLAM2 considers insertions or deletions that are a variable and not incorporated by other algorithms [41]. The DMR sets were uploaded to the GLAM2 website and the best motif produced with the default settings was chosen for each sequence dataset. Previously, motifs were identified using GLAM2 in sets of sequences with vinclozolin-induced transgenerational changes in DNA methylation in the male germline [3,4]. The web-based tool FIMO (Find Individual Motif Occurrences, available online on MEME suite) is a general-purpose web-based tool for identifying candidate binding sites [65]. FIMO assigns scores (p and q values) to individual motif occurrences in a defined set of sequences. FIMO was used in the present study to interrogate whether the motifs previously created with GLAM2 would be overrepresented in the tested sets of sequences versus a random set. FIMO was set with a p-value of 10^{-6} and the scan was on both strands. The sequences with matches of motifs were counted and the percentage of the sequences with matches was calculated in for each DMR dataset. The number of motif matches per sequence was also counted to determine the density of motif matches per DMR dataset. In addition to these analyses, comparison of created motifs with known motifs of transcription factors was performed with STAMP. STAMP is a web-based bioinformatic toolbox used to detect similarities of input motifs to motifs representing transcription factor binding sites, which are present in a database that include binding sites information from several organisms [43]. GLAM2 built motifs were compared to the 'selected eukaryotic' database of transcription factor motifs in STAMP, using the default settings. The top five matches (default setting) were selected (Table 1). These selected motifs were then scanned against the DMR sequence datasets using FIMO, as described above. Another feature of STAMP is that it groups motifs based on similarities using a familial binding analysis. This analysis was performed to

contruct a tree of similarities between the GLAM2 built motifs (Figure 1). The incidence of G quadruplexes and A/T strings in DMR sets were computed with R (R Development CoreTeam (2010), R: A language for statistical computing, R Foun-dation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org). For this, matches for GGGG and WWWW motifs were interrogated, respectively. The number of matches per sequences was then obtained and the density of matches per sequence sets was calculated. Figures showing visualization of all motif incidences (GLAM2 built, GGGG and WWWW) were created using CLC Workbench (Cambridge, MA).

Statistical Analyses

Incidences of motifs between the DMR from different treatments and random sets were tested with Fisher's test. Comparison of density of motif incidences between treatment sets of DMR and random sets were performed with Student's t-test. The average of the distributions of G quadruplexes and A/T strings was also compared with Student's t-test between the treatment DMRs and random sets.

Supporting Information

Figure S1 Validation with DDT DMR data set. (A) Distribution of A/T string (WWWW) incidence in the DDT DMR data set. The percent of DMR with A/T string sequences are presented compared to the random sequence data set. (B) Distribution of G quadruplexes (GGGG) incidence in the DDT DMR data set. The percent of DMR with G-quadruplexes are presented compared to a random sequence data set.
(PDF)

Acknowledgments

We thank the assistance and advice of Mr. Md Haque and Dr. Eric Nilsson, and the assistance of Ms. Heather Johnson in the preparation of the manuscript. Current address for Dr. Carlos Guerrero-Bosagna: Avian Group, Department of Physics, Biology and Chemistry (IFM), Linköping University, Linköping, Sweden.

Author Contributions

Conceived and designed the experiments: MKS CGB. Performed the experiments: CGB SW. Analyzed the data: MKS CGB SW. Wrote the paper: MKS CGB. Edited the Manuscript: MKS CGB SW.

References

1. Anway MD, Cupp AS, Uzumcu M, Skinner MK (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308: 1466–1469.
2. Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner MK (2012) Dioxin (TCDD) induces epigenetic transgenerational inheritance of adult onset disease and sperm epimutations. PLoS One 7: e46249.
3. Guerrero-Bosagna C, Settles M, Lucker B, Skinner M (2010) Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS ONE 5: e13100.
4. Guerrero-Bosagna C, Covert T, Haque MM, Settles M, Nilsson EE, et al. (2012) Epigenetic Transgenerational Inheritance of Vinclozolin Induced Mouse Adult Onset Disease and Associated Sperm Epigenome Biomarkers. Reproductive Toxicology 34: 694–707.
5. Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner M (2012) Pesticide and Insect Repellent Mixture (Permethrin and DEET) Induces Epigenetic Transgenerational Inheritance of Disease and Sperm Epimutations. Reproductive Toxicology 34: 708–719.
6. Manikkam M, Guerrero-Bosagna C, Tracey R, Haque MM, Skinner MK (2012) Transgenerational Actions of Environmental Compounds on Reproductive Disease and Epigenetic Biomarkers of Ancestral Exposures. PLoS ONE 7: e31901.
7. Skinner M, Guerrero-Bosagna C, Haque MM, Nilsson E, Bhandari R, et al. (2013) Environmentally Induced Transgenerational Epigenetic Reprogramming of Primordial Germ Cells and Subsequent Germline. PLoS ONE 8: e66318.
8. Tracey R, Manikkam M, Guerrero-Bosagna C, Skinner M (2013) Hydrocarbon (Jet Fuel JP-8) Induces Epigenetic Transgenerational Inheritance of Adult-Onset Disease and Sperm Epimutations. Reproductive Toxicology 36: 104–116.
9. Skinner MK, Mohan M, Haque MM, Zhang B, Savenkova MI (2012) Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions. Genome Biol 13: R91.
10. Nilsson E, Larsen G, Manikkam M, Guerrero-Bosagna C, Savenkova MI, et al. (2012) Environmentally induced epigenetic transgenerational inheritance of ovarian disease. PLoS One 7: e36129.
11. Arico JK, Katz DJ, van der Vlag J, Kelly WG (2011) Epigenetic patterns maintained in early Caenorhabditis elegans embryos can be established by gene activity in the parental germ cells. PLoS Genet 7: e1001391.
12. Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, et al. (2010) Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell 143: 1084–1096.
13. Dunn GA, Bale TL (2011) Maternal high-fat diet effects on third-generation female body size via the paternal lineage. Endocrinology 152: 2228–2236.
14. Morgan CP, Bale TL (2011) Early prenatal stress epigenetically programs dysmasculinization in second-generation offspring via the paternal lineage. J Neurosci 31: 11748–11755.
15. Saze H (2012) Transgenerational inheritance of induced changes in the epigenetic state of chromatin in plants. Genes Genet Syst 87: 145–152.
16. Lees-Murdock DJ, Walsh CP (2008) DNA methylation reprogramming in the germ line. Epigenetics 3: 5–13.
17. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293: 1089–1093.
18. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, et al. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484: 339–344.
19. Allegrucci C, Thurston A, Lucas E, Young L (2005) Epigenetics and the germline. Reproduction 129: 137–149.
20. Durcova-Hills G, Hajkova P, Sullivan S, Barton S, Surani MA, et al. (2006) Influence of sex chromosome constitution on the genomic imprinting of germ cells. Proc Natl Acad Sci U S A 103: 11184–11188.
21. Hackett JA, Zylicz JJ, Surani MA (2012) Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet 28: 164–174.
22. Skinner MK, Manikkam M, Guerrero-Bosagna C (2010) Epigenetic transgenerational actions of environmental factors in disease etiology. Trends Endocrinol Metab 21: 214–222.
23. Cowin PA, Gold E, Aleksova J, O'Bryan MK, Foster PM, et al. (2010) Vinclozolin exposure in utero induces postpubertal prostatitis and reduces sperm production via a reversible hormone-regulated mechanism. Endocrinology 151: 783–792.
24. Uzumcu M, Suzuki H, Skinner MK (2004) Effect of the anti-androgenic endocrine disruptor vinclozolin on embryonic testis cord formation and postnatal testis development and function. Reprod Toxicol 18: 765–774.
25. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, et al. (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A 102: 4068–4073.
26. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, et al. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480: 490–495.
27. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, et al. (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell 146: 1029–1041.
28. Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, et al. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet 28: 335–343.
29. Ichiyanagi K, Li Y, Watanabe T, Ichiyanagi T, Fukuda K, et al. (2011) Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. Genome Res 21: 2058–2066.
30. Wienholz BL, Kareta MS, Moarefi AH, Gordon CA, Ginno PA, et al. (2010) DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. PLoS Genet 6.
31. Alabert C, Groth A (2012) Chromatin replication and epigenome maintenance. Nat Rev Mol Cell Biol 13: 153–167.
32. Wen B, Wu H, Bjornsson H, Green RD, Irizarry R, et al. (2008) Overlapping euchromatin/heterochromatin- associated marks are enriched in imprinted gene regions and predict allele-specific modification. Genome Res 18: 1806–1813.
33. Das MK, Dai HK (2007) A survey of DNA motif finding algorithms. BMC Bioinformatics 8 Suppl 7: S21.
34. Stormo GD (2000) DNA binding sites: representation and discovery. Bioinformatics 16: 16–23.
35. Elnitski L, Jin VX, Farnham PJ, Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res 16: 1455–1464.

36. Nardone J, Lee DU, Ansel KM, Rao A (2004) Bioinformatics for the 'bench biologist': how to find regulatory regions in genomic DNA. Nat Immunol 5: 768–774.
37. Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5: 276–287.
38. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, et al. (2006) Genome-wide analysis of estrogen receptor binding sites. Nat Genet 38: 1289–1297.
39. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, et al. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128: 1231–1245.
40. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM (2006) DNA motifs associated with aberrant CpG island methylation. Genomics 87: 572–579.
41. Frith MC, Saunders NF, Kobe B, Bailey TL (2008) Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput Biol 4: e1000071.
42. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202–208.
43. Mahony S, Benos PV (2007) STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res 35: W253–258.
44. Nilsson E, Larsen G, Manikkam M, Guerrero-Bosagna C, Savenkova M, et al. (2012) Environmentally Induced Epigenetic Transgenerational Inheritance of Ovarian Disease. PLoS ONE 7: e36129.
45. Guerrero-Bosagna C, Savenkova M, Haque MM, Sadler-Riggleman I, Skinner MK (2013) Environmentally Induced Epigenetic Transgenerational Inheritance of Altered Sertoli Cell Transcriptome and Epigenome: Molecular Etiology of Male Infertility. PLoS ONE 8: e59922.
46. Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner M (2013) Plastics Derived Endocrine Disruptors (BPA, DEHP and DBP) Induce Epigenetic Transgenerational Inheritance of Adult-Onset Disease and Sperm Epimutations. PLoS ONE 8: e55387.
47. Tamaru H, Selker EU (2003) Synthesis of signals for de novo DNA methylation in Neurospora crassa. Mol Cell Biol 23: 2379–2394.
48. Kumar P, Yadav VK, Baral A, Kumar P, Saha D, et al. (2011) Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res 39: 8005–8016.
49. Halder R, Halder K, Sharma P, Garg G, Sengupta S, et al. (2010) Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. Mol Biosyst 6: 2439–2447.
50. Sarkies P, Reams C, Simpson LJ, Sale JE (2010) Epigenetic instability due to defective replication of structured DNA. Mol Cell 40: 703–713.
51. Skinner MK, Manikkam M, Tracey R, Nilsson E, Haque MM, et al. (2013) Ancestral DDT Exposures Promote Epigenetic Transgenerational Inheritance of Obesity BMC Medicine 11: 228.
52. Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, et al. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317: 1760–1764.
53. Sasai N, Nakao M, Defossez PA (2010) Sequence-specific recognition of methylated DNA by human zinc-finger proteins. Nucleic Acids Res 38: 5015–5022.
54. Datta J, Kutay H, Nasser MW, Nuovo GJ, Wang B, et al. (2008) Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis. Cancer Res 68: 5049–5058.
55. Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, et al. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell 19: 667–678.
56. Paeschke K, Bochman ML, Garcia PD, Cejka P, Friedman KL, et al. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. Nature 497: 458–462.
57. Wong C, Kelce WR, Sar M, Wilson EM (1995) Androgen receptor antagonist versus agonist activities of the fungicide vinclozolin relative to hydroxyflutamide. J Biol Chem 270: 19998–20003.
58. Singh S, Li SS (2012) Bisphenol A and phthalates exhibit similar toxicogenomics and health effects. Gene 494: 85–91.
59. Brander SM, He G, Smalling KL, Denison MS, Cherr GN (2012) The in vivo estrogenic and in vitro anti-estrogenic activity of permethrin and bifenthrin. Environ Toxicol Chem 31: 2848–2855.
60. Tanaka J, Yonemoto J, Zaha H, Kiyama R, Sone H (2007) Estrogen-responsive genes newly found to be modified by TCDD exposure in human cell lines and mouse systems. Mol Cell Endocrinol 272: 38–49.
61. Reutman SR, LeMasters GK, Knecht EA, Shukla R, Lockey JE, et al. (2002) Evidence of reproductive endocrine effects in women with occupational fuel and solvent exposures. Environ Health Perspect 110: 805–811.
62. Pitteloud N, Dwyer AA, DeCruz S, Lee H, Boepple PA, et al. (2008) Inhibition of luteinizing hormone secretion by testosterone in men requires aromatization for its pituitary but not its hypothalamic effects: evidence from the tandem study of normal and gonadotropin-releasing hormone-deficient men. J Clin Endocrinol Metab 93: 784–791.
63. Robison AK, Schmidt WA, Stancel GM (1985) Estrogenic activity of DDT: estrogen-receptor profiles and the responses of individual uterine cell types following o,p'-DDT administration. J Toxicol Environ Health 16: 493–508.
64. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, et al. (2004) Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res 32: 1372–1381.
65. Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018.