The effect of experimental lead pollution on DNA methylation in a wild bird population

Hannu Mäkinen[1], Kees van Oers[2], Tapio Eeva[1],Veronika N. Laine[2], Suvi Ruuskanen*[1]

[1]Department of Biology, University of Turku, Finland

[2]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

*corresponding author: skruus@utu.fi

Keywords: lead, *Parus major*, pollution, environmental epigenetics, ecological epigenetics, ecotoxicology

Running title: Pollution and DNA methylation in the wild

Abstract

Anthropogenic pollution is known to negatively influence an organism's physiology, behavior and fitness. Epigenetic regulation, such as DNA methylation, has been hypothesized as one mechanism to mediate such effects, yet studies in wild species are lacking. We first investigated the effects of early-life exposure to the heavy metal lead (Pb) on DNA methylation levels in a wild population of great tits (*Parus major*), by experimentally exposing nestlings to lead at environmentally relevant levels. Secondly, we studied the effects of heavy metal exposure in a population close to a copper smelter, where birds suffer from pollution-related decrease in food quality. For both comparisons, the analysis of about million CpGs covering most of the annotated genes, revealed that regions enriched for developmental processes showed pollution-related changes in DNA methylation, but the results were not consistent with binomial and beta binomial regression. Our study indicates that post-natal anthropogenic heavy metal exposure can affect methylation levels of development related genes in a wild bird population.

Introduction

Epigenetic control of gene expression, such as DNA methylation, is increasingly recognized as playing a major role in many different cellular processes. DNA methylation is the addition of a methyl ($-CH_3$) group to the 5' carbon site of cytosines catalyzed by DNA-methyltransferases that occurs mainly at CpG sites in animals. Especially in CpG islands within promotor regions, DNA methylation is found to be negatively associated with gene expression. Epigenetic changes are linked to variation in phenotype and behavior and are associated with prevalence for various diseases (Angers et al. 2010, Rosenfeld 2010, Skinner et al. 2010).

Methylation patterns can be affected by various environmental factors such as maternal nutrition and maternal care (e.g. Weaver et al. 2004, Heijmans et al. 2008, Faulk and Dolinoy 2011, Feil and Fraga 2012), but also by various pollutants and other early-life stressors, both pre- and postnatally, as discovered in humans and mouse models (reviewed by Cheng et al. 2012, Head et al. 2012, Head 2014, Ray et al. 2014, Ruiz-Hernandez et al. 2015). The potential effects of environmental factors on epigenetic regulation is highly important for ecological and ecotoxicological fields, but research in wild vertebrate populations is only emerging (Bossdorf et al. 2008, Head et al. 2012, Liebl et al. 2013, Wenzel and Piertney 2014, Riyahi et al. 2015, Rubenstein et al. 2016, Verhoeven et al. 2016). However, epigenetics can significantly improve our understanding of the mechanisms underlying natural phenotypic variation and the responses of organisms to environmental change (Verhoeven et al. 2016). Furthermore, potential transgenerational epigenetic effects could explain why populations are slow to recover even after pollution removal (Head 2014). A major barrier for genome-wide methylation studies has been the lack of adequate genomic resources such as sequenced genomes in non-model species (Viitaniemi et al. 2019).

Heavy metals, such as lead (Pb), are global, persistent human-induced pollutants that are among the potential contaminants affecting DNA methylation status (reviewed in Bihaqi

2019). For example in human epidemiological studies, developing fetuses show a decrease in global methylation levels as a result of historical maternal lead exposure and accumulation (Pilsner et al. 2009, Wright et al. 2010). Furthermore, in rat, mouse and monkey models, experimental peri- and post-natal lead exposure decreases DNA methyltransferase activity and affects DNA methylation, which are subsequently related to behavioral alternations (Wu et al. 2008, Faulk et al. 2013, Faulk et al. 2014, Luo et al. 2014, Sanchez-Martin et al. 2015, Singh et al. 2018, Nakayama et al. 2019). In birds, metal exposure has been found to affect offspring growth (Burger and Gochfeld 2000) and multiple aspects of physiology, including stress hormone and stress protein levels (Eeva et al. 2014), oxidative stress levels (Koivula and Eeva 2010) and immune function (reviewed in Boyd 2010). However the potential epigenetic alterations by early-life exposure to metal pollution, potentially underlying such effects in birds, have not been studied.

In addition to the direct effect of metals, large-scale metal pollution can decrease resource availability and quality in wild populations (Eeva and Lehikoinen 2004, Eeva et al. 2005), which could subsequently also influence methylation patterns: Along with toxicants, altered nutrition and diet, especially diet poor in methyl donors, are well-known epigenetic modifiers in animal models (reviewed in Choi and Friso 2010, Rosenfeld 2010, Konycheva et al. 2011). Also protein or lipid-altered diets can cause major changes in the epigenome (Burdge et al. 2007, Aagaard-Tillery et al. 2008, Choi and Friso 2010). All in all, we expect populations inhabiting polluted environments to have altered DNA methylation patterns, either due to direct or indirect pollution effects.

Here we investigated whether experimental and anthropogenic early-life exposure to the heavy metal lead (Pb) alters genome-wide DNA methylation status in a wild population of great tits (*Parus major*). More specifically, we experimentally exposed nestlings to dietary lead at levels found close to active pollution sources in Europe and compared to respective

controls. The exposure covered the whole pre-fledging period. Secondly, we studied methylation patterns using nestlings from a population close to an anthropogenic pollution source, copper smelter (Eeva et al. 1997). Around the smelter, nestlings are exposed to multiple metals (in low concentrations) and experience an altered nutritional quality and quantity compared to controls. In the latter case we therefore studied the combined effect of altered nutrition (indirect effect of metal pollution) and direct, low, metal exposure on DNA methylation. In our recent work using the same experimental protocol we found that lead exposure and altered nutrition during nestling development lead to changes in e.g. growth, oxidative stress markers, stress protein levels and vitamin metabolism, but the mechanisms, potentially epigenetic regulation are not understood (Eeva et al. 2014, Rainio et al. 2015b, Ruiz et al. 2016). We used reduced representation bisulfite sequencing (RRBS) to characterize methylation levels, and the sequenced genome of the study species (Derks et al. 2016, Laine et al. 2016, Verhulst et al. 2016) to assess methylation differences in different gene features and their potential functional significance. Given that there are not yet established golden standards for analyzing methylation data in an ecological context, we used and compared two frequently used analytical tools to detect differentially methylated sites (Wreczycka et al. 2017). To avoid sacrificing the individuals, we used blood as a source tissue; previous studies suggest that blood shows similar methylation patterns as brain tissue in the study species (e.g. 80% similarity between brain and blood methylation in CpGs Derks et al. 2016, Verhulst et al. 2016).

Methods

Study species

The great tit (*Parus major)* is a small passerine bird and a model species in ecological and evolutionary research, with ample ecological and genetic background information available.

It is an insectivorous non-migratory bird that commonly breeds in nest boxes, making it an ideal species for experimental manipulations. Importantly, as one of the only non-domesticated bird species, both the genome and methylome are available (Bouwhuis et al. 2010, Derks et al. 2016, Laine et al. 2016, Verhulst et al. 2016).

Lead treatment protocol and sampling

The lead exposure, dosages and sampling are described in detail in (Eeva et al. 2014, Rainio et al. 2015b, Ruuskanen et al. 2015). Briefly, breeding was monitored to record hatching dates of great tit chicks in a population with low pollution levels in southwestern Finland (Turku, 60°26'N, 22°10'E). From day 3 after hatching (hatch date = 0) until day 14 (i.e. in total 12 days) whole broods were subjected to lead with daily oral dosing with the following treatments: HIGH dose (4 µg Pb/g body mass, N = 15 broods) or CONTROL (distilled water, N = 15 broods). The dose represented environmentally relevant exposure levels occurring in polluted areas in Europe. The exposure period covered most of the post-hatching nestling period, i.e. most important developmental period in altricial birds. The third group (hereafter SMELTER) consisted of nests (N = 19 broods) in the vicinity (<2 km) of Harjavalta copper smelter (61°20'N, 22°10'E), where there is a long-term exposure of several metals (e.g., lead, arsenic, cadmium, copper, nickel) and generally lower food availability and quality (Eeva and Lehikoinen 2004). These nests were monitored in the same way as described above and nestlings were dosed with distilled water.

Blood samples were collected from 7-day old nestlings for sex-determination (following Griffiths et al. 1998) and measurements were taken of multiple physiological indices (Eeva et al. 2014, Rainio et al. 2015a). Fresh fecal samples were collected for measuring metal concentrations (see below). Whole blood samples were collected directly in liquid nitrogen from nestlings at age of 14 days (i.e. after 12 days of treatment) for analyses of DNA methylation status and physiological indices. Samples were stored at −80°C until

analysis. Ten unrelated samples from female nestlings in HIGH and CONTROL groups, and eight samples from SMELTER group were selected randomly (total N = 28 samples). The experiment was conducted under licenses from the Animal Experiment Committee of the State Provincial Office of Southern Finland (license number ESAVI/846/ 04.10.03/2011) and the Centre for Economic Development, Transport and the Environment, ELY Centre Southwest Finland (license number VARELY/149/07.01/2011).

Metal analyses

For detailed analyses, see Eeva et al. (2014). Briefly, two fecal samples (one male and one female) from the same brood were combined to assess brood level metal exposure (total N = 35 broods). The determination of metal concentrations (As, Cd, Cu, Ni, Pb) was conducted with ICP-MS with detection limit of 1 ppt (ng/l) and below. The calibration of the instrument was done with a commercial multi-standard from Ultra Scientific, IMS–102, ICP-MS calibration standard 2 and certified reference materials were used for method validation. Data was analyzed with GLMs (SAS 9.4) with Tukey post-hoc tests.

DNA isolation

DNA isolation was performed at the Center of Evolutionary Applications (University of Turku, Finland). We used whole blood samples, which can be acquired without sacrificing the individuals. In birds erythrocytes have nuclei and therefore >95% of the gained DNA is from erythrocytes. DNA was extracted from c. 10 − 20 µl whole blood using the salt extraction method modified from Aljnabi & Martinez (1997). Extracted DNA was treated with RNase-I according to the manufacturer's protocol. DNA concentrations were measured

fluorometrically with a Qubit High Sensitivity kit (ThermoFisher Scientific) and we assessed DNA integrity by running each DNA sample on an agarose gel.

RRBS library preparation

We used a reduced representation bisulfite sequencing (RRBS) approach, which enriches the regions of the genome that have a high CpG content (Meissner et al. 2005). It was previously shown in the study species that the vast majority of the methylated cytocines (97%) were derived from CpG sites in blood (Derks et al. 2016). Sequencing was conducted at the Finnish Microarray and Sequencing Center in Turku, Finland. The library preparation was started from 200 ng of genomic DNA and was carried out according to a protocol adapted from Boyle et al. (2012). The first step in the workflow involved the fragmentation of genomic DNA with MspI where the cutting pattern of the enzyme (C^CGG) was used to systematically digest DNA to enrich for CpG dinucleotides. After a fragmentation step, a single reaction was carried out to end repair and A-tail (required for the adapter ligation) the MspI digested fragments using Klenow fragment (3’ => 5’ exo), following the purification of A-tailed DNA with bead SPRI clean-up method (AMPure magnetic beads). A unique Illumina TruSeq indexing adapter was then ligated to each sample during adapter ligation step to be able to identify pooled samples of one flow cell lane. To reduce the occurrence of adapter dimers, a lower concentration of adapters (1:10 dilution) was used than recommended by the manufacturer. These ligated DNA fragments were purified with the bead SPRI clean-up method before putting samples through bisulfite conversion to achieve C-to-U conversion of unmethylated cytosines, whereas methylated cytosines remain intact. Bisulfite conversion and sample purification were done according to the Invitrogen MethylCode Bisulfite Conversion Kit. Aliquots of converted DNA were amplified by PCR (16 cycles) with Taq/Pfu Turbo Cx Polymerase, a proofreading PCR enzyme that does not stall when it encounters

uracil, the product of the bisulfite reaction, in the template. PCR-amplified RRBS libraries were purified using two subsequent rounds of SPRI bead clean-ups to minimize primer dimers in the final libraries. The high quality of the libraries was confirmed with Advanced Analytical Fragment Analyzer and the concentrations of the libraries were quantified with Qubit® Fluorometric Quantitation, Life Technologies. We selected fragment sizes ranging between 150 − 1000 bp (average sizes were 250-350 bp) for sequencing.

Sequencing

The samples were normalized and pooled for the automated cluster preparation, which was carried out with an Illumina cBot station. The 28 libraries were randomly combined in three pools, 10 or 8 samples in each pool and sequenced in 3 lanes. The samples were sequenced with an Illumina HiSeq 2500 instrument using TruSeq v3 sequencing chemistry. Paired-end sequencing with 2 x 100 bp read length was used with 6 bp index run.

Bisulfite sequencing analysis

The initial quality check with Fastqc (Andrews, 2010) indicated presence of Illumina universal adapter contamination and low quality (Q < 20) bases in the 3’ end of the raw reads. The adapter sequences were removed with Cutadapt (Martin 2014) and the low quality bases were filtered using Condentri (Smeds and Kunstner 2011) with default settings. The quality filtered reads were then mapped against the Great tit reference genome (Assembly Parus_major1.0.3; NCBI Bioproject PRJNA208335, Laine et al. 2016) using Bismark aligner with default parameters (L, 0, -0.2) allowing 2−3 mismatches or a comparable number of indels per 100 bp read (Krueger and Andrews 2011). Methylation information was extracted from alignment files using the bismark_methylation_extractor tool (Krueger and Andrews 2011). The resulting methylation levels per base pair were inspected to detect potential

methylation bias in the beginning and in the end of read 1 and 2 (Hansen et al. 2012). There was lower methylation in the beginning and higher in the end of read 2. Therefore, the first four bases and the last base were removed from the read 2 for subsequent analyses (supplementary figure 1). On average, we recovered 16.08 million raw reads (range 13.55-21.15) from each RRBS library and after quality filtering 11.72 million reads remained (range 10.12-15.06). On average 6.36 million (54%) of the quality filtered reads were uniquely mapped against the Great tit reference genome. This translates to an average 322.45 million cytosine bases analyzed, of which 179.95 million cytocines (56%) were methylated (Supplementary table 1). We estimated the bisulfite conversion rate by aligning the reads against great tit mitochondrial DNA, which is mostly un-methylated (Mechta et al. 2017) and calculated the conversion rate as 1- methylation% in CpG context. The bisulfite conversion rate was 97.7-99.1%.

In order to call methylated CpG sites from the Bismark methylation extractor files, the function *readBismarkCoverage* in the R package Methylkit (Akalin et al. 2012) was used. Using a minimum coverage threshold of 10, on average 1309860 (range 1062814-1545820) methylated CpG sites were obtained for the CONTROL-HIGH comparison and 1344110 (range 1062814-1774992) CpG sites for the CONTROL-SMELTER comparison. The CpGs were then filtered by extreme coverage to remove e.g. potential PCR duplicates using 99.9% percentile upper threshold as implemented in the function *filterByCoverage* in R package Methylkit. Methylated CpG sites were also median normalized to take into account differing library sizes using *normalizeCoverage* function in R package Methylkit. Finally, CpG sites were united such that the final data set contained only CpG sites covered by a minimum of seven individuals per group. The final data sets comprised 1023725 and 903449 CpG sites for CONTROL-HIGH and CONTROL-SMELTER comparisons, respectively. 879056 CpG sites were shared between these two comparisons.

Two methods were used for the identification of differentially methylated CpG sites. First, a generalized linear model was used as implemented in the R package Methylkit. This method assumes that the methylated and un-methylated counts follow a binomial distribution and the effect of group/treatment can be estimated with a log-likelihood test (Akalin et al. 2012). Second, a generalized linear model assuming beta binomial distribution was used taking into account potential overdispersion by estimating a gene-specific shrinkage operator as implemented in R package dss (Wu et al. 2013; Feng et al. 2014). In both methods, the model was fitted for each CpG site separately and we compared the pairwise methylation differences between the CONTROL and HIGH and CONTROL and SMELTER groups. The model fits were evaluated by inspecting the resulting p-value histograms. Under a proper null model one would expect that the p-value histogram follows a uniform distribution, but if there is an effect of e.g. treatment then a surplus of small p-values is expected (Fodor et al. 2007, Barton et al. 2013, Garamszegi and de Villemereuil 2017). Deviations from the uniform distribution may provide information about the misspecification of the model or problems in the data. Differentially methylated regions (DMRs) or clusters of differentially methylated CpG sites were identified based on the results of both binomial and beta binomial models. The following criteria were used to identify DMRs: (i) minimum size of a DMR = 50 bp, (ii) minimum number of CpGs in the DMR = 3 and (iii) the percentage of CpGs with p-value < 0.01 in the cluster = 50%. The identification of DMRs was conducted in the R package dss.

Annotation of differentially methylated regions

The location and association of the CpGs with a given genomic feature was determined using the Great tit genome assembly and annotation 1.1 (Laine et al. 2016). More specifically, each CpG was annotated with respect to location in genes, exons and in coding sequence (CDS).

The annotation of the DMRs was conducted using the *IntersectBed* options in the BedTools package to identify the overlapping genomic features (Quinlan and Hall 2010). STRING database (Szklarczyk et al. 2015) was used to identify gene ontology categories associated with the DMRs. A hierarchical clustering with a user-specified cutoff value C (0.5) was used as implemented in REVIGO database for merging of semantically similar GO categories corresponding to 1% chance of merging two randomly generated categories (Supek et al. 2011). The lowest p-values of the initial enrichment analyses were used to select a representative GO term for each merged category.

Results

Metal exposure

Our dietary lead treatment (HIGH lead) significantly increased fecal lead concentrations compared to the CONTROL (Table 1). At the SMELTER site, we found intermediate fecal lead levels, not significantly different from either HIGH or CONTROL (Table 1). In the SMELTER area, concentrations of other measured heavy metals (As, Cu, Cd, Ni) were higher than in CONTROL or HIGH treatment (Table 1).

DNA methylation

Overall methylation differences among the treatment groups

The range of mean and median coverage was 25.94-33.31 and 17-27 in the CONTROL-HIGH comparison. The respective statistics were 33.73-41.50 and 27-37 in the CONTROL-SMELTER comparison. The average methylation percentages across all CpGs were 27.97 and 28.01 in CONTROL-HIGH lead comparison, and 26.82 and 26.67 in CONTROL-SMELTER comparison. The mean difference in methylation in CONTROL-HIGH and

CONTROL-SMELTER comparisons were −0.0013% and 0.0015%, respectively. There were no marked differences in the methylation percentage using 2% cutoff between the major chromosomes in either comparison (Supplementary figure 2). Also, there were no clear patterns in sample clustering in either of the comparisons based on hierarchical clustering or principal components analysis (Supplementary figure 3, Supplementary figure 4).

Differentially methylated CpG sites

We identified 96377 differentially methylated CpGs in the CONTROL-HIGH comparison and 129830 CpGs in CONTROL-SMELTER comparison using binomial GLM (p-level 0.05 Table 2). Almost equal proportion of the significant CpGs showed hypomethylation (45.9%) in CONTROL and hypermethylation (54.1%) in HIGH. Similarly, 52.5% of the significant CpGs were hypomethylated and 47.5% were hypermethylated CONTROL-SMELTER comparison (supplementary figure 5). 2789 (2.9%) of the significant CpGs showing hypomethylation, and 3022 (3.1%) showing hypermethylation were shared between CONTROL-HIGH and CONTROL-SMELTER, respectively (supplementary figure 5). Beta binomial GLM identified 16852 differentially methylated CpGs in CONTROL-HIGH comparison and 22669 CpGs in CONTROL-SMELTER comparison at p-level 0.05 (Table 2). For beta binomial regression the original p-values were recalculated based on the test statistics as implemented in the R-package fdrtool (Strimmer 2008b). Since the binomial regression method implemented in the R package Methylkit does not report test statistics the re-estimation of p-values were conducted only for the beta binomial regression. After re-estimation of p-values with fdrtool, 62789 and 67418 CpG sites ($p<0.05$) were differentially methylated in CONTROL-HIGH and CONTROL-SMELTER comparisons, respectively (Table 2). Of the significant CpGs, 28297 (45%) were hypomethylated and 34492 (55%) were hypermethylated in CONTROL-HIGH, and 36968 (55%) were hypomethylated and

30450 (45%) were hypermethylated in CONTROL-SMELTER. 946 (1.5%) of the significant CpGs showing hypomethylation, and 947 (1.5%) showing hypermethylation were shared between CONTROL-HIGH and CONTROL-SMELTER comparisons (Supplementary figure 5).

The test statistics of goodness-of-fit test (Chi-square) of the p-value histograms indicated deviations from the uniform distribution in both methods (Table 2). However, the deviation in the beta binomial regression was smaller than the deviation in the binomial regression and the test statistics were lower when the p-values were re-calculated with fdrtool (Table 2). The deviations from the uniform distribution possibly indicate that our data do not fit to model assumptions or problems with the raw data (Strimmer 2008a, b). Also, methods for multiple testing assume uniform distribution (Strimmer 2008a). Therefore, we further investigated the deviation from uniform distribution by filtering the potentially uninformative CpGs as has been done previously on gene expression count data (Bourgon et al. 2010). More specifically, we calculated the mean of methylated counts (i.e. Cs) for each CpG across all individuals. We applied a threshold for the rank of mean methylated C counts and filtered out those CpGs that were causing the deviation from the uniform distribution (Supplementary figures 6 and 7) by keeping most of the significant CpGs. By removing 30% of the lowest C counts we recovered p-value distribution closer to the uniform distribution and a surplus for small ($p<0.05$) p-values (Figure 2, Table 1). The filtering was carried out using R package genefilter (Gentleman et al. 2018). The filtering approach also increased the number significant CpG sites after controlling for multiple testing (fdr <0.05) in all comparisons (Table 2).

Differentially methylated regions (DMRs)

Altogether, 336 DMRs were detected in binomial regression in CONTROL-HIGH comparison and 781 DMRs in CONTROL-SMELTER comparison 30 DMRs that had exactly

the same starting position were shared between these two comparisons. 72 DMRs were detected in beta binomial regression in CONTROL-HIGH comparison and 159 DMRs in CONTROL-SMELTER comparison. Three DMRs were shared between these two comparisons. Of the DMRs identified in binomial regression, 54% and 40% of were hypomethylated in CONTROL compared to HIGH and SMELTER, respectively. In beta binomial regression 48% and 38% of the DMRs were hypomethylated in CONTROL compared to HIGH and SMELTER, respectively. This suggests that there was no clear pattern of hypo or hypermethylation in respect to pollution.

Annotation of the DMRs and CpGs

Altogether, the CONTROL-HIGH and CONTROL-SMELTER data sets were annotated to 13364 and 12972 genes, respectively. These data sets cover 72% and 70% of the total number of the annotated genes (18550) in Great tit genome. Of the 1023725 CpGs analyzed in the CONTROL-HIGH comparison, 683392 CpGs were found within genes (66.8%), 387735 (37.9%) in coding sequence and 637893 (62.3%) in exons. Of the 903449 CpGs analyzed in the CONTROL-SMELTER comparison 602997 CpGs were found within genes (66.7%), 334558 (37.0%) in coding sequence and 554421 (61.4%) in exons.

The DMRs identified in binomial regression in CONTROL-HIGH were annotated to 124 unique genes and CONTROL-SMELTER DMRs to 281 genes, excluding predicted genes. 46 of these annotated genes were shared between the comparisons. The DMRs from the beta binomial regression were annotated to 33 unique genes in CONTROL-HIGH and to 66 genes in CONTROL-SMELTER. Five of these genes were shared between the comparisons.

Gene enrichment analyses indicated 15 statistically significant (fdr < 0.05) gene ontologies in CONTROL-HIGH comparison (binomial regression) and 62 gene ontologies in CONTROL-SMELTER comparison. No statistically significant gene ontologies were found in either comparison among the DMRs identified in beta binomial regression. After merging semantically similar gene ontologies using REVIGO database, 5 and 11 enrichments remained in CONTROL-HIGH and CONTROL-SMELTER comparisons, respectively (Figure 3, Figure 4). Most of the gene ontologies were associated with developmental processes and were described under GO terms such as "system development" or "nervous system development" (Figure 3, Figure 4). Other categories involved cell-cell signaling or categories involving in transmitting information between cell and its surroundings (Figures 3, 4). Finally, we also report 10 DMRs with the largest differences in methylation levels (Supplementary Table 1). These included 12 genes (*POMC, ITGA11, LEKR1, USH2A, ZPR1, JMJD1C, ADAMTS3, PDE1C, TBP, PAPD4, GCC1* and *UTRN)* that may serve as potential candidates for further studies on the effects of pollution on organisms via DNA methylation.

Discussion

We studied whether early-life exposure to pollution affects DNA methylation patterns in a wild great tit population. We found evidence that both direct lead exposure and experimental and anthropogenic pollution during post-hatching stage affect methylation levels of a small number (0.25-2.1%) genes from which we were able collect data, yet there was no consistent hypo or hypermethylation. The number of DMRs varied between binomial and beta binomial regression to a large extent such that binomial regression was more liberal than beta binomial regression. We found that genes associated with early developmental traits were overrepresented among the DMRs in binomial regression potentially linking methylation differences to biologically meaningful traits in birds living in polluted environments.

Importantly, our results show that post-hatching environment modifies DNA methylation patterns in wild vertebrates.

Direct effects of pollution on DNA methylation

Our data on fecal metal levels presented here, as well as data on bone lead levels (lead accumulates in bone) from the same broods (Eeva et al. 2014, Ruuskanen et al. 2015) shows that the HIGH group was indeed exposed to higher levels of lead than CONTROL during the post-hatching period. The measurements correspond to observed lead levels in polluted environments across Europe (Belskii et al. 1995a, Belskii et al. 1995b, Eeva and Lehikoinen 1996, Belskii et al. 2005, Berglund and Nyholm 2011), thus validating the effectiveness and environmental relevance of the lead exposure treatment.

The HIGH-CONTROL comparison represents direct effects of lead exposure post-hatching. There were no differences in the general methylation levels (hypo or hypermethylation) between the two groups, in contrast for example to previous epidemiological studies in humans (Pilsner et al. 2009, Wright et al. 2010). However, the identified GO terms that were found to be enriched using the binomial regression analysis suggest that high lead exposure affects methylation of genes associated with biological processes such as system development and developmental processes. In previous studies, similar developmental pathways have been identified in rodents, but also sex-specific differences reported (Singh et al. 2018). These results makes sense in the light of what is known from previous studies in the same study system. For example, in the HIGH lead treatment, vitamin A, retinol and stress protein levels were higher than in the CONTROL (Eeva et al. 2014, Ruiz et al. 2016). However, we expect that the patterns that we found in blood tissue are not only restricted to blood cells, but that the majority of the findings are likely to be similar for other tissues, as found previously in the study species (Derks et al. 2016, Verhulst et al. 2016). If the observed methylation differences lead to altered gene

expression at the target genes, our results imply that the effects of pollution on such a variable set of genes may alter various developmental and cellular processes and ultimately health and phenotype.

Indirect effects of pollution on DNA methylation

At the SMELTER site, birds were exposed to various pollutants, such as copper, nickel, cadmium and arsenic, originating from the nearby copper smelter (Eeva et al. 2014), with levels higher compared to CONTROL-HIGH. However, importantly, food quality and availability likely differ between CONTROL-SMELTER, as pollution reduces some important food sources such as caterpillars, and other insects in the area (Eeva and Lehikoinen 1996, Eeva et al. 2003). Furthermore, at the polluted site, individuals may be exposed to pollutants already pre-hatching as well as during post-hatching development. We found very little overlap (~1-3%) in individual CpG sites between the CONTROL-HIGH and CONTROL-SMELTER comparisons. However, on the DMR level and their annotations showed some overlap indicating that the exposure to the lead treatment and to a larger collection of heavy metals at the smelter site can induce similar methylation changes. This suggests some direct effects of metals also at the smelter site. The gene ontology enrichments also mainly pointed that developmental processes were similar in these two comparisons suggesting that the overall effect of pollution is in the same direction. However, the majority of methylation differences in CONTROL and SMELTER are thus likely to be explained by by other elements than lead, and/or indirect effect of food, or their combination, either pre- or postnatally. Given that the metal concentrations observed at the SMELTER area are generally below the critical levels associated with sub-clinical effects (Berglund et al. 2012), indirect pollution effects via lower quality food may be likely (Eeva et al. 2005).

The number of DMRs between CONTROL and SMELTER were considerably higher than in CONTROL-HIGH comparison probably reflecting exposure to a more stressful

environment, both nutritional stress and direct exposure to pollutants of various types. If the observed methylation differences lead to altered gene expression at the target genes (see below), they could contribute to the potential developmental problems associated with poor nutrition. For example, we found that SMELTER group showed lower growth rates, higher antioxidant enzyme and stress hormone levels, lower hematocrit and survival probability than CONTROL (Eeva et al. 2014, Rainio et al. 2015b, Ruiz et al. 2016).

The signal on differential methylation for genes related to nervous system development we detected in CONTROL-SMELTER comparison could potentially point to cognitive or behavioral changes. Parallel to our results, both prenatal lead and malnutrition have recently been found to influence methylation of genes in pathways associated with neuronal proliferation and differentiation in mice and embryonic cell models (Senut et al. 2012, Senut et al. 2014, Weng et al. 2014, Sanchez-Martin et al. 2015, Singh et al. 2018, Dou et al. 2019). In humans, captive animal models and wildlife, both early nutrition and metal exposure, particularly Pb, have well-documented detrimental effects on cognitive abilities and behavior that persist into adulthood (e.g., impaired learning, memory, increased aggression, hyperactivity Brown et al. 1971, Morgan et al. 2000, Burger and Gochfeld 2005, Carere et al. 2005, Arnold et al. 2007, Chen et al. 2012, Ruuskanen et al. 2015). Until now, the role of epigenetic mechanisms underlying such effects has not been thoroughly characterized. Our results can thus stimulate further research on the potential epigenetic mechanisms explaining the long-lasting influences of early-life adverse environment on behavioral and cognitive traits.

Genes at differentially methylated regions

Among the set of 10 most differentially methylated regions across the treatment groups, we identified 12 genes (*POMC, ITGA11, LEKR1, USH2A, ZPR1, JMJD1C, ADAMTS3, PDE1C, TBP, PAPD4, GCC1* and *UTRN).* Of these 12, *POMC*, showed lower methylation (ca. 20%,

thus theoretically higher expression) in methylation *both* in SMELTER and HIGH treatments compared to CONTROL. *POMC* (pro-opiomelanocortin) is a neuronal hormone, which is cleaved to multiple key by-products, including (i) corticotropin (ACHT), controlling the stress response, (ii) appetite control and (iii) b-endorphin (Marco et al. 2016). Methylation of *POMC* has been associated with nutritional state (in rats, Ramamoorthy et al. 2018), maternal under nutrition (in ovine Stevens et al. 2010) and offspring early-life stress (in mice, Wu et al. 2014). Here, we report for the first time that *POMC* methylation may mediate early-life stress (nutritional and/or metal exposure) also in a wild vertebrate population. Furthermore, methylation of another stress related gene, PDEC1 (phosphodiesterase 1C) was also decreased in SMELTER compared to control. Expression of PDEC1 gene has been found to be associated with aldosterone stress hormone in chicken (e.g. Fallahsharoudi et al. 2017).

The other differentially methylated genes in relation to metal exposure were related to (i) DNA damage: *JMJD1C* is a candidate histone demethylase and also plays a role in the pathway DNA-damage response (e.g. Watanabe et al. 2013). Our data suggests that its methylation was decreased (theoretical expression increased) in high lead exposure compared to control, which is logical given that lead exposure is likely to cause more oxidative stress and DNA damage (e.g. Wu et al. 2008, Rainio et al. 2015a). Other studies on Pb exposure also found differences in methylation in detoxification pathways (Sen et al. 2015); (ii) growth and development: *LEKR1* (Freathy et al. 2010), *ADAMTS3* (Janssen et al. 2016), *UTNR* (in mammals, e.g. Schofield et al. 1993). Furthermore, lead has specifically been found to impair neurodevelopment (e.g. Morgan et al. 2000, Burger and Gochfeld 2005). Our data shows that methylation of *ZPR1* (zinc finger protein gene), an important protein in neural development (e.g. Doran et al. 2006) was increased by ca. 20% (theoretical expression decreased) in lead exposure compared to controls, which may warrant further studies on *ZPR1*, lead exposure and neurodevelopment. Previous studies have reported alternation in methylation of other genes related to neurodevelopment, such as another zinc finger protein gene, *Zfp974* and

Zfp787, ARTN, C5aR1 (Dou et al. 2019), *Syt2, Prkg1, Pcdhb20, Slc2a3, Klhl1 and Snap29* (Singh et al. 2018) and *PAX1* and *MSI1* (Senut et al. 2014) (iii) transcription and intracellular processes: *TBP* is universal transcription factor required for all of the eukaryotic RNA polymerases (Shimada et al. 2003), *PAPD4* is a poly(A) RNA polymerase (Burroughs et al. 2010), while *GCC1* is associated with Golgi apparatus structure (Gosavi et al. 2018). Thus, these genes may serve as potential candidates for further studies on the effects of pollution on organisms via DNA methylation.

Functional consequences of varying DNA methylation levels?

Importantly, when interpreting the potential functional consequences of the observed methylation differences, one needs to note that not all these genes with DMRs have been characterized in birds (and annotation has been done using mainly chicken and zebra finch gene models). Thus, the function of these genes is not well understood. Secondly, the link between DNA methylation and gene expression is not always straightforward (Jones 2012). However, we hypothesize that differential methylation at the observed sites affects gene activity and ultimately multiple cellular, developmental and physiological processes. Indirect evidence for a functional interpretation is provided by a recent great tit study using whole-genome bisulphite and RNAseq data. This study showed that across all genes, higher CG methylation at transcription start sites and within gene bodies was associated with lower gene expression (Laine et al. 2016). Finally, without detailed knowledge on gene function or differences in expression, it is difficult to judge whether the observed changes in methylation cause differences in phenotype and physiology we previously observed between HIGH lead exposure and CONTROL groups (Eeva et al. 2014, Rainio et al. 2015b, Ruiz et al. 2016). Therefore, follow-up studies are needed to investigate how the observed parameters are affected by differential methylation in one or more of the regions.

Methodological considerations

We employed two commonly used methods to detect CpGs and evaluated their performance using p-value histograms. Either one of these methods did not recover uniform p-value histograms in our data when applied to overall coverage threshold of 10x, pointing out that the p-values are not reliable as such. When we applied a filtering approach, developed for gene expression count data, we were able to recover uniform distribution for both methods. Thus, it appears that uninformative counts i.e. low counts for methylated state can induce a clear deviation from the uniform distribution. In other words, the small methylation differences between treatment groups are potentially difficult to model using the two statistical approaches. While we applied an overall coverage threshold of 10x to our data, it seems that another filtering step is needed for C counts to recover uniform p-value distribution at least in our data set. Overall, the performance of the binomial and beta binomial regression reflects the outcome of previous studies on simulated and empirical data sets: binomial regression has been found to be more liberal in finding CpGs as compared to beta binomial regression (Dolzhenko and Smith 2014, Park and Wu 2016, Wreczycka et al. 2017). We found considerably more CpGs and DMRs in both comparisons using binomial than beta binomial regression. Overall, the comparison between the two methods is challenging in empirical data sets but both methods seem to recover uniform p-value distribution when uninformative CpGs are filtered out.

Conclusions

In this study, we explored the environmental causes of epigenetic variation in an ecological, non-model organism, which is a novel and emerging research field. We found evidence that

differentially methylated regions contain genes enriched for biologically meaningful processes and suggest potential targets for future research. Although we used a method that does not cover the whole genome, we were able to analyze methylation patterns covering most of the annotated genes in great tit genome. Thus, bisulfite sequencing (RRBS) can be a powerful and cost-efficient method in non-model species. However, the results were not consistent between binomial and beta binomial regression, which warrants caution when selecting analysis methods and interpreting results using different methods. Finally, the functional consequences of variable methylation patterns found in this study are yet to be discovered and a more comprehensive approach combining other molecular levels as well functional studies is needed.

Acknowledgements

We thank Miia Rainio, Salla Koskinen, Tarja Pajari, Marjo Aikko, Orsolyia Palfi, Åsa Berglund and Jorma Nurmi for their efforts in helping us with field work. We also thank Finnish Centre for Scientific Computing (CSC) for providing computational resources, Center for Evolutionary Applications for molecular work and Finnish Functional Genomics Centre for sequencing services. Our study was financed by KONE foundation (SR) and Academy of Finland (TE: project 265859), Finnish Cultural Foundation (HM) and Turku University Foundation (SR and HM)

Conflict of interest

We have no conflict of interest to declare.

Data accessibility

The sequence data are deposited in the SRA database under accession number PRJNA589705.

Author contributions

SR, TE and HM designed the study. SR and TE collected the data. HM, SR and KvO designed the sequencing. HM conducted statistical and bioinformatic analyses. KvO and VL provided the genome resources. SR, HM, KvO and VL interpreted the data. All authors contributed to manuscript preparation.

Tables

Table 1. Metal concentrations (µg/g, dry weight) in feces of seven day old *Parus major* nestlings in the three treatment groups. The values are geometric means with 95% CIs. GLM and Tukey's test: means with the same letter are not significantly different. N indicates number of broods.

Metal	HIGH n = 12	SMELTER n = 11	CONTROL n = 12	$F_{ndf, ddf}$	p
Pb	8.0 (4.8-13.3) a	4.4 (2.6-7.6) ab	2.2 (1.3-3.6) b	$3.70_{2,32}$	0.03
As	0.40 (0.25-	4.3 (2.7-6.9) b	0.60 (0.38-	$24.4_{2,32}$	<0.0001
Cd	0.73 (0.51-	1.93 (1.34-2.78)	0.54 (0.38-	$10.2_{2,32}$	0.0004
Cu	34.6 (27.1-	111 (86-144) b	29.6 (23.2-	$12.9_{2,32}$	<0.0001
Ni	2.24 (1.6-3.1) a	20.6 (14.5-29.2)	1.99 (1.4-2.8) a	$34.5_{2,32}$	<0.0001

Table 2. The number and percentage (in parentheses) of p-values less than 0.05 in different comparisons. Fdr refers to multiple testing correction using Benjamini & Hochberg (1995) method. The test statistics of goodness-of-fit test (χ^2) of the p-value histograms indicated deviations uniform p-value distribution in unfiltered data, which improved with filtering ('filter').

Comparison	p-value < 0.05	fdr < 0.05	χ^2 -statistics
Binomial CONTROL HIGH	96377 (9.4)	9555	793200
Binomial CONTROL SMELTER	129830 (14.4)	25222	851790
Beta-binomial CONTROL HIGH	16852 (1.6)	7	358240
Beta-binomial CONTROL SMELTER	22669 (2.5)	33	236840
fdrtool CONTROL HIGH	62789 (6.1)	520	192880
fdrtool CONTROL SMELTER	67418 (7.5)	1811	149650
Binomial CONTROL HIGH - filter	89486 (12.5)	11334	122810
Binomial CONTROL SMELTER -filter	122463 (19.4)	30782	420650
Beta-binomial CONTROL HIGH - filter	61867 (8.6)	759	31295
Beta-binomial CONTROL SMELTER - filter	66714 (10.5)	2492	67606

Figures

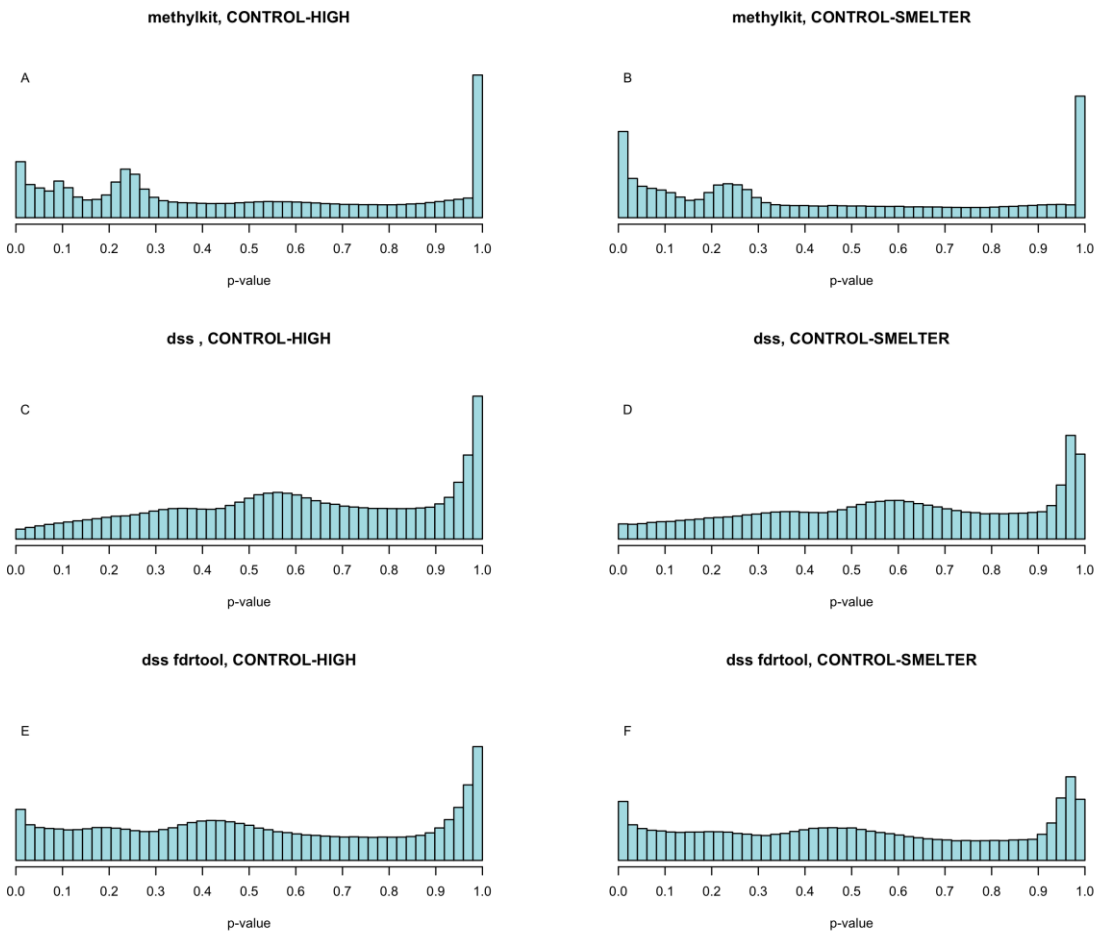


Figure 1

P-value histograms of the binomial (methylkit, A, B) and beta-binomial regression (dss, C, D) in CONTROL-HIGH and CONTROL-SMELTER comparisons. The p-values of the beta binomial regression were re-calculated based on the test statistics as implemented in the fdrtool R-package (E, F).

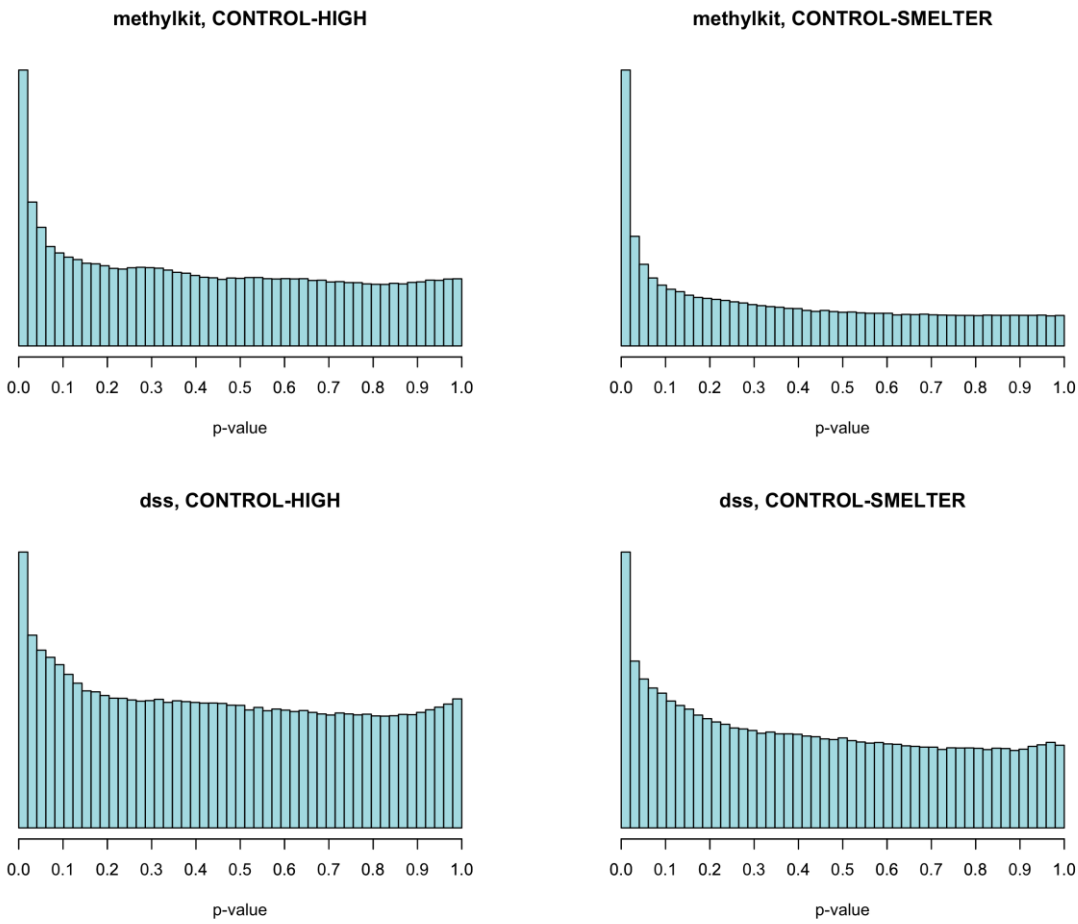


Figure 2. The p-value histograms after low coverage filtering in binomial regression (methylkit) and in beta binomial regression (dss). The p-values in beta binomial regression are based on the re-calculated p-values in fdrtool.

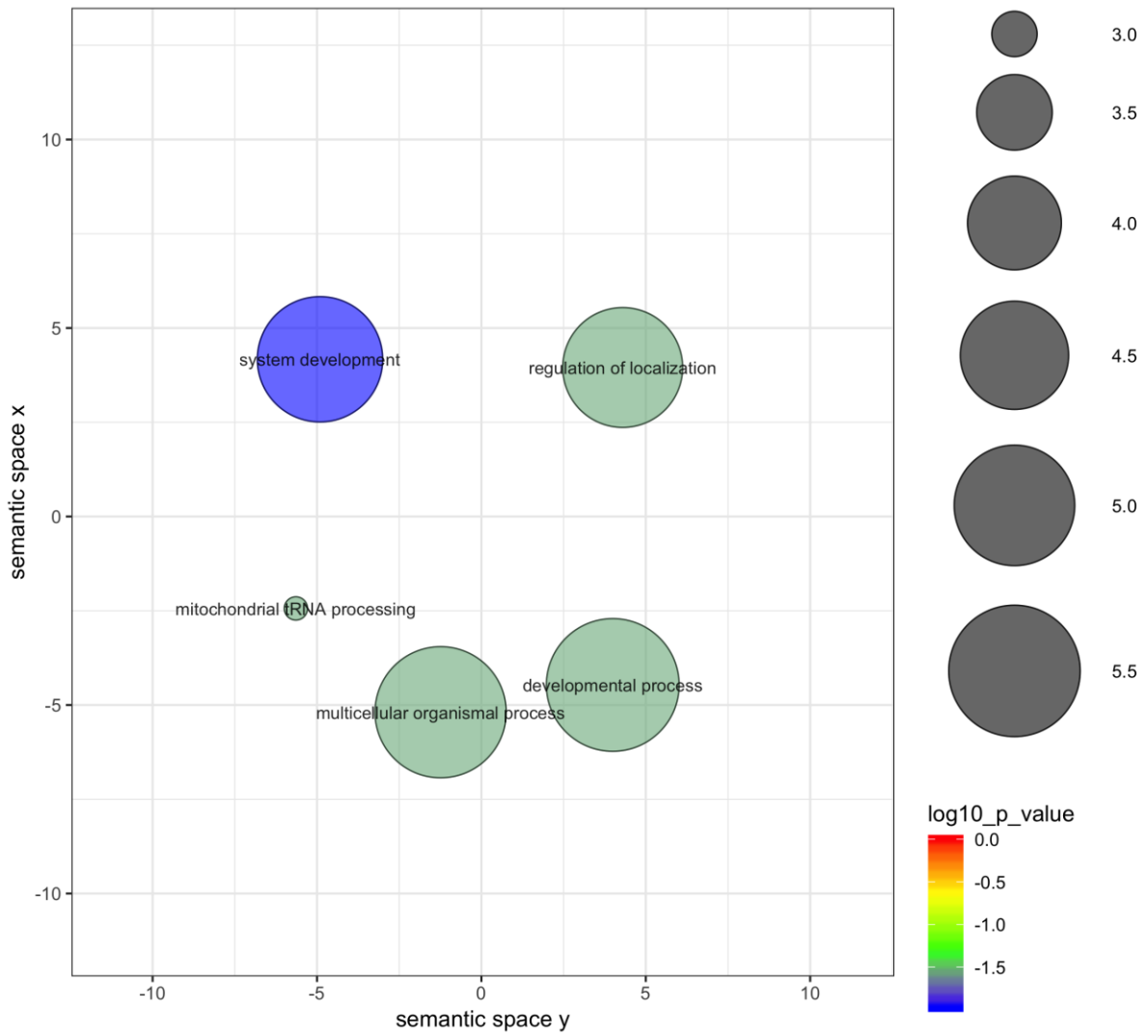


Figure 3. Results of the gene enrichment test for DMRs identified in the binomial regression in the CONTROL-HIGH comparison after merging semantically similar gene ontology categories.

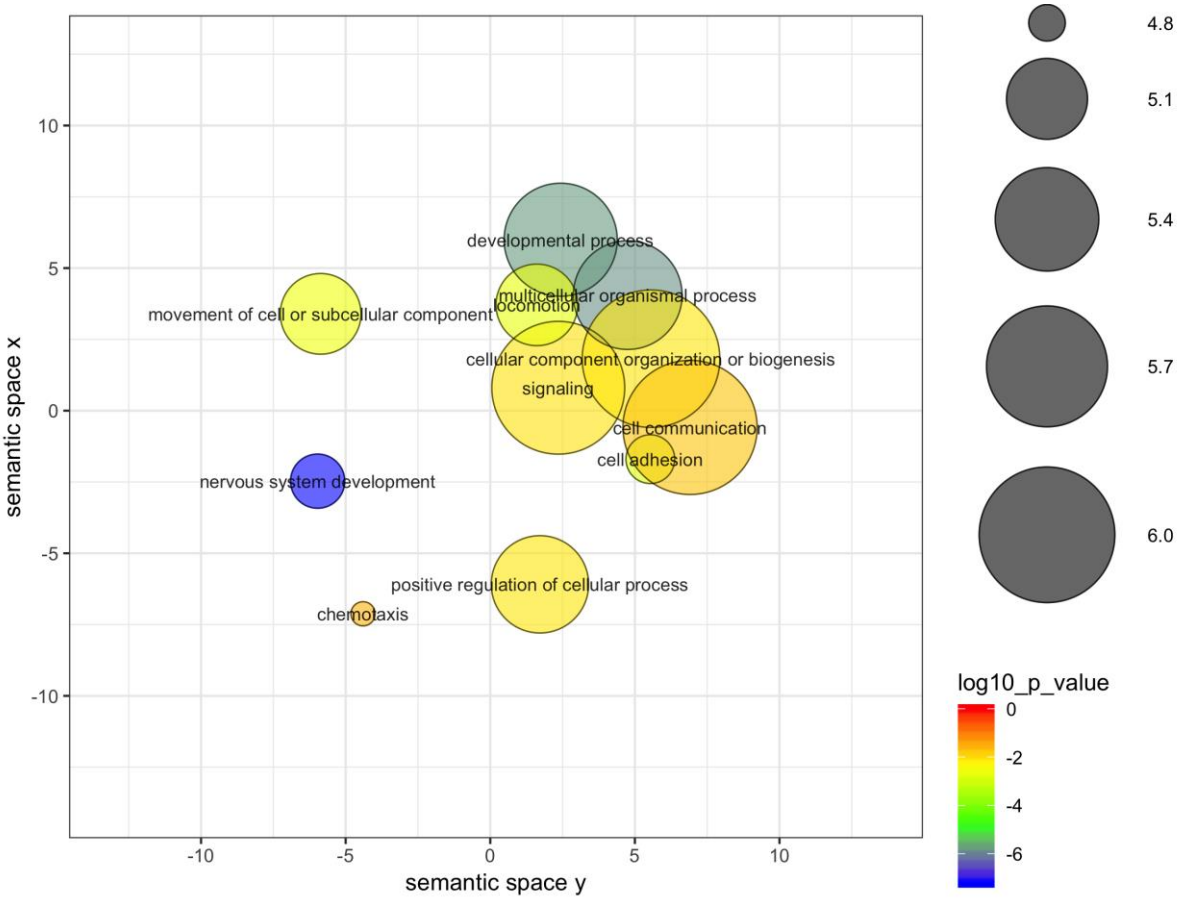
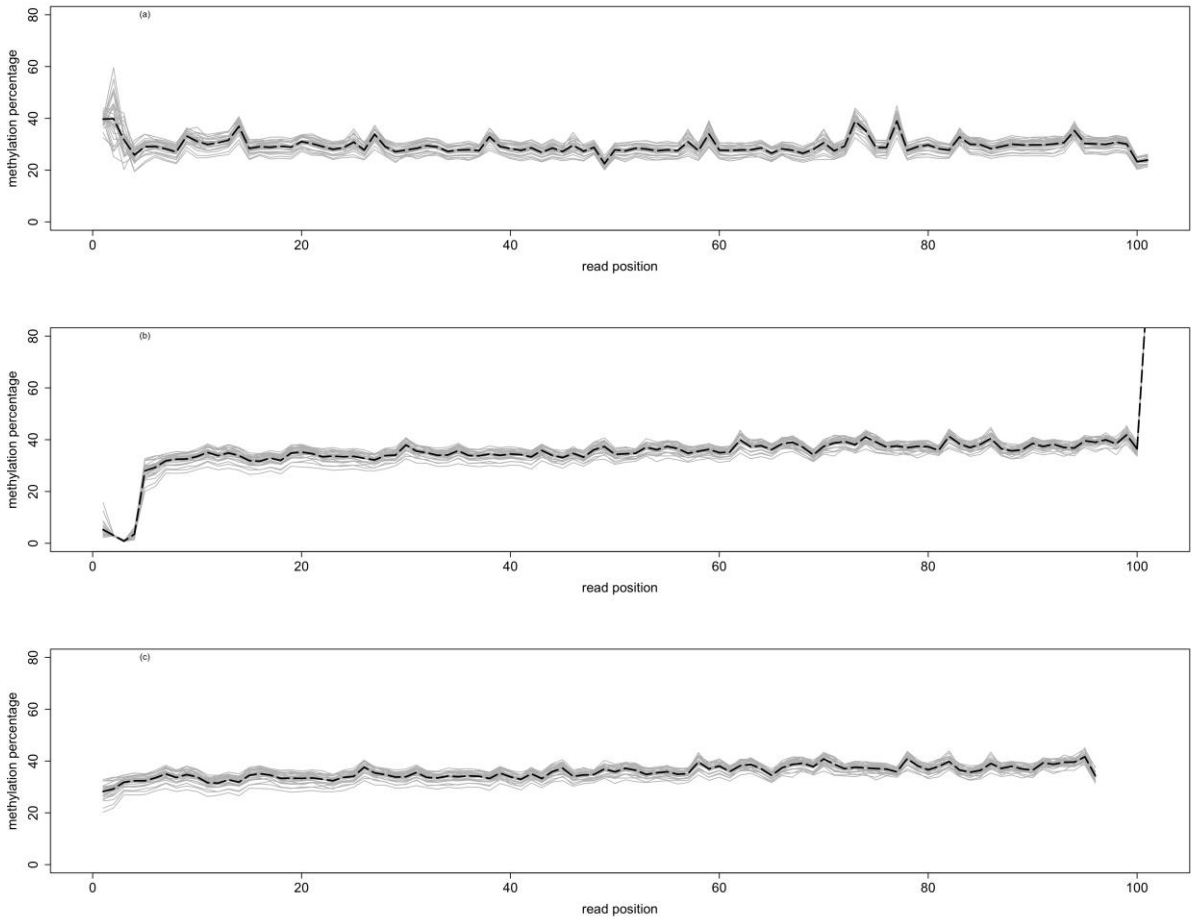
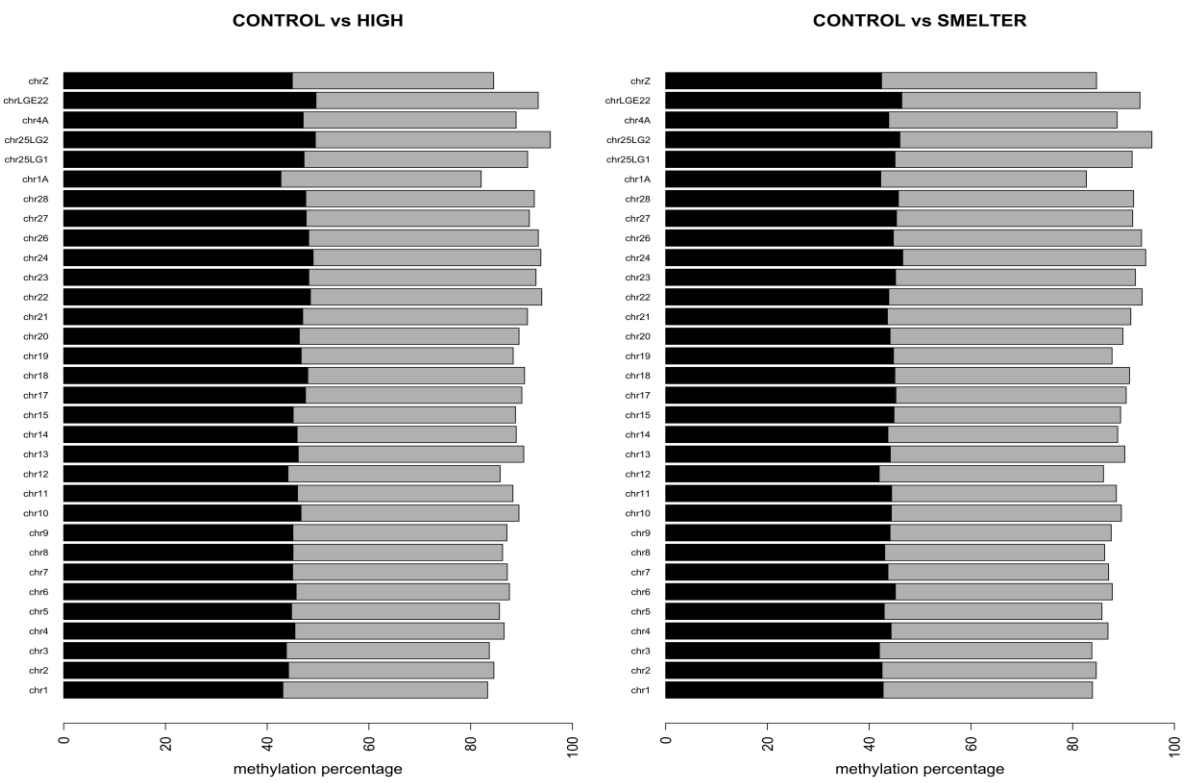


Figure 4. Results of the gene enrichment test for DMRs identified in the binomial regression in the CONTROL-SMELTER comparison after merging semantically similar gene ontology categories.



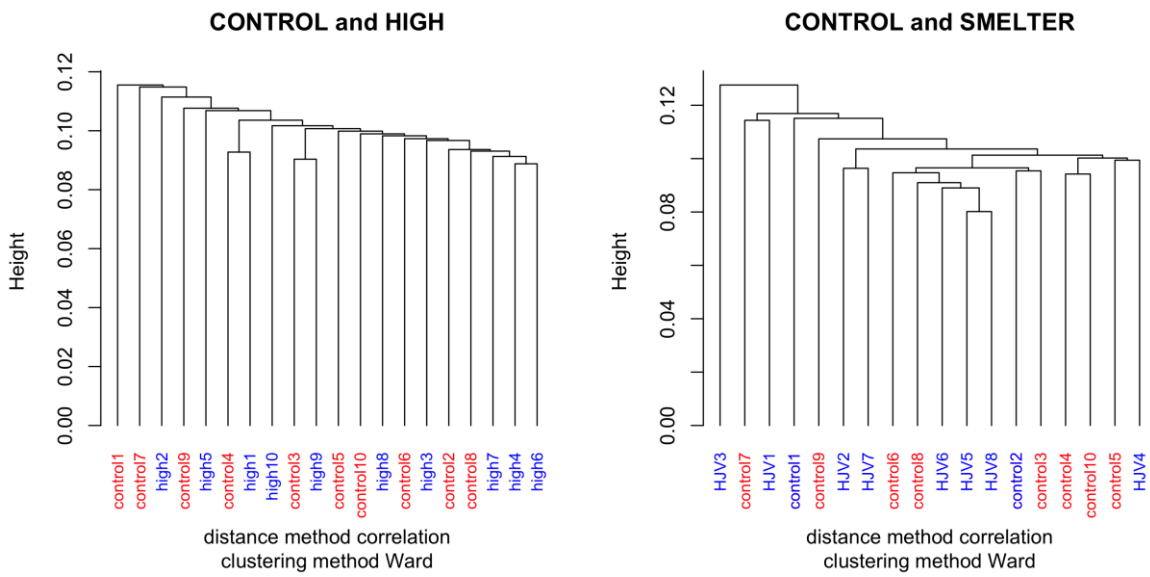
Supplementary figure 1

Methylation bias (M-bias) plots for read 1 (a) and for read 2 before (b) and after (c) cutting the bases showing lower or higher methylation than the other bases in the read. Four bases were cut from the beginning of the read 2 and one from the end of the read 2. The grey lines show the methylation percentage in all libraries along the position in the reads and the black line shows the mean methylation level across all libraries. On the x-axis is the position in the read and on the y-axis the methylation percentage.



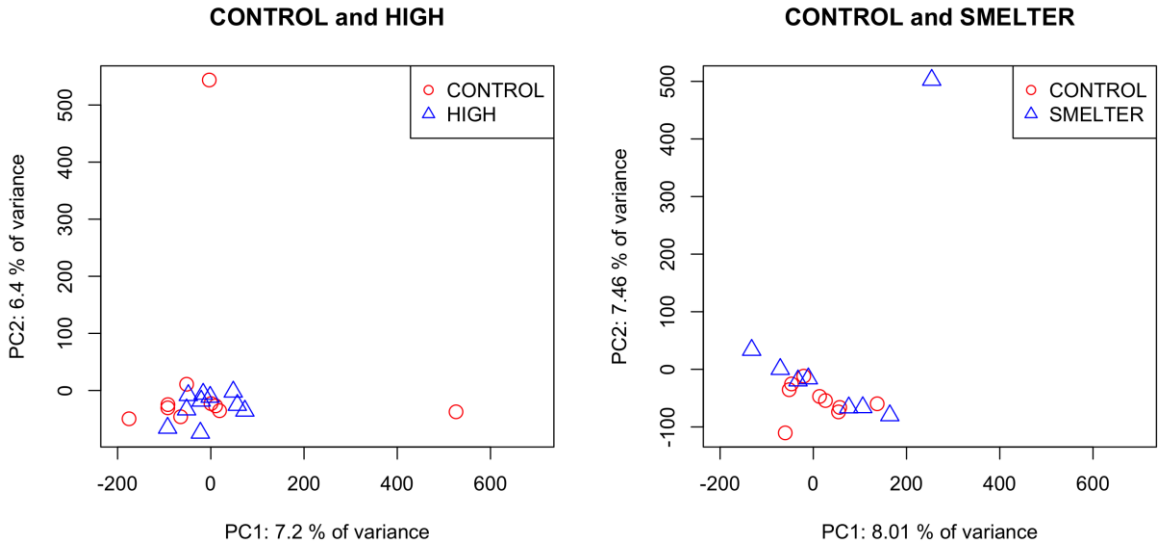
Supplementary figure 2

The percentage of hypo (black) and hypermethylated (grey) CpGs in Great tit major chromosomes. In (a) CONTROL and HIGH and in (b) CONTROL SMELTER comparisons.



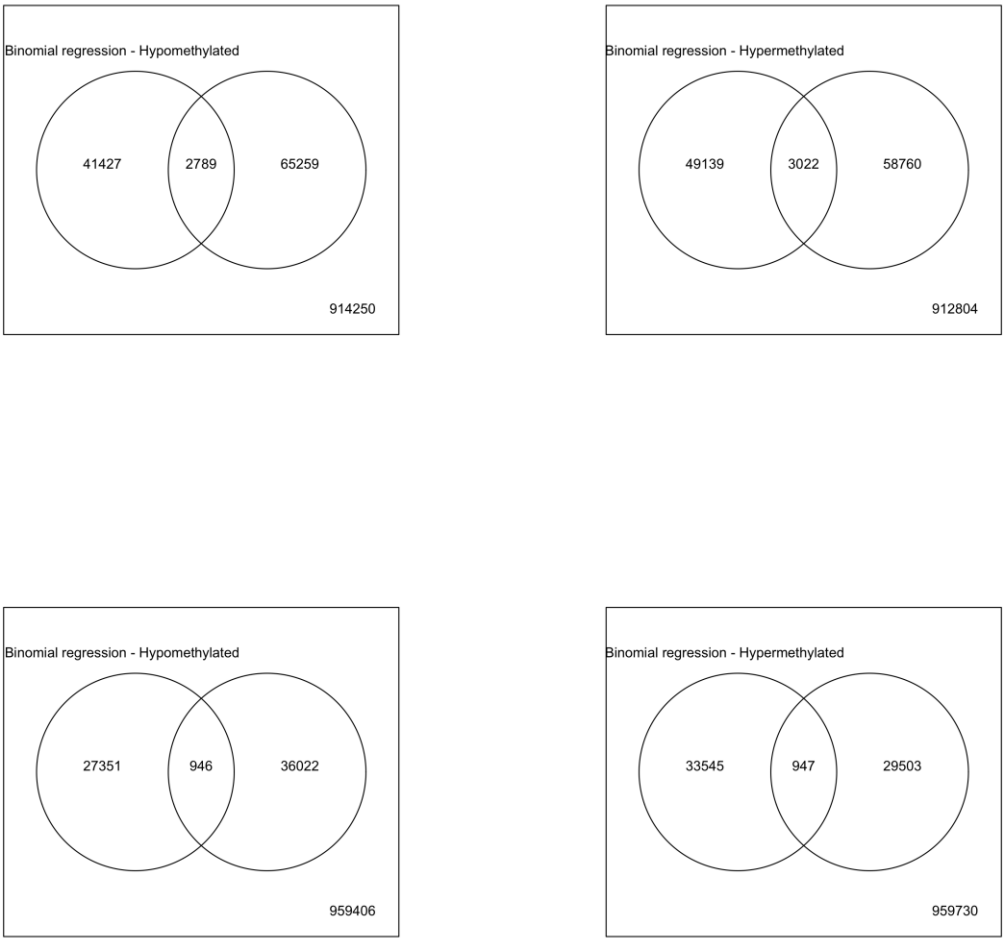
Supplementary figure 3

Hierarchical clustering of all individuals in (a) CONTROL HIGH and (b) CONTROL SMELTER comparisons.



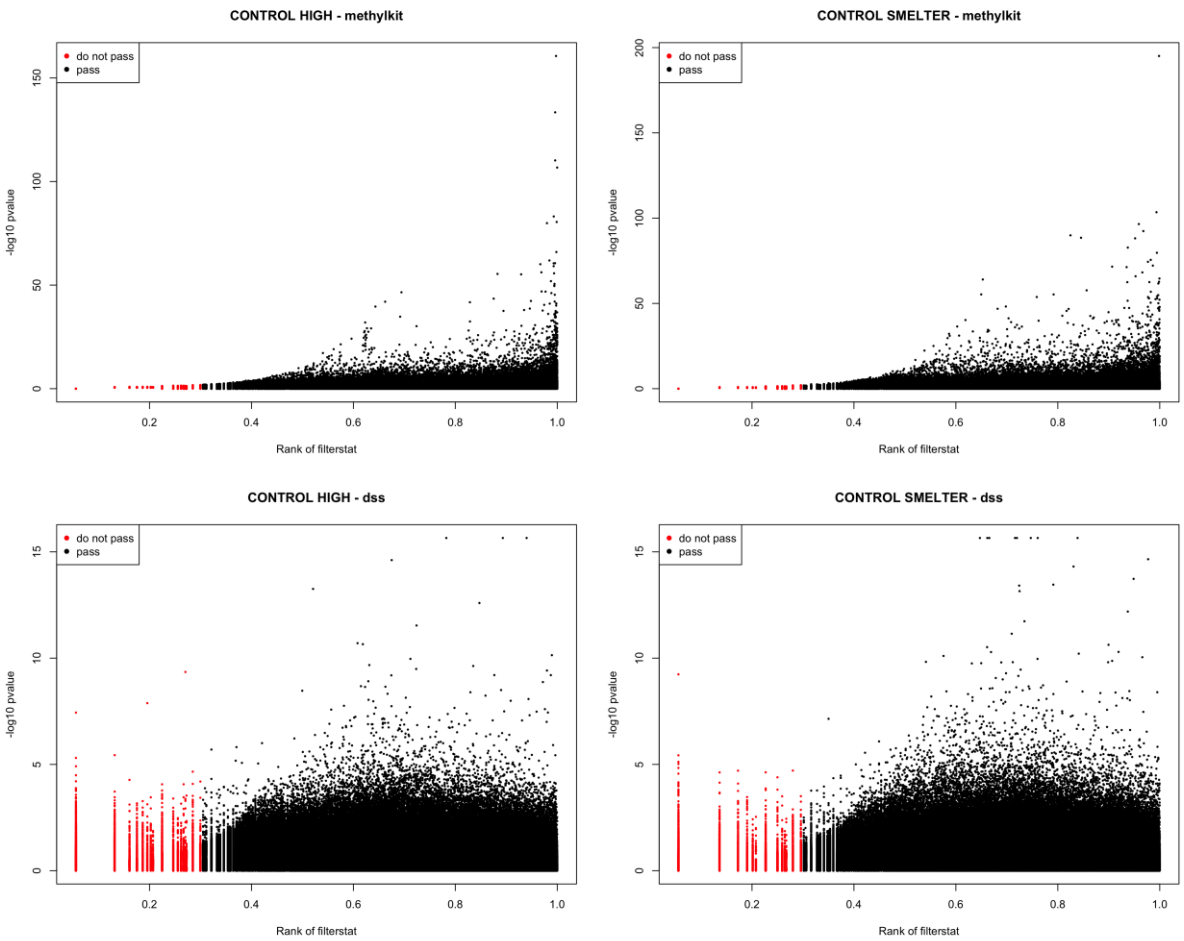
Supplementary figure 4

Principal component analysis of all individuals and CpGs in CONTROL HIGH (a) and in CONTROL SMELTER (b) comparisons. The x-axis shows the variance explained by PC1 and y-axis variance explained by PC2.

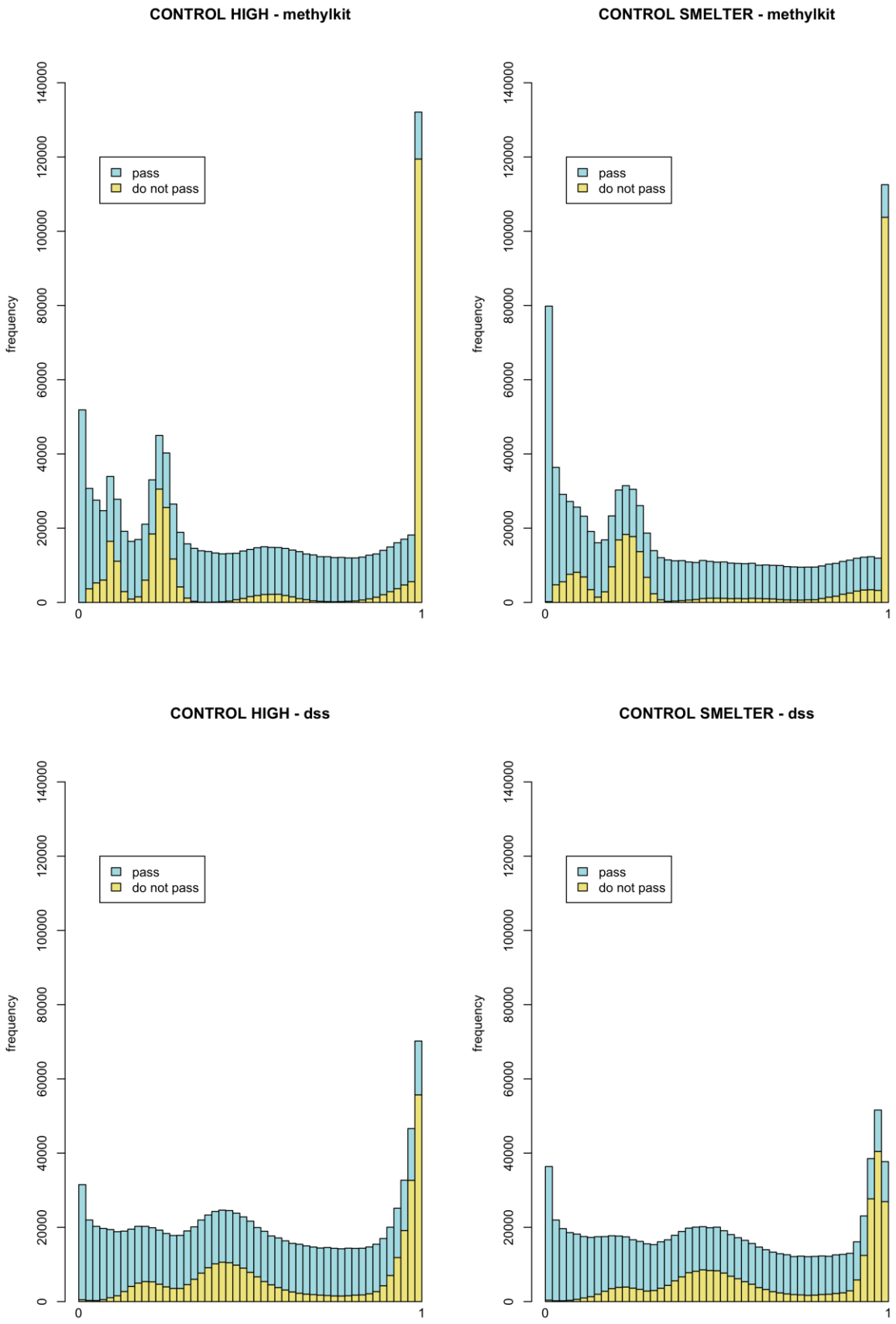


Supplementary figure 5.

Venn diagrams showing the overlap of the significant CpGs between CONTROL-HIGH and CONTROL-SMELTER comparisons. In the upper panel is the overlap in binomial regression and in the lower panel the overlap in binomial regression.



Supplementary figure 6. The principle of filtering uninformative CpGs. On the x-axis is the rank of filter statistics i.e. the mean coverage of Cs in each analyzed CpG. On the y-axis is the -log10 p-value obtained either from binomial regression (methylkit) or from beta binomial regression (dss). The red filled circles are CpGs not passing the filtering threshold (30%) while the filled black circles are CpGs passing the filtering threshold.



Supplementary figure 7.

The filtering approach shown as p-value histogram. The yellow color shows the effect of removing low coverage CpG sites along the p-value distribution. The blue colour shows the p-values remaining after the low coverage filtering.

References

Aagaard-Tillery, K. M., K. Grove, J. Bishop, X. R. Ke, Q. Fu, R. McKnight, and R. H. Lane. 2008. Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome. Journal of Molecular Endocrinology **41**:91-102.

Aljanabi, S. M., and I. Martinez. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research **25**:4692-4693.

Angers, B., E. Castonguay, and R. Massicotte. 2010. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Molecular Ecology **19**:1283-1295.

Arnold, K. E., S. L. Ramsay, C. Donaldson, and A. Adam. 2007. Parental prey selection affects risk-taking behaviour and spatial learning in avian offspring. Proceedings of the Royal Society B-Biological Sciences **274**:2563-2569.

Barton, S. J., S. R. Crozier, K. A. Lillycrop, K. M. Godfrey, and H. M. Inskip. 2013. Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. Bmc Genomics **14**.

Belskii, E. A., V. S. Bezel, and A. G. Lyakhov. 1995a. CHARACTERISTICS OF THE REPRODUCTIVE INDEXES OF BIRDS NESTING IN TREE HOLLOWS UNDER CONDITIONS OF TECHNOGENIC POLLUTION. Russian Journal of Ecology **26**:126-131.

Belskii, E. A., V. S. Bezel, and E. A. Polents. 1995b. EARLY STAGES OF THE NESTING PERIOD OF HOLLOW-NESTING BIRDS UNDER CONDITIONS OF INDUSTRIAL-POLLUTION. Russian Journal of Ecology **26**:38-43.

Belskii, E. A., N. V. Lugas'kova, and A. A. Karfidova. 2005. Reproductive parameters of adult birds and morphophysiological characteristics of chicks in the pied flycatcher (Ficedula hypoleuca Pall.) in technogenically polluted habitats. Russian Journal of Ecology **36**:329-335.

Benjamini, Y., and Y. Hochberg. 1995. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. Journal of the Royal Statistical Society Series B-Statistical Methodology **57**:289-300.

Berglund, A. M. M., and N. E. I. Nyholm. 2011. Slow improvements of metal exposure, health- and breeding conditions of pied flycatchers (Ficedula hypoleuca) after decreased industrial heavy metal emissions. Science of the Total Environment **409**:4326-4334.

Berglund, A. M. M., M. J. Rainio, and T. Eeva. 2012. Decreased metal accumulation in passerines as a result of reduced emissions. Environmental Toxicology and Chemistry **31**:1317-1323.

Bihaqi, S. W. 2019. Early life exposure to lead (Pb) and changes in DNA methylation: relevance to Alzheimer's disease. Reviews on Environmental Health **34**:187-195.

Bossdorf, O., C. L. Richards, and M. Pigliucci. 2008. Epigenetics for ecologists. Ecology Letters **11**:106-115.

Bourgon, R., R. Gentleman, and W. Huber. 2010. Independent filtering increases detection power for high-throughput experiments. Proceedings of the National Academy of Sciences of the United States of America **107**:9546-9551.

683 Bouwhuis, S., A. J. van Noordwijk, B. C. Sheldon, S. Verhulst, and M. E. Visser. 2010. Similar patterns
684 of age-specific reproduction in an island and mainland population of great tits Parus major.
685 Journal of Avian Biology **41**:615-620.
686 Boyd, R. S. 2010. Heavy Metal Pollutants and Chemical Ecology: Exploring New Frontiers. Journal of
687 Chemical Ecology **36**:46-58.
688 Boyle, P., K. Clement, H. C. Gu, Z. D. Smith, M. Ziller, J. L. Fostel, L. Holmes, J. Meldrim, F. Kelley, A.
689 Gnirke, and A. Meissner. 2012. Gel-free multiplexed reduced representation bisulfite
690 sequencing for large-scale DNA methylation profiling. Genome Biology **13**:10.
691 Brown, S., Dragann, and W. H. Vogel. 1971. EFFECTS OF LEAD ACETATE ON LEARNING AND MEMORY
692 IN RATS. Archives of Environmental Health **22**:370-&.
693 Burdge, G. C., M. A. Hanson, J. L. Slater-Jefferies, and K. A. Lillycrop. 2007. Epigenetic regulation of
694 transcription: a mechanism for inducing variations in phenotype (fetal programming) by
695 differences in nutrition during early life? British Journal of Nutrition **97**:1036-1046.
696 Burger, J., and M. Gochfeld. 2005. Effects of lead on learning in herring gulls: An avian wildlife model
697 for neurobehavioral deficits. Neurotoxicology **26**:615-624.
698 Burroughs, A. M., Y. Ando, M. J. L. de Hoon, Y. Tomaru, T. Nishibu, R. Ukekawa, T. Funakoshi, T.
699 Kurokawa, H. Suzuki, Y. Hayashizaki, and C. O. Daub. 2010. A comprehensive survey of 3 '
700 animal miRNA modification events and a possible role for 3 ' adenylation in modulating
701 miRNA targeting effectiveness. Genome Research **20**:1398-1410.
702 Carere, C., P. J. Drent, J. M. Koolhaas, and T. G. G. Groothuis. 2005. Epigenetic effects on personality
703 traits: early food provisioning and sibling competition. Behaviour **142**:1329-1355.
704 Chen, J. F., Y. H. Chen, W. Liu, C. L. Bai, X. X. Liu, K. Liu, R. Li, J. H. Zhu, and C. J. Huang. 2012.
705 Developmental lead acetate exposure induces embryonic toxicity and memory deficit in
706 adult zebrafish. Neurotoxicology and Teratology **34**:581-586.
707 Cheng, T. F., S. Choudhuri, and K. Muldoon-Jacobs. 2012. Epigenetic targets of some toxicologically
708 relevant metals: a review of the literature. Journal of Applied Toxicology **32**:643-653.
709 Choi, S. W., and S. Friso. 2010. Epigenetics: A New Bridge between Nutrition and Health. Advances in
710 Nutrition **1**:8-16.
711 Derks, M. F. L., K. M. Schachtschneider, O. Madsen, E. Schijlen, K. J. F. Verhoeven, and K. van Oers.
712 2016. Gene and transposable element methylation in great tit (Parus major) brain and blood.
713 Bmc Genomics **17**:13.
714 Dolzhenko, E., and A. D. Smith. 2014. Using beta-binomial regression for high-precision differential
715 methylation analysis in multifactor whole-genome bisulfite sequencing experiments. Bmc
716 Bioinformatics **15**.
717 Doran, B., N. Gherbesi, G. Hendricks, R. A. Flavell, R. J. Davis, and L. Gangwani. 2006. Deficiency of
718 the zinc finger protein ZPR1 causes neurodegeneration. Proceedings of the National
719 Academy of Sciences of the United States of America **103**:7471-7475.
720 Dou, J. F., Z. Farooqui, C. D. Faulk, A. K. Barks, T. Jones, D. C. Dolinoy, and K. M. Bakulski. 2019.
721 Perinatal Lead (Pb) Exposure and Cortical Neuron-Specific DNA Methylation in Male Mice.
722 Genes **10**:15.
723 Eeva, T., and E. Lehikoinen. 1996. Growth and mortality of nestling great tits (Parus major) and pied
724 flycatchers (Ficedula hypoleuca) in a heavy metal pollution gradient. Oecologia **108**:631-639.
725 Eeva, T., and E. Lehikoinen. 2004. Rich calcium availability diminishes heavy metal toxicity in Pied
726 Flycatcher. Functional Ecology **18**:548-553.
727 Eeva, T., E. Lehikoinen, and M. Nikinmaa. 2003. Pollution-induced nutritional stress in birds: An
728 experimental study of direct and indirect effects. Ecological Applications **13**:1242-1249.
729 Eeva, T., E. Lehikoinen, and T. Pohjalainen. 1997. Pollution-related variation in food supply and
730 breeding success in two hole-nesting passerines. Ecology **78**:1120-1131.
731 Eeva, T., M. Rainio, A. Berglund, M. Kanerva, J. Stauffer, M. Stoewe, and S. Ruuskanen. 2014.
732 Experimental manipulation of dietary lead levels in great tit nestlings: limited effects on
733 growth, physiology and survival. Ecotoxicology **23**:914-928.

734 Eeva, T., M. Ryoma, and J. Riihimaki. 2005. Pollution-related changes in diets of two insectivorous
735 passerines. Oecologia **145**:629-639.
736 Fallahsharoudi, A., N. de Kock, M. Johnsson, L. Bektic, S. Ubhayasekera, J. Bergquist, D. Wright, and
737 P. Jensen. 2017. Genetic and Targeted eQTL Mapping Reveals Strong Candidate Genes
738 Modulating the Stress Response During Chicken Domestication. G3-Genes Genomes
739 Genetics **7**:497-504.
740 Faulk, C., A. Barks, K. Liu, J. M. Goodrich, and D. C. Dolinoy. 2013. Early-life lead exposure results in
741 dose- and sex-specific effects on weight and epigenetic gene regulation in weanling mice.
742 Epigenomics **5**:487-500.
743 Faulk, C., and D. C. Dolinoy. 2011. Timing is everything The when and how of environmentally
744 induced changes in the epigenome of animals. Epigenetics **6**:791-797.
745 Faulk, C., K. Liu, A. Barks, J. M. Goodrich, and D. C. Dolinoy. 2014. Longitudinal epigenetic drift in
746 mice perinatally exposed to lead. Epigenetics **9**:934-941.
747 Feil, R., and M. F. Fraga. 2012. Epigenetics and the environment: emerging patterns and implications.
748 Nature Reviews Genetics **13**:97-109.
749 Fodor, A. A., T. L. Tickle, and C. Richardson. 2007. Towards the uniform distribution of null p-values
750 on Affymetrix microarrays. Genome Biology **8**.
751 Freathy, R. M., D. O. Mook-Kanamori, U. Sovio, I. Prokopenko, N. J. Timpson, D. J. Berry, N. M.
752 Warrington, E. Widen, J. J. Hottenga, M. Kaakinen, L. A. Lange, J. P. Bradfield, M. Kerkhof, J.
753 A. Marsh, R. Magi, C. M. Chen, H. N. Lyon, M. Kirin, L. S. Adair, Y. S. Aulchenko, A. J. Bennett,
754 J. B. Borja, N. Bouatia-Naji, P. Charoen, L. J. M. Coin, D. L. Cousminer, E. J. C. de Geus, P.
755 Deloukas, P. Elliott, D. M. Evans, P. Froguel, B. Glaser, C. J. Groves, A. L. Hartikainen, N.
756 Hassanali, J. N. Hirschhorn, A. Hofman, J. M. P. Holly, E. Hypponen, S. Kanoni, B. A. Knight, J.
757 Laitinen, C. M. Lindgren, W. L. McArdle, P. F. O'Reilly, C. E. Pennell, D. S. Postma, A. Pouta, A.
758 Ramasamy, N. W. Rayner, S. M. Ring, F. Rivadeneira, B. M. Shields, D. P. Strachan, I. Surakka,
759 A. Taanila, C. Tiesler, A. G. Uitterlinden, C. M. van Duijn, A. H. Wijga, G. Willemsen, H. T.
760 Zhang, J. H. Zhao, J. F. Wilson, E. A. P. Steegers, A. T. Hattersley, J. G. Eriksson, L. Peltonen, K.
761 L. Mohlke, S. F. A. Grant, H. Hakonarson, G. H. Koppelman, G. V. Dedoussis, J. Heinrich, M.
762 W. Gillman, L. J. Palmer, T. M. Frayling, D. I. Boomsma, G. D. Smith, C. Power, V. W. V.
763 Jaddoe, M. R. Jarvelin, M. I. McCarthy, G. Consortium, Magic, Wtccc, and E. G. G.
764 Consortium. 2010. Variants in ADCY5 and near CCNL1 are associated with fetal growth and
765 birth weight. Nature Genetics **42**:430-U473.
766 Garamszegi, L. Z., and P. de Villemereuil. 2017. Perturbations on the uniform distribution of p-values
767 can lead to misleading inferences from null-hypothesis testing. Trends in Neuroscience and
768 Education **8-9**:18-27.
769 Gentleman, Carey V, Huber W, and H. F. 2018. *genefilter: methods for filtering genes from high-*
770 *throughput experiments. R* package version 1.64.0.
771 Gosavi, P., F. J. Houghton, P. J. McMillan, E. Hanssen, and P. A. Gleeson. 2018. The Golgi ribbon in
772 mammalian cells negatively regulates autophagy by modulating mTOR activity. Journal of
773 Cell Science **131**:15.
774 Griffiths, R., M. C. Double, K. Orr, and R. J. G. Dawson. 1998. A DNA test to sex most birds. Molecular
775 Ecology **7**:1071-1075.
776 Hansen, K. D., B. Langmead, and R. A. Irizarry. 2012. BSmooth: from whole genome bisulfite
777 sequencing reads to differentially methylated regions. Genome Biology **13**.
778 Head, J. A. 2014. Patterns of DNA Methylation in Animals: An Ecotoxicological Perspective.
779 Integrative and Comparative Biology **54**:77-86.
780 Head, J. A., D. C. Dolinoy, and N. Basu. 2012. Epigenetics for ecotoxicologists. Environmental
781 toxicology and chemistry / SETAC **31**:221-227.
782 Heijmans, B. T., E. W. Tobi, A. D. Stein, H. Putter, G. J. Blauw, E. S. Susser, P. E. Slagboom, and L. H.
783 Lumey. 2008. Persistent epigenetic differences associated with prenatal exposure to famine
784 in humans. Proceedings of the National Academy of Sciences of the United States of America
785 **105**:17046-17049.

786 Janssen, L., L. Dupont, M. Bekhouche, A. Noel, C. Leduc, M. Voz, B. Peers, D. Cataldo, S. S. Apte, J.
787 Dubail, and A. Colige. 2016. ADAMTS3 activity is mandatory for embryonic
788 lymphangiogenesis and regulates placental angiogenesis. Angiogenesis **19**:53-65.
789 Jones, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature
790 Reviews Genetics **13**:484-492.
791 Koivula, M. J., and T. Eeva. 2010. Metal-related oxidative stress in birds. Environmental Pollution
792 **158**:2359-2370.
793 Konycheva, G., M. A. Dziadek, L. R. Ferguson, C. U. Krageloh, M. W. Coolen, M. Davison, and B. H.
794 Breier. 2011. Dietary methyl donor deficiency during pregnancy in rats shapes learning and
795 anxiety in offspring. Nutrition Research **31**:790-804.
796 Krueger, F., and S. R. Andrews. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-
797 Seq applications. Bioinformatics **27**:1571-1572.
798 Laine, V. N., T. I. Gossmann, K. M. Schachtschneider, C. J. Garroway, O. Madsen, K. J. F. Verhoeven,
799 V. de Jager, H. J. Megens, W. C. Warren, P. Minx, R. Crooijmans, P. Corcoran, B. C. Sheldon, J.
800 Slate, K. Zeng, K. van Oers, M. E. Visser, M. A. M. Groenen, and C. Great Tit HapMap. 2016.
801 Evolutionary signals of selection on cognition from the great tit genome and methylome.
802 Nature Communications **7**.
803 Liebl, A. L., A. W. Schrey, C. L. Richards, and L. B. Martin. 2013. Patterns of DNA Methylation
804 Throughout a Range Expansion of an Introduced Songbird. Integrative and Comparative
805 Biology **53**:351-358.
806 Luo, M., Y. Xu, R. Cai, Y. Q. Tang, M. M. Ge, Z. H. Liu, L. Xu, F. Hu, D. Y. Ruan, and H. L. Wang. 2014.
807 Epigenetic histone modification regulates developmental lead exposure induced
808 hyperactivity in rats. Toxicology Letters **225**:78-85.
809 Marco, A., A. Weller, and N. Meiri. 2016. Epigenetic Programming of Hypothalamic Pomc Regulates
810 Feeding and Obesity. Pages 135-163 *in* D. Spengler and E. Binder, editors. Epigenetics and
811 Neuroendocrinology: Clinical Focus on Psychiatry, Vol 1.
812 Mechta, M., L. R. Ingerslev, O. Fabre, M. Picard, and R. Barres. 2017. Evidence Suggesting Absence of
813 Mitochondrial DNA Methylation. Frontiers in Genetics **8**:9.
814 Morgan, R. E., D. A. Levitsky, and B. J. Strupp. 2000. Effects of chronic lead exposure on learning and
815 reaction time in a visual discrimination task. Neurotoxicology and Teratology **22**:337-345.
816 Nakayama, S. M. M., H. Nakata, Y. Ikenaka, J. Yabe, B. Oroszlany, Y. B. Yohannes, N. Bortey-Sam, K.
817 Muzandu, K. Choongo, T. Kuritani, M. Nakagawa, and M. Ishizuka. 2019. One year exposure
818 to Cd- and Pb-contaminated soil causes metal accumulation and alteration of global DNA
819 methylation in rats. Environmental Pollution **252**:1267-1276.
820 Park, Y., and H. Wu. 2016. Differential methylation analysis for BS-seq data under general
821 experimental design. Bioinformatics **32**:1446-1453.
822 Pilsner, J. R., H. Hu, A. Ettinger, B. N. Sanchez, R. O. Wright, D. Cantonwine, A. Lazarus, H. Lamadrid-
823 Figueroa, A. Mercado-Garcia, M. M. Tellez-Rojo, and M. Hernandez-Avila. 2009. Influence of
824 Prenatal Lead Exposure on Genomic Methylation of Cord Blood DNA. Environmental Health
825 Perspectives **117**:1466-1471.
826 Rainio, M. J., T. Eeva, T. Lilley, J. Stauffer, and S. Ruuskanen. 2015a. Effects of early-life lead exposure
827 on oxidative status and phagocytosis activity in great tits (Parus major). Comparative
828 Biochemistry and Physiology C-Toxicology & Pharmacology **167**:24-34.
829 Rainio, M. J., T. Eeva, T. Lilley, J. Stauffer, and S. Ruuskanen. 2015b. Effects of early-life lead exposure
830 on oxidative status and phagocytosis activity in great tits (Parus major). Comparative
831 biochemistry and physiology. Toxicology & pharmacology : CBP **167**:24-34.
832 Ramamoorthy, T. G., T. J. Allen, A. Davies, E. Harno, C. Sefton, C. Murgatroyd, and A. White. 2018.
833 Maternal overnutrition programs epigenetic changes in the regulatory regions of
834 hypothalamic Pomc in the offspring of rats. International Journal of Obesity **42**:1431-1444.
835 Ray, P. D., A. Yosim, and R. C. Fry. 2014. Incorporating epigenetic data into the risk assessment
836 process for the toxic metals arsenic, cadmium, chromium, lead, and mercury: strategies and
837 challenges. Frontiers in Genetics **5**:201.

838 Riyahi, S., M. Sanchez-Delgado, F. Calafell, D. Monk, and J. C. Senar. 2015. Combined epigenetic and
839 intraspecific variation of the DRD4 and SERT genes influence novelty seeking behavior in
840 great tit Parus major. Epigenetics **10**:516-525.
841 Rosenfeld, C. S. 2010. Animal Models to Study Environmental Epigenetics. Biology of Reproduction
842 **82**:473-488.
843 Rubenstein, D. R., H. Skolnik, A. Berrio, F. A. Champagne, S. Phelps, and J. Solomon. 2016. Sex-
844 specific fitness effects of unpredictable early life conditions are associated with DNA
845 methylation in the avian glucocorticoid receptor. Molecular Ecology **25**:1714-1728.
846 Ruiz, S., S. Espin, M. Rainio, S. Ruuskanen, J. P. Salminen, T. M. Lilley, and T. Eeva. 2016. Effects of
847 dietary lead exposure on vitamin levels in great tit nestlings - An experimental manipulation.
848 Environmental Pollution **213**:688-697.
849 Ruiz-Hernandez, A., C. C. Kuo, P. Rentero-Garrido, W. Y. Tang, J. Redon, J. M. Ordovas, A. Navas-
850 Acien, and M. Tellez-Plaza. 2015. Environmental chemicals and DNA methylation in adults: a
851 systematic review of the epidemiologic evidence. Clinical Epigenetics **7**:24.
852 Ruuskanen, S., T. Eeva, P. Kotitalo, J. Stauffer, and M. Rainio. 2015. No delayed behavioral and
853 phenotypic responses to experimental early-life lead exposure in great tits (Parus major).
854 Environmental Science and Pollution Research **22**:2610-2621.
855 Sanchez-Martin, F. J., D. M. Lindquist, J. Landero-Figueroa, X. Zhang, J. Chen, K. M. Cecil, M.
856 Medvedovic, and A. Puga. 2015. Sex- and tissue-specific methylome changes in brains of
857 mice perinatally exposed to lead. Neurotoxicology **46**:92-100.
858 Schofield, J., D. Houzelstein, K. Davies, M. Buckingham, and Y. H. Edwards. 1993. EXPRESSION OF THE
859 DYSTROPHIN-RELATED PROTEIN (UTROPHIN) GENE DURING MOUSE EMBRYOGENESIS.
860 Developmental Dynamics **198**:254-264.
861 Sen, A., N. Heredia, M. C. Senut, M. Hess, S. Land, W. Qu, K. Hollacher, M. O. Dereski, and D. M.
862 Ruden. 2015. Early life lead exposure causes gender-specific changes in the DNA methylation
863 profile of DNA extracted from dried blood spots. Epigenomics **7**:379-393.
864 Senut, M. C., P. Cingolani, A. Sen, A. Kruger, A. Shaik, H. Hirsch, S. T. Suhr, and D. Ruden. 2012.
865 Epigenetics of early-life lead exposure and effects on brain development. Epigenomics
866 **4**:665-674.
867 Senut, M. C., A. Sen, P. Cingolani, A. Shaik, S. J. Land, and D. M. Ruden. 2014. Lead Exposure Disrupts
868 Global DNA Methylation in Human Embryonic Stem Cells and Alters Their Neuronal
869 Differentiation. Toxicological Sciences **139**:142-161.
870 Shimada, M., T. Nakadai, and T. A. Tamura. 2003. TATA-binding protein-like protein (TLP/TRF2/TLF)
871 negatively regulates cell cycle progression and is required for the stress-mediated G(2)
872 checkpoint. Molecular and Cellular Biology **23**:4107-4120.
873 Singh, G., V. Singh, Z. X. Wang, G. Voisin, F. Lefebvre, J. M. Navenot, B. Evans, M. Verma, D. W.
874 Anderson, and J. S. Schneider. 2018. Effects of developmental lead exposure on the
875 hippocampal methylome: Influences of sex and timing and level of exposure. Toxicology
876 Letters **290**:63-72.
877 Skinner, M. K., M. Manikkam, and C. Guerrero-Bosagna. 2010. Epigenetic transgenerational actions
878 of environmental factors in disease etiology. Trends in Endocrinology and Metabolism
879 **21**:214-222.
880 Smeds, L., and A. Kunstner. 2011. CONDETRI - A Content Dependent Read Trimmer for Illumina Data.
881 Plos One **6**:6.
882 Stevens, A., G. Begum, A. Cook, K. Connor, C. Rumball, M. Oliver, J. Challis, F. Bloomfield, and A.
883 White. 2010. Epigenetic Changes in the Hypothalamic Proopiomelanocortin and
884 Glucocorticoid Receptor Genes in the Ovine Fetus after Periconceptional Undernutrition.
885 Endocrinology **151**:3652-3664.
886 Strimmer, K. 2008a. A unified approach to false discovery rate estimation. Bmc Bioinformatics **9**:14.
887 Strimmer, K. 2008b. fdrtool: a versatile R package for estimating local and tail area-based false
888 discovery rates. Bioinformatics **24**:1461-1462.

Supek, F., M. Bosnjak, N. Skunca, and T. Smuc. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. Plos One **6**.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, and C. von Mering. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research **43**:D447-D452.

Verhoeven, K. J. F., B. M. Vonholdt, and V. L. Sork. 2016. Epigenetics in ecology and evolution: what we know and what we need to know INTRODUCTION. Molecular Ecology **25**:1631-1638.

Verhulst, E. C., A. C. Mateman, M. V. Zwier, S. P. Caro, K. J. F. Verhoeven, and K. Van Oers. 2016. Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. Molecular Ecology **25**:1801-1811.

Viitaniemi, H. M., I. Verhagen, M. E. Visser, A. Honkela, K. van Oers, and A. Husby. 2019. Seasonal Variation in Genome-Wide DNA Methylation Patterns and the Onset of Seasonal Timing of Reproduction in Great Tits. Genome Biology and Evolution **11**:970-983.

Watanabe, S., K. Watanabe, V. Akimov, J. Bartkova, B. Blagoev, J. Lukas, and J. Bartek. 2013. JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. Nature Structural & Molecular Biology **20**:1425-1433.

Weaver, I. C. G., N. Cervoni, F. A. Champagne, A. C. D'Alessio, S. Sharma, Seckl, Jr., S. Dymov, M. Szyf, and M. J. Meaney. 2004. Epigenetic programming by maternal behavior. Nature Neuroscience **7**:847-854.

Weng, X. L., D. Z. Zhou, F. T. Liu, H. Zhang, J. Y. Ye, Z. Zhang, D. Zhang, Y. N. Wang, L. M. Tao, L. Cao, M. Y. Kan, T. Wang, G. Y. Feng, X. L. Qin, J. H. Sun, L. He, and Y. Liu. 2014. DNA methylation profiling in the thalamus and hippocampus of postnatal malnourished mice, including effects related to long-term potentiation. Bmc Neuroscience **15**:8.

Wenzel, M. A., and S. B. Piertney. 2014. Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (Lagopus lagopus scotica). Molecular Ecology **23**:4256-4273.

Wreczycka, K., A. Gosdschan, D. Yusuf, B. Gruning, Y. Assenov, and A. Akalin. 2017. Strategies for analyzing bisulfite sequencing data. Journal of Biotechnology **261**:105-115.

Wright, R. O., J. Schwartz, R. J. Wright, V. Bollati, L. Tarantini, S. K. Park, H. Hu, D. Sparrow, P. Vokonas, and A. Baccarelli. 2010. Biomarkers of Lead Exposure and DNA Methylation within Retrotransposons. Environmental Health Perspectives **118**:790-795.

Wu, J., M. R. Basha, B. Brock, D. P. Cox, F. Cardozo-Pelaez, C. A. McPherson, J. Harry, D. C. Rice, B. Maloney, D. Chen, D. K. Lahiri, and N. H. Zawia. 2008. Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (pb): Evidence for a developmental origin and environmental link for AD. Journal of Neuroscience **28**:3-9.

Wu, Y. H., A. V. Patchev, G. Daniel, O. F. X. Almeida, and D. Spengler. 2014. Early-Life Stress Reduces DNA Methylation of the Pomc Gene in Male Mice. Endocrinology **155**:1751-1762.

Supplementary Table 1. Top 10 DMRs for CONTROL-HIGH and CO-SMELTER comparisons, separately for binomial and betabinomial models.Mean Met% = average methylation percentage. Met% difference = difference, in percentages, in methylation between two groups. CO = control, HI = HIGH, SME = SMELTER

CHROMOSOME	POSITION (bp)		ANNOTATION	DMR lenght	No. CpGs	Mean met% CO	Mean met% HI/SME	Met% difference
A) CO-HIGH binomial								
chr10	19104012	19104107	ITGA11	96	4	80.08	49.96	30.12
chr9	24013175	24013233	LEKR1	59	7	45.12	69.51	-24.38
chr13	12002300	12002394	.	95	8	31.02	54.51	-23.49
chr28	1285543	1285681	.	139	4	60.02	83.44	-23.42
chr3	20274597	20274654	USH2A	58	6	51.39	73.61	-22.22
chr24	788116	788169	ZPR1	54	4	40.76	61.20	-20.44
chr22	703067	703161	.	95	8	37.47	18.45	19.02
chr4	1948465	1948599	ADAMTS3	135	11	33.67	52.10	-18.43
chr2	81838048	81838308	.	261	18	36.30	18.74	17.56
chrZ	65661754	65661805	.	52	4	40.04	22.79	17.25
B) CO-SMELTER betabinomial								
chrZ	19688875	19689005	.	131	14	63.56	23.61	39.95
chr2	53829014	53829116	PDE1C	103	6	94.64	61.62	33.02
Scaffold1294	1027	1153	.	127	6	42.69	67.69	-24.99
chrZ	21902415	21902582	PAPD4	168	13	8.46	32.57	-24.11
chr7	9525141	9525265	.	125	7	66.83	44.56	22.26
chr3	48718277	48718390	UTRN	114	8	74.32	52.12	22.20
chr28	1285543	1285688	.	146	5	67.49	89.66	-22.17
chr3	41373464	41375606	TBP	2143	10	38.19	16.11	22.08
chr6	1317263	1317448	LOC107207020	186	5	85.72	63.69	22.03
chr15	13816009	13816974	POMC	966	40	52.83	31.41	21.42

CHROMOSOME	POSITION (bp)		ANNOTATION	DMR lenght	No. CpGs	Mean met% CO	Mean met% HI/SME	Met% difference
C) CO-SMELTER binomial								
chr10	19104012	19104107	ITGA11	96	4	80.08	49.96	30.12
chr9	24013175	24013233	LEKR1	59	7	45.12	69.51	-24.38
chr13	12002300	12002394	.	95	8	31.02	54.51	-23.49
chr3	20274597	20274654	USH2A	58	6	51.39	73.61	-22.22
chr24	788116	788169	ZPR1	54	4	40.76	61.20	-20.44
chr15	13816420	13816728	POMC	309	25	51.26	30.95	20.31
Scaffold306	178895	179055	.	161	16	45.08	65.33	-20.25
chr6	4145270	4145351	JMJD1C	82	5	71.25	51.89	19.36
chr22	703067	703161	.	95	8	37.47	18.45	19.02
chr4	1948465	1948599	ADAMTS3	135	11	33.67	52.10	-18.43
D) CO-SMELTER betabinomial								
chrZ	19688875	19689005	.	131	14	63.56	23.61	39.95
chr2	53829014	53829116	PDE1C	103	6	94.64	61.62	33.02
chr3	41373476	41375606	TBP	2131	8	32.04	5.65	26.39
chr6	1317265	1317448	LOC107207020	184	4	83.96	59.20	24.76
chrZ	21902415	21902570	PAPD4	156	10	6.63	31.20	-24.56
chr15	13816009	13816877	POMC	869	36	51.97	29.03	22.94
Scaffold525	2758	2857	GCC1	100	28	53.35	30.49	22.86
chr3	161130	161201	.	72	10	61.32	83.76	-22.44
chr7	9525141	9525265	.	125	7	66.83	44.56	22.26
chr3	48718277	48718390	UTRN	114	8	74.32	52.12	22.20