

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Mäkinen, Hannu; van Oers, Kees; Eeva, Tapio; Ruuskanen, Suvi

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

Year: 2022

Version: Accepted version (Final draft)

Copyright: © 2021 Informa UK Limited, trading as Taylor & Francis Group

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Mäkinen, H., van Oers, K., Eeva, T., & Ruuskanen, S. (2022). The effect of experimental lead pollution on DNA methylation in a wild bird population. *Epigenetics*, 17(6), 625-641.

<https://doi.org/10.1080/15592294.2021.1943863>

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

2

3 Hannu Mäkinen¹, Kees van Oers², Tapio Eeva¹, Suvi Ruuskanen*¹

4

5 ¹Department of Biology, University of Turku, Finland

6 ²Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW),
7 Wageningen, The Netherlands

8 *corresponding author: hamakine@gmail.com

9

10 Keywords: Pb, *Parus major*, pollution, environmental epigenetics, ecological epigenetics,
11 ecotoxicology

12 Running title: Pollution and DNA methylation in the wild

13

14 Abstract

15 Anthropogenic pollution is known to negatively influence an organism's physiology,
16 behavior and fitness. Epigenetic regulation, such as DNA methylation, has been hypothesized
17 as a potential mechanism to mediate such effects, yet studies in wild species are lacking. We
18 first investigated the effects of early-life exposure to the heavy metal lead (Pb) on DNA
19 methylation levels in a wild population of great tits (*Parus major*), by experimentally
20 exposing nestlings to Pb at environmentally relevant levels. Secondly, we compared nestling
21 DNA methylation from a population exposed to long-term heavy metal pollution (close to a
22 copper smelter), where birds suffer from pollution-related decrease in food quality, and a
23 control population. For both comparisons, the analysis of about one million CpGs covering
24 most of the annotated genes revealed that pollution-related changes in DNA methylation
25 were not genome wide, but enriched for genes underlying developmental processes.
26 However, the results were not consistent when using binomial or beta binomial regression
27 highlighting the difficulty of modeling variance in CpGs. Our study indicates that post-natal
28 anthropogenic heavy metal exposure can affect methylation levels of development related
29 genes in a wild bird population.

30

31

32

33

34

35

36

37 **Introduction**

38 Epigenetic control of gene expression, such as DNA methylation, is increasingly recognized
39 as playing a major role in many different cellular processes. DNA methylation is the addition
40 of a methyl (-CH₃) group to the 5' carbon site of cytosines catalyzed by DNA-
41 methyltransferases that occurs mainly at CpG sites in animals. Especially in CpG islands
42 within promotor regions, DNA methylation is found to be negatively associated with gene
43 expression. Epigenetic changes are linked to variation in phenotype and behavior and are
44 associated with prevalence for various diseases (Angers et al. 2010, Rosenfeld 2010, Skinner
45 et al. 2010).

46 Methylation patterns can be affected by various environmental factors such as
47 maternal nutrition and maternal care (e.g. Weaver et al. 2004, Heijmans et al. 2008, Faulk
48 and Dolinoy 2011, Feil and Fraga 2012), but also by various pollutants and other early-life
49 stressors, both pre- and postnatally, as discovered in humans and mouse models (reviewed by
50 Cheng et al. 2012, Head et al. 2012, Head 2014, Ray et al. 2014, Ruiz-Hernandez et al.
51 2015). The potential effects of environmental factors on epigenetic regulation are highly
52 important for ecological and ecotoxicological fields, but research in wild vertebrate
53 populations is only emerging (Bossdorf et al. 2008, Head et al. 2012, Liebl et al. 2013,
54 Wenzel and Piertney 2014, Riyahi et al. 2015, Rubenstein et al. 2016, Verhoeven et al. 2016,
55 Sepers et al. 2019). However, epigenetics can significantly improve our understanding of the
56 mechanisms underlying natural phenotypic variation and the responses of organisms to
57 environmental change (Verhoeven et al. 2016). Furthermore, potential transgenerational
58 epigenetic effects could explain why populations are slow to recover even after pollution
59 removal (Head 2014).

60 Heavy metals, such as Pb, are global, persistent human-induced pollutants that are
61 among the potential contaminants affecting DNA methylation status (reviewed in Bihaghi
62 2019). For example in human epidemiological studies, developing fetuses show a decrease in

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

74 In addition to the direct effect of metals, large-scale metal pollution can decrease
75 resource availability and quality in wild populations (Eeva and Lehikoinen 2004, Eeva et al.
76 2005), which could subsequently also influence methylation patterns: Along with toxicants,
77 altered nutrition and diet, especially diet poor in methyl donors, are well-known epigenetic
78 modifiers in animal models (reviewed in Choi and Friso 2010, Rosenfeld 2010, Konycheva et
79 al. 2011). Also protein or lipid-altered diets can cause major changes in the epigenome
80 (Burdge et al. 2007, Aagaard-Tillery et al. 2008, Choi and Friso 2010). All in all, we expect
81 populations inhabiting polluted environments to have altered DNA methylation patterns,
82 either due to direct or indirect pollution effects.

83 Here we investigated whether experimental and anthropogenic early-life exposure to
84 the heavy metal Pb alters genome-wide DNA methylation status in a wild population of great
85 tits (*Parus major*). First, we experimentally exposed nestlings to dietary Pb at levels found
86 close to active pollution sources in Europe and compared to respective controls. The
87 exposure covered the whole postnatal pre-fledging period. Second, we compared methylation

88 patterns from nestlings in a population close to an anthropogenic pollution source, copper
89 smelter (Eeva et al. 1997) to nestlings from an unpolluted population. Around the smelter,
90 nestlings are exposed to multiple metals (in low concentrations) pre-and postnatally and
91 experience an altered nutritional quality and quantity compared to controls. In our recent
92 work using the same experimental protocol we found that Pb exposure and altered nutrition
93 during nestling development lead to changes in e.g. growth, oxidative stress markers, stress
94 protein levels and vitamin metabolism, but the mechanisms, potentially epigenetic regulation
95 are not understood (Eeva et al. 2014, Rainio et al. 2015b, Ruiz et al. 2016).

96 Our objectives are two-fold. First, by comparing nestlings exposed to the
97 experimental Pb treatment to the control group we aim to detect methylation changes directly
98 induced by Pb during post-natal stage. Second, by comparing nestling methylation patterns
99 from an environment with a long history of metal pollution, to an unpolluted population we
100 aim to detect long term effects of pollutants on great tit methylation levels, that could be
101 mediated via prenatal or postnatal metal exposure or via altered nutrition (latter coined as
102 indirect effect of metal pollution). By these two comparisons we expect to find differentially
103 methylated regions associated with genes that can potentially explain how wild bird
104 populations cope with anthropogenic pollution. Given that there are not yet established
105 golden standards for analyzing methylation data in an ecological context, we used and
106 compared two frequently used analytical tools to detect differentially methylated sites
107 (Wreczycka et al. 2017, Zhang et al. 2018).

108

109 **Methods**

110 *Study species*

111 The great tit is a small passerine bird and a model species in ecological and evolutionary
112 research, with ample ecological and genetic background information available. It is an

113 insectivorous non-migratory bird that commonly breeds in nest boxes, making it an ideal
114 species for experimental manipulations. Importantly, as one of the only non-domesticated
115 bird species, both the genome and methylome are available (Derks et al. 2016, Laine et al.
116 2016, Verhulst et al. 2016).

117 ***(1) Experimental Pb treatment and control population***

118 The Pb exposure, dosages and sampling are described in detail in (Eeva et al. 2014, Rainio et
119 al. 2015b, Ruuskanen et al. 2015). Briefly, breeding was monitored to record hatching dates
120 of great tit chicks in a population with low pollution levels in southwestern Finland (Turku,
121 60°26'N, 22°10'E). From day 3 after hatching (hatch date = 0) until day 14 (i.e. in total 12
122 days) whole broods were subjected to Pb with daily oral dosing with the following
123 treatments: HIGH dose (4 µg Pb/g body mass, N = 15 broods) or CONTROL (distilled water,
124 N = 15 broods). Based on calculations of fecal Pb levels in passerines at several polluted and
125 reference sites across Europe (Nyholm 1994, Belskii et al. 1995b Eeva and Lehikoinen 1996;
126 Eeva et al. 2005b; Dauwe et al. 2000; Dauwe et al. 2004, Eeva et al. 2009a, Berglund et al.
127 2010) the estimated Pb intake averages 2.2–8.5 µg/g body mass daily in polluted and 0.2–0.5
128 µg/g in control areas. Therefore, the dose represented environmentally relevant exposure
129 levels occurring in polluted areas in Europe. The exposure period covered most of the post-
130 hatching nestling period, i.e. most important developmental period in altricial birds.

131

132 ***(2) Polluted environment: Copper smelter population***

133

134 The Harjavalta copper smelter (61°20'N, 22°10'E) was built in 1945 and great tits have been
135 monitored since 1991 in the vicinity (<2 km) of the pollution source. Thus, this study site is
136 suitable for investigating long-term exposure of several metals (e.g., Pb, arsenic, cadmium,
137 copper, nickel) as well as lower food availability and quality (Eeva and Lehikoinen 2004).

138 The Mean fecal Pb concentrations have varied between 8-10 $\mu\text{g/g}$ (dry mass, geometric
139 mean) (Berglund et al. 2015) in nestlings in SMELTER. Assuming two year generation time
140 interval, great tits have been exposed to environmental pollution for about 12 generations
141 after the study population was established by placing nest boxes. These nests were monitored
142 in the same way as in HIGH and CONTROL and nestlings were dosed with distilled water.
143 The sample in the SMELTER consisted of 19 nests. The smelter site is ca 95 km from the
144 experimental site. The breeding habitat was similar in both CONTROL and SMELTER,
145 representing pine dominated forests with mixed spruce and birch. However, in the
146 CONTROL area some oaks were scattered in the forest. The same control group was used for
147 HIGH and SMELTER comparisons.

148

149 *Blood sampling protocol, DNA isolation and RRBS library preparation*

150

151 Blood samples were collected from 7-day old nestlings for sex-determination (following
152 Griffiths et al. 1998) and only females were chosen for CONTROL, HIGH and SMELTER
153 groups. Measurements were taken of multiple physiological indices (Eeva et al. 2014, Rainio
154 et al. 2015a). Fresh fecal samples were collected for measuring metal concentrations (see
155 below). Whole blood samples were collected directly in liquid nitrogen from nestlings at age
156 of 14 days (i.e. after 12 days of treatment) for analyses of DNA methylation status and
157 physiological indices. Samples were stored at -80°C until analysis. Ten unrelated (one
158 nestling/brood) samples from female nestlings in HIGH and CONTROL groups, and eight
159 samples from SMELTER group were selected randomly (total N = 28 samples). The
160 experiment was conducted under licenses from the Animal Experiment Committee of the
161 State Provincial Office of Southern Finland (license number ESAVI/846/ 04.10.03/2011) and
162 the Centre for Economic Development, Transport and the Environment, ELY Centre

163 Southwest Finland (license number VARELY/149/07.01/2011). The details of the DNA
164 isolation and RRBS library preparation are given in the supplementary material.

165

166 *Metal analyses*

167 For detailed analyses, see Eeva et al. (2014). Briefly, two fecal samples (one male and one
168 female) from the same brood were combined to assess brood level metal exposure (total N =
169 35 broods). The determination of metal concentrations (As, Cd, Cu, Ni, Pb) was conducted
170 with ICP-MS with detection limit of 1 ppt (ng/l) and below. The calibration of the instrument
171 was done with a commercial multi-standard from Ultra Scientific, IMS-102, ICP-MS
172 calibration standard 2 and certified reference materials were used for method validation. Data
173 was analyzed with GLMs (SAS 9.4) with Tukey post-hoc tests.

174

175 *Bisulfite sequencing analysis*

176 The initial quality check with Fastqc (Andrews 2010) indicated presence of Illumina
177 universal adapter contamination and low quality ($Q < 20$) bases in the 3' end of the raw
178 reads. The adapter sequences were removed with Cutadapt (Martin 2014) and the low quality
179 bases were filtered using Condenti (Smeds and Kunstner 2011) with default settings. The
180 quality filtered reads were then mapped against the Great tit reference genome (Assembly
181 Parus_major1.0.3; NCBI Bioproject PRJNA208335, Laine et al. 2016) using Bismark aligner
182 with default parameters (L, 0, -0.2) allowing 2–3 mismatches or a comparable number of
183 indels per 100 bp read (Krueger and Andrews 2011). Methylation information was extracted
184 from alignment files using the bismark_methylation_extractor tool (Krueger and Andrews
185 2011). The resulting methylation levels per base pair were inspected to detect potential
186 methylation bias in the beginning and in the end of read 1 and 2 (Hansen et al. 2012). There

187 was lower methylation in the beginning and higher in the end of read 2. Therefore, the first
188 four bases and the last base were removed from the read 2 for subsequent analyses
189 (supplementary figure 1). On average, we recovered 16.08 million raw reads (range 13.55-
190 21.15) from each RRBS library and after quality filtering 11.72 million reads remained
191 (range 10.12-15.06). On average 6.36 million (54%) of the quality filtered reads were
192 uniquely mapped against the Great tit reference genome. This translates to an average 322.45
193 million cytosine bases analyzed, of which 179.95 million cytosines (56%) were methylated.
194 We estimated the bisulfite conversion rate by aligning the reads against great tit
195 mitochondrial DNA, which is mostly un-methylated (Mechta et al. 2017) and calculated the
196 conversion rate as 1- methylation% in CpG context. The bisulfite conversion rate was 97.7-
197 99.1%.

198 In order to call methylated CpG sites from the Bismark methylation extractor files,
199 the function *readBismarkCoverage* in the R package MethyKit (Akalın et al. 2012) was used.
200 Using a minimum coverage threshold of 10, on average 1 309 860 (range 1 062 814-1 545
201 820) methylated CpG sites were obtained for the CONTROL-HIGH comparison and
202 1344110 (range 1062814-1774992) CpG sites for the CONTROL-SMELTER comparison.
203 The CpGs were then filtered by extreme coverage to remove e.g. potential PCR duplicates
204 using 99.9% percentile upper threshold as implemented in the function *filterByCoverage* in R
205 package MethyKit. Methylated CpG sites were also median normalized to take into account
206 differing library sizes using *normalizeCoverage* function in R package MethyKit. Finally,
207 CpG sites were united such that the data set contained only CpG sites covered by a minimum
208 of seven individuals per group. The subsequent data sets comprised 1023725 and 903449
209 CpG sites for CONTROL-HIGH and CONTROL-SMELTER comparisons, respectively.
210 879056 CpG sites were shared between these two comparisons. The range of mean and

211 median coverage was 25.94-33.31 and 17-27 in the CONTROL- HIGH comparison. The
212 respective statistics were 33.73-41.50 and 27-37 in the CONTROL- SMELTER comparison.

213 *Statistical analyses*

214 Two commonly used methods were used for the identification of differentially methylated
215 CpG sites. First, a generalized linear model was used as implemented in the R package
216 MethylKit. This method assumes that the methylated and un-methylated counts follow a
217 binomial distribution and the effect of group/treatment can be estimated with a log-likelihood
218 test (Akalin et al. 2012). Second, a generalized linear model assuming beta binomial
219 distribution was used taking into account potential overdispersion by estimating a gene-
220 specific shrinkage operator as implemented in R package dss (Feng et al. 2014). Currently,
221 potential overdispersion has not thoroughly been tested in MethylKit and thus was not
222 applied. In both methods, the model was fitted for each CpG site separately and we compared
223 the pairwise methylation differences between the CONTROL and HIGH and CONTROL and
224 SMELTER groups. For beta binomial regression the original p-values were recalculated
225 based on the test statistics as implemented in the R-package fdrtool (Strimmer 2008b). Since
226 the binomial regression method implemented in the R package MethylKit does not report test
227 statistics and the method is not recommended to be used with U-shaped p-value distributions
228 (Strimmer 2008b), the re-estimation of p-values were conducted only for the beta binomial
229 regression. The model fits were evaluated by inspecting the resulting p-value histograms.
230 Under a proper null model one would expect that the p-value histogram follows
231 approximately uniform distribution, but if there is an effect of treatment then a surplus of
232 small p-values is expected (Fodor et al. 2007, Barton et al. 2013, Garamszegi and de
233 Villemereuil 2017). Deviations from the uniform distribution may provide information about
234 the misspecification of the model or problems in the data.

235 The test statistics of goodness-of-fit test (Chi-square) of the p-value histograms and
236 visual inspection indicated deviations from the uniform distribution in both methods (Fig 1,
237 Table 1). However, the deviation in the beta binomial regression was smaller than the
238 deviation in the binomial regression and the test statistics were lower when the p-values were
239 re-calculated with *fdrtool* (Table 2). The deviations from the uniform distribution possibly
240 indicate that our data do not fit to model assumptions or problems with the raw data
241 (Strimmer 2008a, b). Also, methods for multiple testing assume uniform distribution
242 (Strimmer 2008a). Therefore, we further investigated the deviation from uniform distribution
243 by filtering the potentially uninformative CpGs. We used independent filtering approach
244 where those p-values deviating from the uniform are filtered out based on appropriate filter
245 statistics. In gene expression data the mean count for each transcript across all samples has
246 been successfully used as a filter statistics (Bourgon et al. 2010). Using similar approach, we
247 calculated the mean of methylated counts (i.e. Cs) for each CpG across all individuals. Note
248 that the count of C determines relative to the count of T (un-methylated) the methylation
249 level of a given CpG. We applied a threshold for the rank of mean methylated C counts and
250 filtered out those CpGs that were causing the deviation from the uniform distribution
251 (Supplementary figures 6 and 7) by keeping most of the significant CpGs. By removing 30%
252 of the lowest C counts we recovered p-value distribution closer to the uniform distribution
253 and a surplus for small ($p < 0.05$) p-values (Figure 2, Table 2). The filtering was carried out
254 using R package *genefilter* (Gentleman et al. 2018). The filtering approach also increased the
255 number significant CpG sites after controlling for multiple testing ($fdr < 0.05$) in all
256 comparisons (Table 2).

257 *Identification of DMRs*

258 Differentially methylated regions (DMRs) or clusters of differentially methylated CpG sites
259 were identified based on the results of both binomial and beta binomial models. The

260 following criteria were used to identify DMRs: (i) minimum size of a DMR = 50 bp, (ii)
261 minimum number of CpGs in the DMR = 3 and (iii) the percentage of CpGs with $\text{fdr} < 0.05 =$
262 50% , (v) and more liberally by including the percentage of CpGs with $\text{p-value} < 0.01$ in the
263 cluster = 50%. The identification of DMRs was conducted in the R package *dss*.

264 *Annotation of differentially methylated regions*

265 The location and association of the all CpGs with a given genomic feature was determined
266 using the Great tit genome assembly and annotation 1.1 (Laine et al. 2016). More
267 specifically, each CpG was annotated with respect to location in genes and promoter regions,
268 which were defined as 3 kb upstream from the gene start (Viitaniemi et al. 2019). We
269 excluded alternative transcripts in defining promoters. Some CpGs annotated to both genes
270 and promoters. The annotation was conducted using the *IntersectBed* option in the *BedTools*
271 package to identify the only the overlapping genomic features (Quinlan and Hall 2010).
272 Altogether, CpGs present in the CONTROL-HIGH and CONTROL-SMELTER comparisons
273 were annotated to 13364 and 12972 genes, respectively. These data sets cover 72% and 70%
274 of the total number of the annotated genes (18550) in the Great tit genome. Of all the
275 1023725 CpGs analyzed in the CONTROL-HIGH comparison, 683392 CpGs were found
276 within genes (66.8%) and 451843 (44.1%) within promoters. Of the 903449 CpGs in
277 CONTROL-SMELTER comparison 602997 (66.7%) were located within genes, and 398381
278 (44.1%) CpGs within promoters. The annotation includes also the overlapping parts of gene
279 bodies and promoters.

280 STRING database (Szklarczyk et al. 2015) was used to identify gene ontology categories
281 associated with the DMRs. Annotated genes without generic names i.e. genes with LOC
282 identifier were excluded from the analyses. A hierarchical clustering with a user-specified
283 cutoff value C (0.5) was used as implemented in REVIGO database for merging of
284 semantically similar GO categories corresponding to 1% chance of merging two randomly

285 generated categories (Supek et al. 2011). The lowest fdr corrected p-values of the initial
286 enrichment analyses were used to select a representative GO term for each merged category.

287 **Results**

288 *Metal exposure*

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

294

295 *DNA methylation*

296 *Descriptive methylation patterns among the treatment groups*

297 The average methylation percentages across all CpGs were 27.97 and 28.01 in CONTROL-
298 HIGH Pb comparison, and 26.82 and 26.67 in CONTROL-SMELTER comparison. The
299 mean difference in methylation in CONTROL-HIGH and CONTROL-SMELTER
300 comparisons were -0.0013% and 0.0015%, respectively. There were no marked differences
301 in the methylation percentage using 2% cutoff between the major chromosomes in either
302 comparison (Supplementary figure 2). Also, there were no clear patterns in sample clustering
303 in either of the comparisons based on hierarchical clustering or principal components analysis
304 (Supplementary figure 3, Supplementary figure 4).

305 *Differentially methylated CpG sites and DMRs in CONTROL HIGH comparison*

306 We identified 96 377 (9.4%) differentially methylated CpGs in the CONTROL-HIGH
307 comparison using binomial GLM (Table 2) at p-level 0.05, and after fdr correction (<0.05)
308 9555 CpGs remained. Almost equal proportion of the significant CpGs showed
309 hypomethylation (45.9%) in CONTROL and hypermethylation (54.1%) in HIGH
310 (supplementary figure 5). Beta binomial GLM identified 16852 (1.6%) differentially
311 methylated CpGs in CONTROL-HIGH comparison at p-level 0.05 (Table 2) and seven CpGs
312 after fdr correction. Altogether, 336 DMRs were detected in binomial regression and 72
313 DMRs in beta binomial regression in CONTROL-HIGH comparison.

314 *Differentially methylated CpG sites and DMRs in CONTROL SMELTER comparison*

315 We identified 129 830 (14.4%) CpGs in CONTROL-SMELTER comparison using binomial
316 GLM at p-level 0.05 and 25222 CpGs remained after fdr correction. 52.5% of the significant
317 CpGs were hypomethylated and 47.5% were hypermethylated (supplementary figure 5). Beta
318 binomial GLM identified 22669 (2.5%) CpGs in CONTROL-SMELTER comparison at p-
319 level <0.05 and 33 remained after fdr correction. Altogether, 781 DMRs were detected in
320 binomial regression and 159 DMRs in beta binomial regression in CONTROL-SMELTER
321 comparison.

322 *The overlap between differentially methylated CpG sites and DMRs between the two* 323 *comparisons*

324 2789 (2.9%) of the significant CpGs showing hypomethylation, and 3022 (3.1%) showing
325 hypermethylation were shared between CONTROL-HIGH and CONTROL-SMELTER,
326 respectively (supplementary figure 5). 946 (1.5%) of the significant CpGs showing
327 hypomethylation, and 947 (1.5%) showing hypermethylation were shared between
328 CONTROL-HIGH and CONTROL-SMELTER comparisons (Supplementary figure 5).
329 Using binomial regression, 30 DMRs that had exactly the same starting position were shared

330 between CONTROL-HIGH and CONTROL-SMELTER comparisons. Using beta binomial
331 regression, three DMRs were shared between these two comparisons. Of the DMRs identified
332 in binomial regression, 54% and 40% of were hypomethylated in CONTROL compared to
333 HIGH and SMELTER, respectively. In beta binomial regression 48% and 38% of the DMRs
334 were hypomethylated in CONTROL compared to HIGH and SMELTER, respectively. This
335 suggests that there was no clear pattern of hypo- or hypermethylation in respect to pollution.

336 *Annotation of the DMRs*

337 The mean methylation in DMRs (identified by binomial regression) was higher for gene
338 bodies than promoters in CONTROL (50.4% vs. 25.7%) in HIGH (49.7% vs. 24.9%). Similar
339 pattern was found in DMRs between CONTROL (48.4% vs. 30.1%) and SMELTER (46.7%
340 vs. 26.6%). These differences were also statistically significant (permutation test, 1000
341 replicates, $p = 0.001$). The DMRs identified in binomial regression in CONTROL-HIGH
342 were annotated to 123 unique genes and 53 promoter regions excluding predicted genes.
343 CONTROL-SMELTER DMRs were annotated to 281 genes, and 115 promoter regions. The
344 DMRs from the beta binomial regression were annotated to 33 unique genes and, 8 promoter
345 regions in. CONTROL-HIGH and to 66 genes and 34 promoter regions in CONTROL-
346 SMELTER. In CONTROL HIGH comparison, the number of DMRs showing hyper or hypo
347 methylation in gene bodies and promoters were similar. In SMELTER site, gene bodies and
348 promoters shows tendency for hypermethylation (57% of the DMRs in gene bodies and 67%
349 in promoters, respectively).

350 Gene enrichment analyses indicated 15 statistically significant ($fdr < 0.05$) gene
351 ontologies in CONTROL-HIGH comparison and 62 gene ontologies in CONTROL-
352 SMELTER comparison, when using DMRs from the binomial regression. No statistically
353 significant gene ontologies were found in either comparison among the DMRs identified in
354 beta binomial regression. After merging semantically similar gene ontologies using REVIGO

355 database, 5 and 11 enrichments remained in CONTROL-HIGH and CONTROL-SMELTER
356 comparisons, respectively (Figure 3, Figure 4). Most of the gene ontologies were associated
357 with developmental processes and were described under GO terms such as “system
358 development” or “nervous system development” (Figure 3, Figure 4). Other categories
359 involved cell-cell signaling or categories involving in transmitting information between cell
360 and its surroundings (Figures 3, 4). Finally, we also report 10 DMRs with the largest
361 differences in methylation levels (Supplementary Table 1). These included 12 genes (*POMC*,
362 *ITGA11*, *LEKR1*, *USH2A*, *ZPR1*, *JMJD1C*, *ADAMTS3*, *PDE1C*, *TBP*, *PAPD4*, *GCC1* and
363 *UTRN*) that may serve as potential candidates for further studies on the effects of pollution on
364 organisms via DNA methylation.

365

366 **Discussion**

367 We studied whether early-life exposure to pollution affects DNA methylation patterns in wild
368 great tit populations. We found evidence that both direct Pb exposure during post-hatching
369 stage and long-term anthropogenic pollution affect methylation levels of a small number
370 (0.25-2.1%) genes from which we were able collect data. The number of CpGs and DMRs
371 varied between binomial and beta binomial regression to a large extent such that binomial
372 regression was more liberal than beta binomial regression. We found that genes associated
373 with early developmental traits were enriched among the DMRs in binomial regression
374 potentially linking methylation differences to biologically meaningful traits in birds living in
375 polluted environments. However, this result was not consistent between the two statistical
376 methods highlighting the difficulty of modeling the variance in the CpGs. Nevertheless, our
377 results suggest that post-hatching, not only prenatal, environment modifies DNA methylation
378 patterns in wild vertebrates.

379 *Causal and direct effects of Pb pollution on DNA methylation: CONTROL-HIGH comparison*

380 Our data on fecal metal levels presented here, as well as data on bone Pb levels (Pb
381 accumulates in bone) from the same broods (Eeva et al. 2014, Ruuskanen et al. 2015) shows
382 that the HIGH group was indeed exposed to higher levels of Pb than CONTROL during the
383 post-hatching period. The measurements correspond to observed Pb levels in polluted
384 environments across Europe (Belskii et al. 1995a, Belskii et al. 1995b, Eeva and Lehikoinen
385 1996, Belskii et al. 2005, Berglund and Nyholm 2011), thus validating the effectiveness and
386 environmental relevance of the Pb exposure treatment.

387 The HIGH-CONTROL comparison represents direct effects of Pb exposure post-
388 hatching. There were no differences in the general methylation levels (hypo- or
389 hypermethylation) between the two groups in either at the CpG or DMRs, in contrast for
390 example to previous epidemiological studies in humans (Pilsner et al. 2009, Wright et al.
391 2010). The identified GO terms that were found to be enriched using the binomial regression
392 analysis suggest that high Pb exposure may affect methylation of genes associated with
393 biological processes such as system development and developmental processes. In previous
394 studies, similar developmental pathways have been identified in rodents, but also sex-specific
395 differences reported (Singh et al. 2018). These results makes sense in the light of what is
396 known from previous studies in the same study system. For example, in the HIGH Pb
397 treatment, vitamin A, retinol and stress protein levels were higher than in the CONTROL
398 (Eeva et al. 2014, Ruiz et al. 2016). However, we acknowledge that the patterns that we
399 found in blood tissue can be different in other tissues, but that the majority of the findings are
400 likely to be similar for other tissues, as found previously in the study species (Derks et al.
401 2016, Lindner et al. 2021 Verhulst et al. 2016, Husby 2020). Also, relative little is known
402 about the temporal stability of CpG methylation in great tits or other birds (Sepers et al.
403 2019). Viitaniemi et al. (2019) found that the majority of the CpGs showed stable

404 methylation during the breeding season of an experimental great tit population. We cannot
405 however, rule out the possibility that these changes are transient, and different methylation
406 patterns may emerge in another life stage than 14 day-old offspring. Furthermore, If the
407 observed methylation differences Pb to altered gene expression at the target genes, our results
408 imply that the effects of pollution on such a variable set of genes may alter various
409 developmental and cellular processes and ultimately health and phenotype.

410 *Effects of long-term environmental pollution on DNA methylation: CONTROL-SMELTER*
411 *comparison*

412 At the SMELTER site, a population residing in an environment with long-term pollution
413 exposure, birds were exposed to various pollutants, such as copper, nickel, cadmium and
414 arsenic, originating from the nearby copper smelter (Eeva et al. 2014), with levels higher
415 compared to CONTROL population. However, importantly, food quality and availability
416 likely differed between CONTROL and SMELTER, as pollution reduces some important
417 food sources such as caterpillars, and other insects in the area (Eeva and Lehikoinen 1996,
418 Eeva et al. 2003). *Detailed studies on invertebrate abundance at polluted and control sites*
419 *have shown that especially the amount of nutritious (e.g. rich in carotenoids) caterpillar*
420 *larvae of moths have decreased in the polluted area (Eeva and Lehikoinen 1997; Sillanpää et*
421 *al. 2009). This difference in abundance can be seen in the diet of great tit nestlings: the diet*
422 *of great tit nestlings contained ca. 20 % less (based on biomass) moths and caterpillars as*
423 *compared to the control area (Eeva et al. 2005). Decreased caterpillar availability manifest*
424 *in inferior growth and less yellow plumage of great tit nestlings (Eeva et al. 2009). The metal*
425 *concentrations observed at the SMELTER area are generally below the critical levels*
426 *associated with sub-clinical effects (Berglund et al. 2012), suggesting that indirect pollution*
427 *effects via lower quality food is more likely (Eeva et al. 2005). The CONTROL and*
428 *SMELTER population are not likely to differ genetically (given the low genetic*

429 differentiation in this species even at European scale, e.g. Lemoine et al. 2016) and the
430 habitats are relatively similar. However, at the SMELTER site, individuals are be exposed to
431 pollutants already pre-hatching (e.g. Ruuskanen et al. 2015) contrary to experimental birds,
432 but the importance of timing of the exposure is not well understood.

433 Contrary to predictions we did not find general differences in hypo/hypermethylation
434 in CpGs or DMRs. Only when the DMRs annotated to genes were inspected, a tendency for
435 hypermethylation was detected especially in the promoters of 155 genes (binomial) and 34
436 genes (beta-binomial) was detected. We detected a signal on differential methylation for
437 genes related to nervous system in CONTROL-SMELTER comparison, which could
438 potentially point to cognitive or behavioral changes. Parallel to our results, both prenatal Pb
439 and malnutrition have recently been found to influence methylation of genes in pathways
440 associated with neuronal proliferation and differentiation in mice and embryonic cell models
441 (Senut et al. 2012, Senut et al. 2014, Weng et al. 2014, Sanchez-Martin et al. 2015, Singh et
442 al. 2018, Dou et al. 2019). In humans, captive animal models and wildlife, both early
443 nutrition and metal exposure, particularly Pb, have well-documented detrimental effects on
444 cognitive abilities and behavior that persist into adulthood (e.g., impaired learning, memory,
445 increased aggression, hyperactivity Brown et al. 1971, Morgan et al. 2000, Burger and
446 Gochfeld 2005, Carere et al. 2005, Arnold et al. 2007, Chen et al. 2012, Ruuskanen et al.
447 2015). Until now, the role of epigenetic mechanisms underlying such effects has not been
448 thoroughly characterized. Our results can thus stimulate further research on the potential
449 epigenetic mechanisms explaining the long-lasting influences of early-life adverse
450 environment on behavioral and cognitive traits. If the observed methylation differences Pb to
451 altered gene expression at the target genes (see below), they could contribute to the potential
452 developmental problems associated with poor nutrition. For example, we found that
453 SMELTER group showed lower growth rates, higher antioxidant enzyme and stress hormone

454 levels, lower hematocrit and survival probability than CONTROL (Eeva et al. 2014, Rainio et
455 al. 2015b, Ruiz et al. 2016).

456

457 *Overlaps between the two comparisons*

458

459 We found very little overlap (~1-3%) in methylation of individual CpG sites between the
460 CONTROL-HIGH and CONTROL-SMELTER comparisons. However, on the DMR level
461 and their annotations showed some overlap indicating that the exposure to the Pb treatment
462 and metal-polluted at the smelter site can induce similar methylation changes. This suggests
463 some direct effects of metals also at the smelter site. The gene ontology enrichments also
464 mainly pointed that developmental processes were similar in these two comparisons
465 suggesting that the overall effect of pollution is in the same direction. However, the majority
466 of methylation differences in CONTROL and SMELTER are thus likely to be explained by
467 (i) other elements than Pb, (ii) and/or indirect effect of food, (iii) or their combination, or by
468 (iv) SMELTER site birds exposed to metals prenatally, compared to only postnatal exposure
469 in the HIGH group. Currently, we cannot distinguish between these alternatives. In general,
470 the number of DMRs between CONTROL and SMELTER were considerably higher than in
471 CONTROL-HIGH comparison probably reflecting exposure to a more stressful environment,
472 both nutritional stress and direct exposure to pollutants of various types pre and post-natally.

473

474 *Functional consequences of varying DNA methylation levels?*

475 Importantly, when interpreting the potential functional consequences of the observed
476 methylation differences, one needs to note that not all these genes with DMRs have been
477 characterized in birds (and annotation has been done using mainly chicken and zebra finch

478 gene models). Thus, the function of these genes is not well understood. Secondly, the link
479 between DNA methylation and gene expression is not always straightforward depending on
480 the genomic feature where the methylation changes occur (Jones 2012). Promoter
481 methylation has been found to be inversely correlated with gene expression (Lou et al. 2014),
482 but gene body methylation can have similar effects (Dixon et al. 2018). However, we
483 hypothesize that differential methylation at the observed sites affects gene activity and
484 ultimately multiple cellular, developmental and physiological processes. Indirect evidence for
485 a functional interpretation is provided by a recent great tit study using whole-genome
486 bisulfite and RNA-seq data. This study showed that across all genes, higher CG methylation
487 at transcription start sites and within gene bodies was associated with lower gene expression
488 (Laine et al. 2016). Finally, without detailed knowledge on gene function or differences in
489 expression, it is difficult to judge whether the observed changes in methylation cause
490 differences in phenotype and physiology we previously observed between HIGH Pb exposure
491 and CONTROL groups (Eeva et al. 2014, Rainio et al. 2015b, Ruiz et al. 2016). Therefore,
492 follow-up studies are needed to investigate how the observed parameters are affected by
493 differential methylation in one or more of the regions.

494 The performance of the binomial and beta-binomial models

495 We employed two commonly used methods to detect CpGs and evaluated their performance
496 using p-value histograms. Either one of these methods did not recover uniform p-value
497 histograms in our data when applied to overall coverage threshold of 10x, pointing out that
498 the p-values are not reliable as such. When we applied a filtering approach, developed for
499 gene expression count data, we were able to recover uniform distribution for both methods.
500 Thus, it appears that uninformative counts i.e. low counts for methylated state can induce a
501 clear deviation from the uniform distribution. In other words, the small methylation
502 differences between treatment groups are potentially difficult to model using the two

503 statistical approaches. While we applied an overall coverage threshold of 10x to our data, it
504 seems that another filtering step is needed for C counts to recover uniform p-value
505 distribution at least in our data set.

506 Overall, the performance of the binomial and beta binomial regression reflects the
507 outcome of previous studies on simulated and empirical data sets: binomial regression has
508 been found to be more liberal in finding CpGs as compared to beta binomial regression
509 (Dolzhenko and Smith 2014, Park and Wu 2016, Wreczycka et al. 2017), and we also found
510 considerably more CpGs and DMRs in both comparisons using binomial than beta binomial
511 regression. The likelihood of false negatives is higher when stringency of accounting for the
512 number of tests is lower. Also, this is more apparent in the individual CpG site analysis than
513 in the DMR analysis. Yet, in this novel and explorative study, identifying a set of potentially
514 affected CpGs and DMRs, very strict correction for the number of tests would lead to a large
515 type II error and thus failure to recognize differentially methylated sites or regions as
516 significant. Overall, the comparison between the two methods is challenging in empirical
517 data sets but both methods seem to recover uniform p-value distribution when uninformative
518 CpGs are filtered out. We view our analysis as a starting point for further functional
519 validation of our findings.

520 *Conclusions*

521 In this study, we explored the environmental causes of epigenetic variation in an ecological
522 model organism, which is a novel and emerging research field. We found evidence that
523 differentially methylated regions contain genes enriched for biologically meaningful
524 processes and suggest potential targets for future research. Although we used a method that
525 does not cover the whole genome, we were able to analyze methylation patterns covering
526 most of the annotated genes in great tit genome. However, the results were not consistent
527 between binomial and beta binomial regression, which warrants caution when selecting

528 analysis methods and interpreting results using different methods. Finally, the functional
529 consequences of variable methylation patterns found in this study are yet to be discovered
530 and a more comprehensive approach combining other molecular levels as well functional
531 studies is needed.

532

533 **Acknowledgements**

534 We thank Miia Rainio, Salla Koskinen, Tarja Pajari, Marjo Aikko, Orsolyia Palfi, Åsa
535 Berglund and Jorma Nurmi for their efforts in helping us with field work. We thank Veronika
536 Laine and two anonymous reviewers for constructive comments on earlier versions of this
537 manuscript. We also thank Finnish Centre for Scientific Computing (CSC) for providing
538 computational resources, Center for Evolutionary Applications for molecular work and
539 Finnish Functional Genomics Centre for sequencing services. Our study was financed by
540 KONE foundation (SR) and Academy of Finland (TE: project 265859), Finnish Cultural
541 Foundation (HM) and Turku University Foundation (SR and HM)

542

543 **Disclosure of interest**

544 We have no conflict of interest to declare.

545

546 **Data accessibility**

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

549

550

551 **Author contributions**

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

556 **References**

- 557 Aagaard-Tillery, K. M., K. Grove, J. Bishop, X. R. Ke, Q. Fu, R. McKnight, and R. H. Lane.
 558 2008. Developmental origins of disease and determinants of chromatin structure: maternal
 559 diet modifies the primate fetal epigenome. *Journal of Molecular Endocrinology* 41:91-102.
- 560 Akalin, A., M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, and
 561 C. E. Mason. 2012. methylKit: a comprehensive R package for the analysis of genome-wide
 562 DNA methylation profiles. *Genome Biology* 13:9.
- 563 Aljanabi, S. M., and I. Martinez. 1997. Universal and rapid salt-extraction of high quality
 564 genomic DNA for PCR-based techniques. *Nucleic Acids Research* 25:4692-4693.
- 565 Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data .
 566 Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- 567 Angers, B., E. Castonguay, and R. Massicotte. 2010. Environmentally induced phenotypes
 568 and DNA methylation: how to deal with unpredictable conditions until the next generation
 569 and after. *Molecular Ecology* 19:1283-1295.
- 570 Arnold, K. E., S. L. Ramsay, C. Donaldson, and A. Adam. 2007. Parental prey selection
 571 affects risk-taking behaviour and spatial learning in avian offspring. *Proceedings of the Royal
 572 Society B-Biological Sciences* 274:2563-2569.
- 573 Barton, S. J., S. R. Crozier, K. A. Lillycrop, K. M. Godfrey, and H. M. Inskip. 2013.
 574 Correction of unexpected distributions of P values from analysis of whole genome arrays by
 575 rectifying violation of statistical assumptions. *Bmc Genomics* 14.
- 576 Belskii, E. A., V. S. Bezel, and A. G. Lyakhov. 1995a. Characteristics of the reproductive
 577 indexes of birds nesting in tree hollows under conditions of technogenic pollution. *Russian
 578 Journal of Ecology* 26:126-131.
- 579 Belskii, E. A., V. S. Bezel, and E. A. Polents. 1995b. Early stages of the nesting period of
 580 hollow-nesting birds under conditions of industrial-pollution. *Russian Journal of Ecology*
 581 26:38-43.
- 582 Belskii, E. A., N. V. Lugas'kova, and A. A. Karfidova. 2005. Reproductive parameters of
 583 adult birds and morphophysiological characteristics of chicks in the pied flycatcher (*Ficedula
 584 hypoleuca* Pall.) in technogenically polluted habitats. *Russian Journal of Ecology* 36:329-
 585 335.
- 586 Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate - a practical and
 587 powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-
 588 Statistical Methodology* 57:289-300.

- 589 Berglund, A. M. M., and N. E. I. Nyholm. 2011. Slow improvements of metal exposure,
590 health- and breeding conditions of pied flycatchers (*Ficedula hypoleuca*) after decreased
591 industrial heavy metal emissions. *Science of the Total Environment* 409:4326-4334.
- 592 Berglund, A. M. M., M. J. Rainio, and T. Eeva. 2012. Decreased metal accumulation in
593 passerines as a result of reduced emissions. *Environmental Toxicology and Chemistry*
594 31:1317-1323.
- 595 Bihaqi, S. W. 2019. Early life exposure to lead (Pb) and changes in DNA methylation:
596 relevance to Alzheimer's disease. *Reviews on Environmental Health* 34:187-195.
- 597 Bossdorf, O., C. L. Richards, and M. Pigliucci. 2008. Epigenetics for ecologists. *Ecology*
598 *Letters* 11:106-115.
- 599 Bourgon, R., R. Gentleman, and W. Huber. 2010. Independent filtering increases detection
600 power for high-throughput experiments. *Proceedings of the National Academy of Sciences of*
601 *the United States of America* 107:9546-9551.
- 602 Boyd, R. S. 2010. Heavy Metal Pollutants and Chemical Ecology: Exploring New Frontiers.
603 *Journal of Chemical Ecology* 36:46-58.
- 604 Boyle, P., K. Clement, H. C. Gu, Z. D. Smith, M. Ziller, J. L. Fostel, L. Holmes, J. Meldrim,
605 F. Kelley, A. Gnirke, and A. Meissner. 2012. Gel-free multiplexed reduced representation
606 bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biology* 13:10.
- 607 Brown, S., Dragann, and W. H. Vogel. 1971. Effects of lead acetate on learning and memory
608 in rats. *Archives of Environmental Health* 22:370-&.
- 609 Burdge, G. C., M. A. Hanson, J. L. Slater-Jefferies, and K. A. Lillycrop. 2007. Epigenetic
610 regulation of transcription: a mechanism for inducing variations in phenotype (fetal
611 programming) by differences in nutrition during early life? *British Journal of Nutrition*
612 97:1036-1046.
- 613 Burger, J., and M. Gochfeld. 2005. Effects of lead on learning in herring gulls: An avian
614 wildlife model for neurobehavioral deficits. *Neurotoxicology* 26:615-624.
- 615 Burroughs, A. M., Y. Ando, M. J. L. de Hoon, Y. Tomaru, T. Nishibu, R. Ukekawa, T.
616 Funakoshi, T. Kurokawa, H. Suzuki, Y. Hayashizaki, and C. O. Daub. 2010. A
617 comprehensive survey of 3' animal miRNA modification events and a possible role for 3'
618 adenylation in modulating miRNA targeting effectiveness. *Genome Research* 20:1398-1410.
- 619 Carere, C., P. J. Drent, J. M. Koolhaas, and T. G. G. Groothuis. 2005. Epigenetic effects on
620 personality traits: early food provisioning and sibling competition. *Behaviour* 142:1329-1355.
- 621 Chen, J. F., Y. H. Chen, W. Liu, C. L. Bai, X. X. Liu, K. Liu, R. Li, J. H. Zhu, and C. J.
622 Huang. 2012. Developmental lead acetate exposure induces embryonic toxicity and memory
623 deficit in adult zebrafish. *Neurotoxicology and Teratology* 34:581-586.
- 624 Cheng, T. F., S. Choudhuri, and K. Muldoon-Jacobs. 2012. Epigenetic targets of some
625 toxicologically relevant metals: a review of the literature. *Journal of Applied Toxicology*
626 32:643-653.
- 627 Choi, S. W., and S. Friso. 2010. Epigenetics: A New Bridge between Nutrition and Health.
628 *Advances in Nutrition* 1:8-16.

- 629 Derks, M. F. L., K. M. Schachtschneider, O. Madsen, E. Schijlen, K. J. F. Verhoeven, and K.
630 van Oers. 2016. Gene and transposable element methylation in great tit (*Parus major*) brain
631 and blood. *Bmc Genomics* 17:13.
- 632 Dixon, G., Y. Liao, L. K. Bay, and M. V. Matz. 2018. Role of gene body methylation in
633 acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of
634 Sciences of the United States of America* 115:13342-13346.
- 635 Doran, B., N. Gherbesi, G. Hendricks, R. A. Flavell, R. J. Davis, and L. Gangwani. 2006.
636 Deficiency of the zinc finger protein ZPR1 causes neurodegeneration. *Proceedings of the
637 National Academy of Sciences of the United States of America* 103:7471-7475.
- 638 Dou, J. F., Z. Farooqui, C. D. Faulk, A. K. Barks, T. Jones, D. C. Dolinoy, and K. M.
639 Bakulski. 2019. Perinatal Lead (Pb) Exposure and Cortical Neuron-Specific DNA
640 Methylation in Male Mice. *Genes* 10:15.
- 641 Eeva, T., and E. Lehikoinen. 1996. Growth and mortality of nestling great tits (*Parus major*)
642 and pied flycatchers (*Ficedula hypoleuca*) in a heavy metal pollution gradient. *Oecologia*
643 108:631-639.
- 644 Eeva, T., and E. Lehikoinen. 2004. Rich calcium availability diminishes heavy metal toxicity
645 in Pied Flycatcher. *Functional Ecology* 18:548-553.
- 646 Eeva, T., E. Lehikoinen, and M. Nikinmaa. 2003. Pollution-induced nutritional stress in
647 birds: An experimental study of direct and indirect effects. *Ecological Applications* 13:1242-
648 1249.
- 649 Eeva, T., E. Lehikoinen, and T. Pohjalainen. 1997. Pollution-related variation in food supply
650 and breeding success in two hole-nesting passerines. *Ecology* 78:1120-1131.
- 651 Eeva, T., M. Rainio, A. Berglund, M. Kanerva, J. Stauffer, M. Stoewe, and S. Ruuskanen.
652 2014. Experimental manipulation of dietary lead levels in great tit nestlings: limited effects
653 on growth, physiology and survival. *Ecotoxicology* 23:914-928.
- 654 Eeva, T., M. Ryoma, and J. Riihimaki. 2005. Pollution-related changes in diets of two
655 insectivorous passerines. *Oecologia* 145:629-639.
- 656 Eeva, T., Sillanpää, S. and Salminen, J.-P. (2009): The effects of diet quality and quantity on
657 plumage colour and growth of Great tit nestlings: a food manipulation experiment along a
658 pollution gradient. *Journal of Avian Biology* 40: 1-9.
- 659 Fallahsharoudi, A., N. de Kock, M. Johnsson, L. Bektic, S. Ubhayasekera, J. Bergquist, D.
660 Wright, and P. Jensen. 2017. Genetic and Targeted eQTL Mapping Reveals Strong Candidate
661 Genes Modulating the Stress Response During Chicken Domestication. *G3-Genes Genomes
662 Genetics* 7:497-504.
- 663 Faulk, C., A. Barks, K. Liu, J. M. Goodrich, and D. C. Dolinoy. 2013. Early-life lead
664 exposure results in dose- and sex-specific effects on weight and epigenetic gene regulation in
665 weanling mice. *Epigenomics* 5:487-500.
- 666 Faulk, C., and D. C. Dolinoy. 2011. Timing is everything The when and how of
667 environmentally induced changes in the epigenome of animals. *Epigenetics* 6:791-797.
- 668 Faulk, C., K. Liu, A. Barks, J. M. Goodrich, and D. C. Dolinoy. 2014. Longitudinal
669 epigenetic drift in mice perinatally exposed to lead. *Epigenetics* 9:934-941.

- 670 Feil, R., and M. F. Fraga. 2012. Epigenetics and the environment: emerging patterns and
671 implications. *Nature Reviews Genetics* 13:97-109.
- 672 Feng, H., K. N. Conneely, and H. Wu. 2014. A Bayesian hierarchical model to detect
673 differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic
674 Acids Research* 42:11.
- 675 Fodor, A. A., T. L. Tickle, and C. Richardson. 2007. Towards the uniform distribution of null
676 p-values on Affymetrix microarrays. *Genome Biology* 8.
- 677 Freathy, R. M., D. O. Mook-Kanamori, U. Sovio, I. Prokopenko, N. J. Timpson, D. J. Berry,
678 N. M. Warrington, E. Widen, J. J. Hottenga, M. Kaakinen, L. A. Lange, J. P. Bradfield, M.
679 Kerkhof, J. A. Marsh, R. Magi, C. M. Chen, H. N. Lyon, M. Kirin, L. S. Adair, Y. S.
680 Aulchenko, A. J. Bennett, J. B. Borja, N. Bouatia-Naji, P. Charoen, L. J. M. Coin, D. L.
681 Cousminer, E. J. C. de Geus, P. Deloukas, P. Elliott, D. M. Evans, P. Froguel, B. Glaser, C. J.
682 Groves, A. L. Hartikainen, N. Hassanali, J. N. Hirschhorn, A. Hofman, J. M. P. Holly, E.
683 Hypponen, S. Kanoni, B. A. Knight, J. Laitinen, C. M. Lindgren, W. L. McArdle, P. F.
684 O'Reilly, C. E. Pennell, D. S. Postma, A. Pouta, A. Ramasamy, N. W. Rayner, S. M. Ring, F.
685 Rivadeneira, B. M. Shields, D. P. Strachan, I. Surakka, A. Taanila, C. Tiesler, A. G.
686 Uitterlinden, C. M. van Duijn, A. H. Wijga, G. Willemsen, H. T. Zhang, J. H. Zhao, J. F.
687 Wilson, E. A. P. Steegers, A. T. Hattersley, J. G. Eriksson, L. Peltonen, K. L. Mohlke, S. F.
688 A. Grant, H. Hakonarson, G. H. Koppelman, G. V. Dedoussis, J. Heinrich, M. W. Gillman, L.
689 J. Palmer, T. M. Frayling, D. I. Boomsma, G. D. Smith, C. Power, V. W. V. Jaddoe, M. R.
690 Jarvelin, M. I. McCarthy, G. Consortium, MAGIC, WTCCC, and E. G. G. Consortium. 2010.
691 Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight.
692 *Nature Genetics* 42:430-U473.
- 693 Garamszegi, L. Z., and P. de Villemereuil. 2017. Perturbations on the uniform distribution of
694 p-values can lead to misleading inferences from null-hypothesis testing. *Trends in
695 Neuroscience and Education* 8-9:18-27.
- 696 Gentleman, Carey V, Huber W, and H. F. 2018. *genefilter*: methods for filtering genes from
697 high-throughput experiments. R package version 1.64.0.
- 698 Gosavi, P., F. J. Houghton, P. J. McMillan, E. Hanssen, and P. A. Gleeson. 2018. The Golgi
699 ribbon in mammalian cells negatively regulates autophagy by modulating mTOR activity.
700 *Journal of Cell Science* 131:15.
- 701 Griffiths, R., M. C. Double, K. Orr, and R. J. G. Dawson. 1998. A DNA test to sex most
702 birds. *Molecular Ecology* 7:1071-1075.
- 703 Hansen, K. D., B. Langmead, and R. A. Irizarry. 2012. BSmooth: from whole genome
704 bisulfite sequencing reads to differentially methylated regions. *Genome Biology* 13.
- 705 Head, J. A. 2014. Patterns of DNA Methylation in Animals: An Ecotoxicological
706 Perspective. *Integrative and Comparative Biology* 54:77-86.
- 707 Head, J. A., D. C. Dolinoy, and N. Basu. 2012. Epigenetics for ecotoxicologists.
708 *Environmental toxicology and chemistry / SETAC* 31:221-227.
- 709 Heijmans, B. T., E. W. Tobi, A. D. Stein, H. Putter, G. J. Blauw, E. S. Susser, P. E.
710 Slagboom, and L. H. Lumey. 2008. Persistent epigenetic differences associated with prenatal
711 exposure to famine in humans. *Proceedings of the National Academy of Sciences of the
712 United States of America* 105:17046-17049.

- 713 Husby A (2020) On the use of blood samples for measuring DNA methylation in ecological
714 epigenetic studies. *Integrative and Comparative Biology*, 60, 1558–1566.
- 715 Janssen, L., L. Dupont, M. Bekhouche, A. Noel, C. Leduc, M. Voz, B. Peers, D. Cataldo, S.
716 S. Apte, J. Dubail, and A. Colige. 2016. ADAMTS3 activity is mandatory for embryonic
717 lymphangiogenesis and regulates placental angiogenesis. *Angiogenesis* 19:53-65.
- 718 Jones, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and
719 beyond. *Nature Reviews Genetics* 13:484-492.
- 720 Koivula, M. J., and T. Eeva. 2010. Metal-related oxidative stress in birds. *Environmental*
721 *Pollution* 158:2359-2370.
- 722 Konycheva, G., M. A. Dziadek, L. R. Ferguson, C. U. Krageloh, M. W. Coolen, M. Davison,
723 and B. H. Breier. 2011. Dietary methyl donor deficiency during pregnancy in rats shapes
724 learning and anxiety in offspring. *Nutrition Research* 31:790-804.
- 725 Krueger, F., and S. R. Andrews. 2011. Bismark: a flexible aligner and methylation caller for
726 Bisulfite-Seq applications. *Bioinformatics* 27:1571-1572.
- 727 Laine, V. N., T. I. Gossmann, K. M. Schachtschneider, C. J. Garroway, O. Madsen, K. J. F.
728 Verhoeven, V. de Jager, H. J. Megens, W. C. Warren, P. Minx, R. Crooijmans, P. Corcoran,
729 B. C. Sheldon, J. Slate, K. Zeng, K. van Oers, M. E. Visser, M. A. M. Groenen, and C. Great
730 Tit HapMap. 2016. Evolutionary signals of selection on cognition from the great tit genome
731 and methylome. *Nature Communications* 7.
- 732 Liebl, A. L., A. W. Schrey, C. L. Richards, and L. B. Martin. 2013. Patterns of DNA
733 Methylation Throughout a Range Expansion of an Introduced Songbird. *Integrative and*
734 *Comparative Biology* 53:351-358.
- 735 Lindner, M., Verhagen, I., Viitaniemi, H. M., Laine, V. N., Visser, M. E., Husby, A., & van
736 Oers, K. (2021). Temporal changes in DNA methylation and RNA expression in a small song
737 bird: within- and between-tissue comparisons. *BMC Genomics*, 22(1), 1–16.
- 738 Lou, S. K., H. M. Lee, H. Qin, J. W. Li, Z. B. Gao, X. Liu, L. L. Chan, V. K. L. Lam, W. Y.
739 So, Y. Wang, S. Lok, J. Wang, R. C. W. Ma, S. K. W. Tsui, J. C. N. Chan, T. F. Chan, and K.
740 Y. Yip. 2014. Whole-genome bisulfite sequencing of multiple individuals reveals
741 complementary roles of promoter and gene body methylation in transcriptional regulation.
742 *Genome Biology* 15:21.
- 743 Luo, M., Y. Xu, R. Cai, Y. Q. Tang, M. M. Ge, Z. H. Liu, L. Xu, F. Hu, D. Y. Ruan, and H.
744 L. Wang. 2014. Epigenetic histone modification regulates developmental lead exposure
745 induced hyperactivity in rats. *Toxicology Letters* 225:78-85.
- 746 Marco, A., A. Weller, and N. Meiri. 2016. Epigenetic Programming of Hypothalamic Pomc
747 Regulates Feeding and Obesity. Pages 135-163 in D. Spengler and E. Binder, editors.
748 *Epigenetics and Neuroendocrinology: Clinical Focus on Psychiatry*, Vol 1.
- 749 Mechta, M., L. R. Ingerslev, O. Fabre, M. Picard, and R. Barres. 2017. Evidence Suggesting
750 Absence of Mitochondrial DNA Methylation. *Frontiers in Genetics* 8:9.
- 751 Meissner, A., A. Gnirke, G. W. Bell, B. Ramsahoye, E. S. Lander, and R. Jaenisch. 2005.
752 Reduced representation bisulfite sequencing for comparative high-resolution DNA
753 methylation analysis. *Nucleic Acids Research* 33:5868-5877.

- 754 Morgan, R. E., D. A. Levitsky, and B. J. Strupp. 2000. Effects of chronic lead exposure on
755 learning and reaction time in a visual discrimination task. *Neurotoxicology and Teratology*
756 22:337-345.
- 757 Nakayama, S. M. M., H. Nakata, Y. Ikenaka, J. Yabe, B. Oroszlany, Y. B. Yohannes, N.
758 Bortey-Sam, K. Muzandu, K. Choongo, T. Kuritani, M. Nakagawa, and M. Ishizuka. 2019.
759 One year exposure to Cd- and Pb-contaminated soil causes metal accumulation and alteration
760 of global DNA methylation in rats. *Environmental Pollution* 252:1267-1276.
- 761 Nyholm N.E.I. 1994. Heavy metal tissue levels, impact on breeding and nestling
762 development in natural populations of pied flycatchers (*Aves*) in the pollution gradient from a
763 smelter. In: Donker M, Eijsackers H, Heimback F (eds) *Ecotoxicology of soil organisms*.
764 Lewis Publishers, Boca Raton, pp 373–382.
- 765 Pilsner, J. R., H. Hu, A. Ettinger, B. N. Sanchez, R. O. Wright, D. Cantonwine, A. Lazarus,
766 H. Lamadrid-Figueroa, A. Mercado-Garcia, M. M. Tellez-Rojo, and M. Hernandez-Avila.
767 2009. Influence of Prenatal Lead Exposure on Genomic Methylation of Cord Blood DNA.
768 *Environmental Health Perspectives* 117:1466-1471.
- 769 Rainio, M. J., T. Eeva, T. Lilley, J. Stauffer, and S. Ruuskanen. 2015a. Effects of early-life
770 lead exposure on oxidative status and phagocytosis activity in great tits (*Parus major*).
771 *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* 167:24-34.
- 772 Rainio, M. J., T. Eeva, T. Lilley, J. Stauffer, and S. Ruuskanen. 2015b. Effects of early-life
773 lead exposure on oxidative status and phagocytosis activity in great tits (*Parus major*).
774 *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP* 167:24-34.
- 775 Ramamoorthy, T. G., T. J. Allen, A. Davies, E. Harno, C. Sefton, C. Murgatroyd, and A.
776 White. 2018. Maternal overnutrition programs epigenetic changes in the regulatory regions of
777 hypothalamic *Pomc* in the offspring of rats. *International Journal of Obesity* 42:1431-1444.
- 778 Ray, P. D., A. Yosim, and R. C. Fry. 2014. Incorporating epigenetic data into the risk
779 assessment process for the toxic metals arsenic, cadmium, chromium, lead, and mercury:
780 strategies and challenges. *Frontiers in Genetics* 5:201.
- 781 Riyahi, S., M. Sanchez-Delgado, F. Calafell, D. Monk, and J. C. Senar. 2015. Combined
782 epigenetic and intraspecific variation of the *DRD4* and *SERT* genes influence novelty
783 seeking behavior in great tit *Parus major*. *Epigenetics* 10:516-525.
- 784 Rosenfeld, C. S. 2010. Animal Models to Study Environmental Epigenetics. *Biology of*
785 *Reproduction* 82:473-488.
- 786 Rubenstein, D. R., H. Skolnik, A. Berrio, F. A. Champagne, S. Phelps, and J. Solomon. 2016.
787 Sex-specific fitness effects of unpredictable early life conditions are associated with DNA
788 methylation in the avian glucocorticoid receptor. *Molecular Ecology* 25:1714-1728.
- 789 Ruiz, S., S. Espin, M. Rainio, S. Ruuskanen, J. P. Salminen, T. M. Lilley, and T. Eeva. 2016.
790 Effects of dietary lead exposure on vitamin levels in great tit nestlings - An experimental
791 manipulation. *Environmental Pollution* 213:688-697.
- 792 Ruiz-Hernandez, A., C. C. Kuo, P. Rentero-Garrido, W. Y. Tang, J. Redon, J. M. Ordovas,
793 A. Navas-Acien, and M. Tellez-Plaza. 2015. Environmental chemicals and DNA methylation
794 in adults: a systematic review of the epidemiologic evidence. *Clinical Epigenetics* 7:24.

- 795 Ruuskanen, S., T. Eeva, P. Kotitalo, J. Stauffer, and M. Rainio. 2015. No delayed behavioral
796 and phenotypic responses to experimental early-life lead exposure in great tits (*Parus major*).
797 *Environmental Science and Pollution Research* 22:2610-2621.
- 798 Sanchez-Martin, F. J., D. M. Lindquist, J. Landero-Figueroa, X. Zhang, J. Chen, K. M. Cecil,
799 M. Medvedovic, and A. Puga. 2015. Sex- and tissue-specific methylome changes in brains of
800 mice perinatally exposed to lead. *Neurotoxicology* 46:92-100.
- 801 Schofield, J., D. Houzelstein, K. Davies, M. Buckingham, and Y. H. Edwards. 1993.
802 Expression of the dystrophin-related protein (utrophin) gene during mouse embryogenesis.
803 *Developmental Dynamics* 198:254-264.
- 804 Sen, A., N. Heredia, M. C. Senut, M. Hess, S. Land, W. Qu, K. Hollacher, M. O. Dereski,
805 and D. M. Ruden. 2015. Early life lead exposure causes gender-specific changes in the DNA
806 methylation profile of DNA extracted from dried blood spots. *Epigenomics* 7:379-393.
- 807 Senut, M. C., P. Cingolani, A. Sen, A. Kruger, A. Shaik, H. Hirsch, S. T. Suhr, and D. Ruden.
808 2012. Epigenetics of early-life lead exposure and effects on brain development. *Epigenomics*
809 4:665-674.
- 810 Senut, M. C., A. Sen, P. Cingolani, A. Shaik, S. J. Land, and D. M. Ruden. 2014. Lead
811 Exposure Disrupts Global DNA Methylation in Human Embryonic Stem Cells and Alters
812 Their Neuronal Differentiation. *Toxicological Sciences* 139:142-161.
- 813 Sepers, B., K. van den Heuvel, M. Lindner, H. Viitaniemi, A. Husby, and K. van Oers. 2019.
814 Avian ecological epigenetics: pitfalls and promises. *Journal of Ornithology* 160:1183-1203.
- 815 Shimada, M., T. Nakadai, and T. A. Tamura. 2003. TATA-binding protein-like protein
816 (TLP/TRF2/TLF) negatively regulates cell cycle progression and is required for the stress-
817 mediated G(2) checkpoint. *Molecular and Cellular Biology* 23:4107-4120.
- 818 Singh, G., V. Singh, Z. X. Wang, G. Voisin, F. Lefebvre, J. M. Navenot, B. Evans, M.
819 Verma, D. W. Anderson, and J. S. Schneider. 2018. Effects of developmental lead exposure
820 on the hippocampal methylome: Influences of sex and timing and level of exposure.
821 *Toxicology Letters* 290:63-72.
- 822 Sillanpää, S. Salminen, J.-P. and Eeva, T. (2009): Breeding success and lutein availability in
823 Great tit (*Parus major*). *Acta Oecologica* 35: 805-810.
- 824 Skinner, M. K., M. Manikkam, and C. Guerrero-Bosagna. 2010. Epigenetic transgenerational
825 actions of environmental factors in disease etiology. *Trends in Endocrinology and*
826 *Metabolism* 21:214-222.
- 827 Smeds, L., and A. Kunstner. 2011. CONDETRI - A Content Dependent Read Trimmer for
828 Illumina Data. *Plos One* 6:6.
- 829 Stevens, A., G. Begum, A. Cook, K. Connor, C. Rumball, M. Oliver, J. Challis, F.
830 Bloomfield, and A. White. 2010. Epigenetic Changes in the Hypothalamic
831 Proopiomelanocortin and Glucocorticoid Receptor Genes in the Ovine Fetus after
832 Periconceptual Undernutrition. *Endocrinology* 151:3652-3664.
- 833 Strimmer, K. 2008a. A unified approach to false discovery rate estimation. *Bmc*
834 *Bioinformatics* 9:14.

- 835 Strimmer, K. 2008b. fdrtool: a versatile R package for estimating local and tail area-based
836 false discovery rates. *Bioinformatics* 24:1461-1462.
- 837 Supek, F., M. Bosnjak, N. Skunca, and T. Smuc. 2011. REVIGO Summarizes and Visualizes
838 Long Lists of Gene Ontology Terms. *Plos One* 6.
- 839 Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M.
840 Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, and C. von
841 Mering. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of
842 life. *Nucleic Acids Research* 43:D447-D452.
- 843 Verhoeven, K. J. F., B. M. Vonholdt, and V. L. Sork. 2016. Epigenetics in ecology and
844 evolution: what we know and what we need to know INTRODUCTION. *Molecular Ecology*
845 25:1631-1638.
- 846 Verhulst, E. C., A. C. Mateman, M. V. Zwier, S. P. Caro, K. J. F. Verhoeven, and K. Van
847 Oers. 2016. Evidence from pyrosequencing indicates that natural variation in animal
848 personality is associated with DRD4 DNA methylation. *Molecular Ecology* 25:1801-1811.
- 849 Viitaniemi, H. M., I. Verhagen, M. E. Visser, A. Honkela, K. van Oers, and A. Husby. 2019.
850 Seasonal Variation in Genome-Wide DNA Methylation Patterns and the Onset of Seasonal
851 Timing of Reproduction in Great Tits. *Genome Biology and Evolution* 11:970-983.
- 852 Watanabe, S., K. Watanabe, V. Akimov, J. Bartkova, B. Blagoev, J. Lukas, and J. Bartek.
853 2013. JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin
854 response to DNA breaks. *Nature Structural & Molecular Biology* 20:1425-1433.
- 855 Weaver, I. C. G., N. Cervoni, F. A. Champagne, A. C. D'Alessio, S. Sharma, Seckl, Jr., S.
856 Dymov, M. Szyf, and M. J. Meaney. 2004. Epigenetic programming by maternal behavior.
857 *Nature Neuroscience* 7:847-854.
- 858 Weng, X. L., D. Z. Zhou, F. T. Liu, H. Zhang, J. Y. Ye, Z. Zhang, D. Zhang, Y. N. Wang, L.
859 M. Tao, L. Cao, M. Y. Kan, T. Wang, G. Y. Feng, X. L. Qin, J. H. Sun, L. He, and Y. Liu.
860 2014. DNA methylation profiling in the thalamus and hippocampus of postnatal
861 malnourished mice, including effects related to long-term potentiation. *Bmc Neuroscience*
862 15:8.
- 863 Wenzel, M. A., and S. B. Piertney. 2014. Fine-scale population epigenetic structure in
864 relation to gastrointestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Molecular*
865 *Ecology* 23:4256-4273.
- 866 Wreczycka, K., A. Goidschan, D. Yusuf, B. Gruning, Y. Assenov, and A. Akalin. 2017.
867 Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology* 261:105-115.
- 868 Wright, R. O., J. Schwartz, R. J. Wright, V. Bollati, L. Tarantini, S. K. Park, H. Hu, D.
869 Sparrow, P. Vokonas, and A. Baccarelli. 2010. Biomarkers of Lead Exposure and DNA
870 Methylation within Retrotransposons. *Environmental Health Perspectives* 118:790-795.
- 871 Wu, J., M. R. Basha, B. Brock, D. P. Cox, F. Cardozo-Pelaez, C. A. McPherson, J. Harry, D.
872 C. Rice, B. Maloney, D. Chen, D. K. Lahiri, and N. H. Zawia. 2008. Alzheimer's disease
873 (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead
874 (pb): Evidence for a developmental origin and environmental link for AD. *Journal of*
875 *Neuroscience* 28:3-9.

876 Wu, Y. H., A. V. Patchev, G. Daniel, O. F. X. Almeida, and D. Spengler. 2014. Early-Life
877 Stress Reduces DNA Methylation of the Pomc Gene in Male Mice. *Endocrinology* 155:1751-
878 1762.

879 Zhang, Y., S. Baheti, and Z. F. Sun. 2018. Statistical method evaluation for differentially
880 methylated CpGs in base resolution next-generation DNA sequencing data. *Briefings in*
881 *Bioinformatics* 19:374-386.

882

883

884 **Tables**

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

Metal	HIGH n = 12	SMELTER n = 11	CONTROL n = 12	$F_{\text{ndf, ddf}}$	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 (0.25-0.62) a	4.3 (2.7-6.9) b	0.60 (0.38-0.95) a	24.4 _{2,32}	<0.0001
Cd	0.73 (0.51-1.04) a	1.93 (1.34-2.78) b	0.54 (0.38-0.76) a	10.2 _{2,32}	0.0004
Cu	34.6 (27.1-44.1) a	111 (86-144) b	29.6 (23.2-29.6) a	12.9 _{2,32}	<0.0001
Ni	2.24 (1.6-3.1) a	20.6 (14.5-29.2) b	1.99 (1.4-2.8) a	34.5 _{2,32}	<0.0001

889

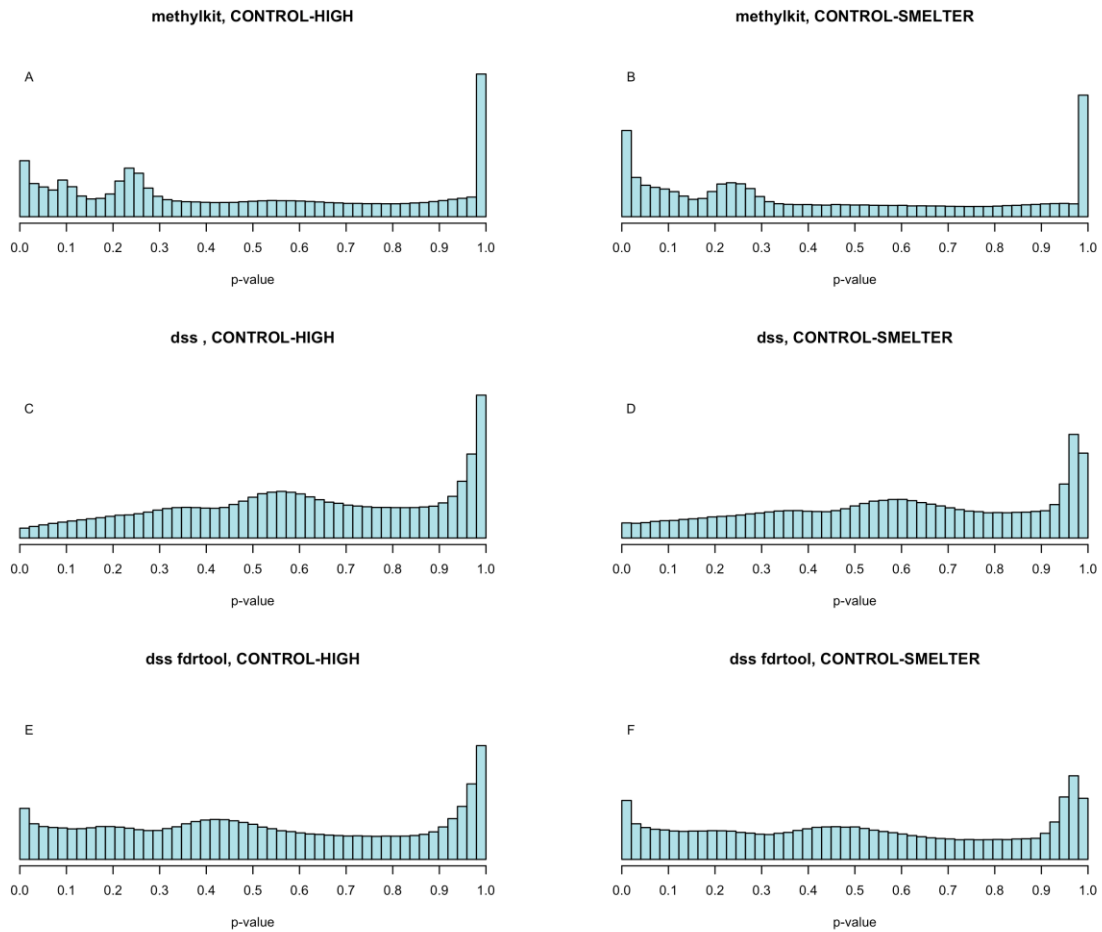
890

891 **Table 2.** The number and percentage (in parentheses) of p-values for CpG sites less than 0.05
 892 in different comparisons. Fdr refers to multiple testing correction using Benjamini &
 893 Hochberg (1995) method. The test statistics of goodness-of-fit test (χ^2) of the p-value
 894 histograms indicated deviations uniform p-value distribution in unfiltered data, which
 895 improved (i.e. higher χ^2 values) 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 (beta-binomial)	62789 (6.1)	520	192880
fdrtool CONTROL SMELTER (beta-binomial)	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

896

898

899 **Figures**

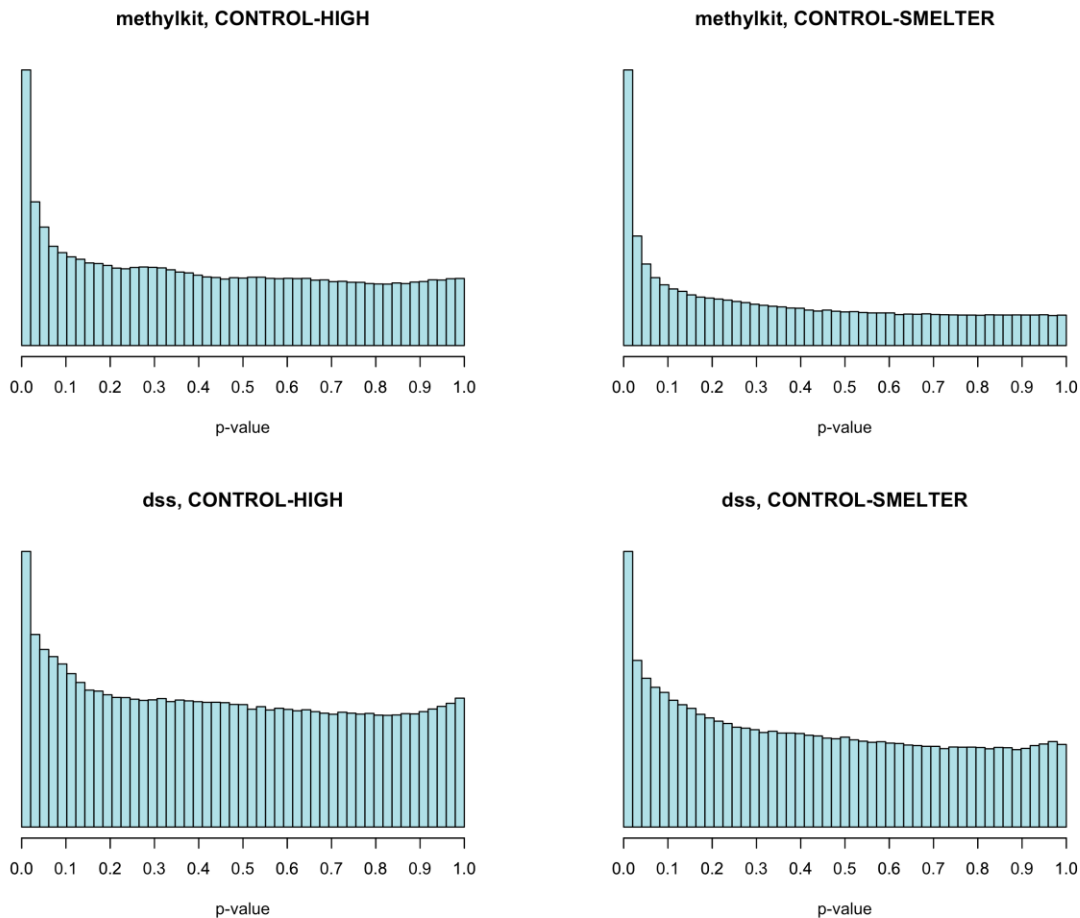
900

901 **Figure 1**

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

906

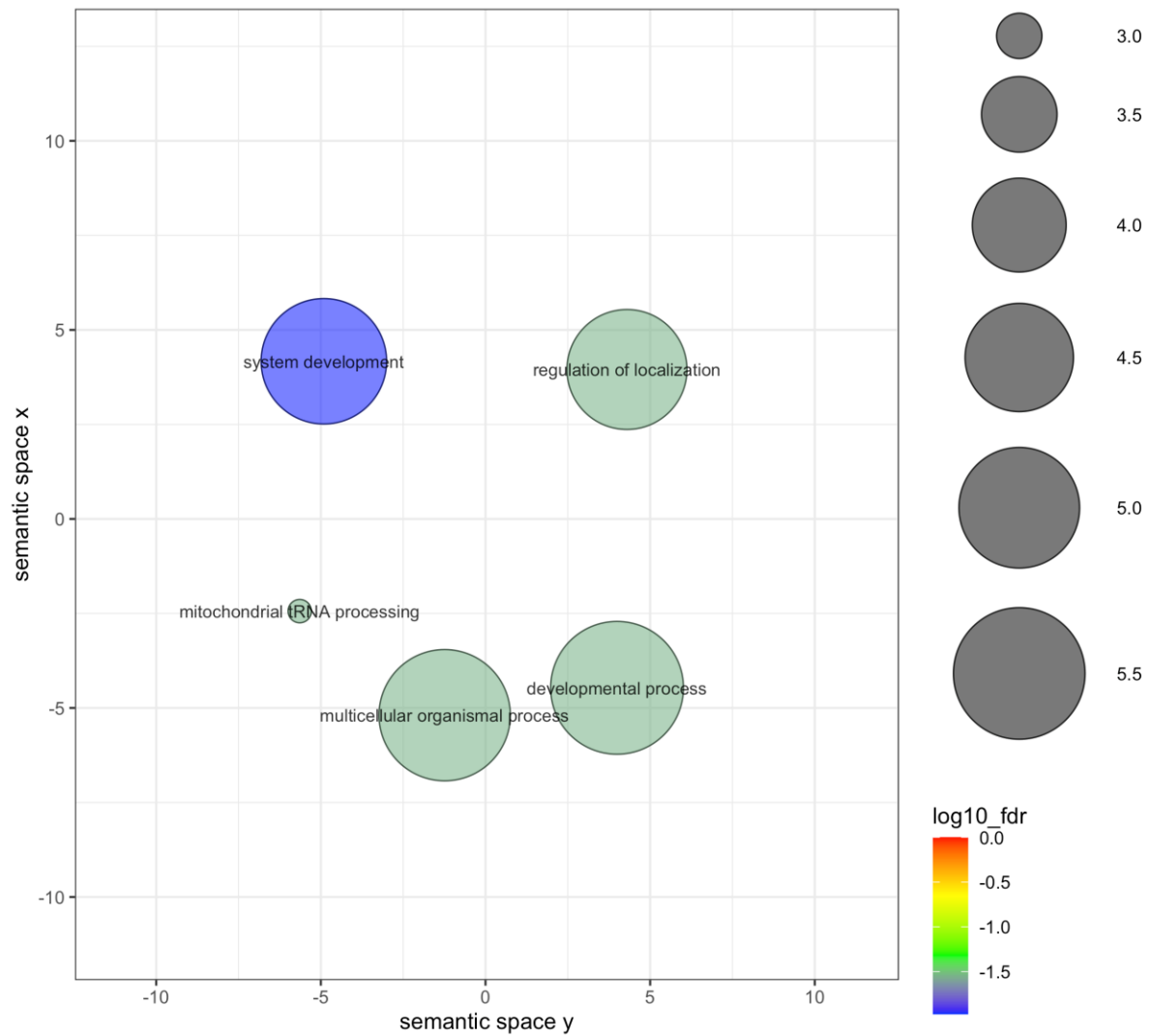
907



908

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

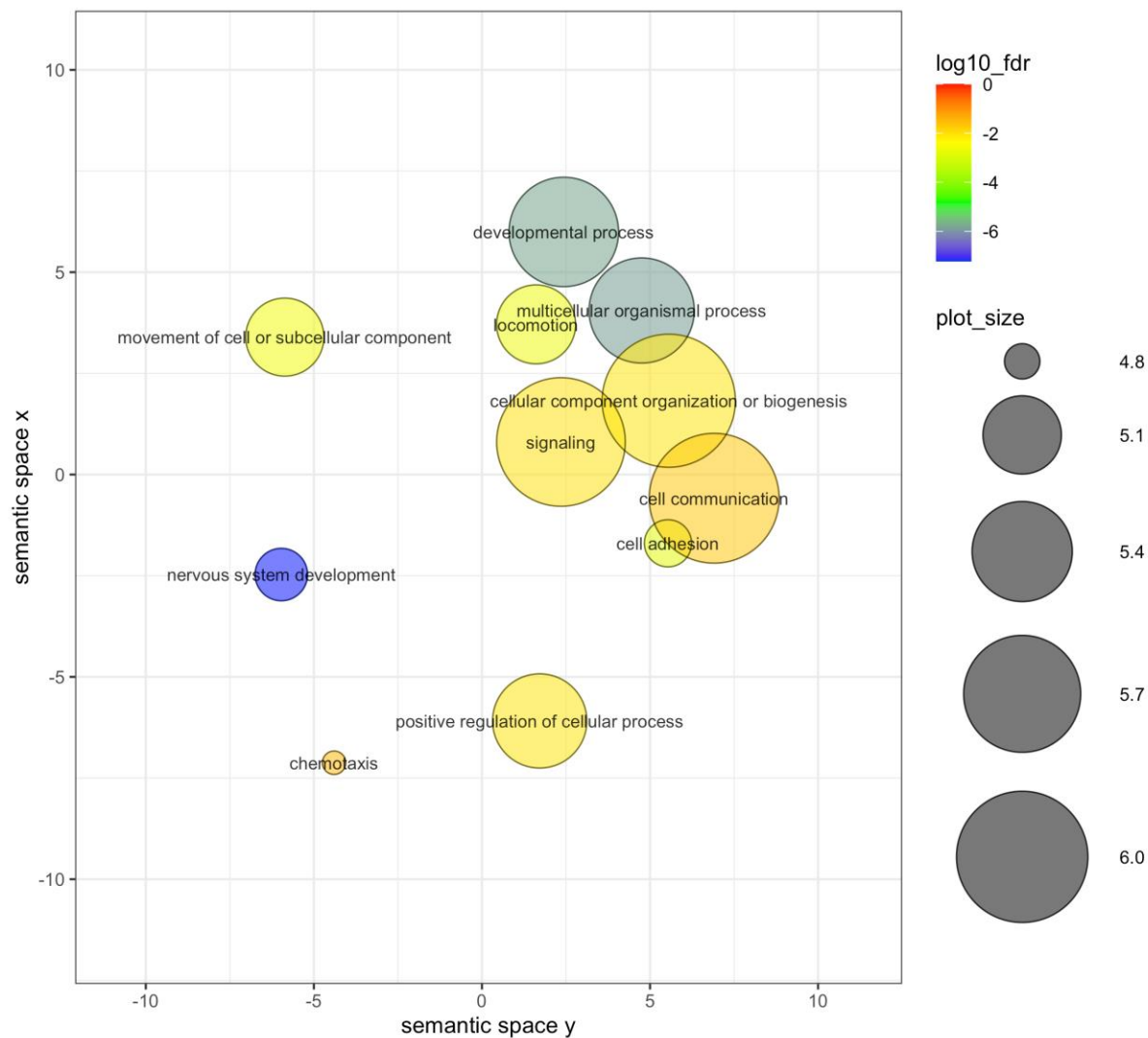
912



913

914

915 Figure 3. Results of the gene enrichment test for DMRs identified in the binomial regression
 916 in the CONTROL-HIGH comparison after merging semantically similar gene ontology
 917 categories. Circle size indicates the frequency of the GO term in the underlying GO database
 918 (bubbles of more general terms are larger; <http://revigo.irb.hr/>) and color scale shows the fdr
 919 (log10 scale) of the representative GO term for each merged category.



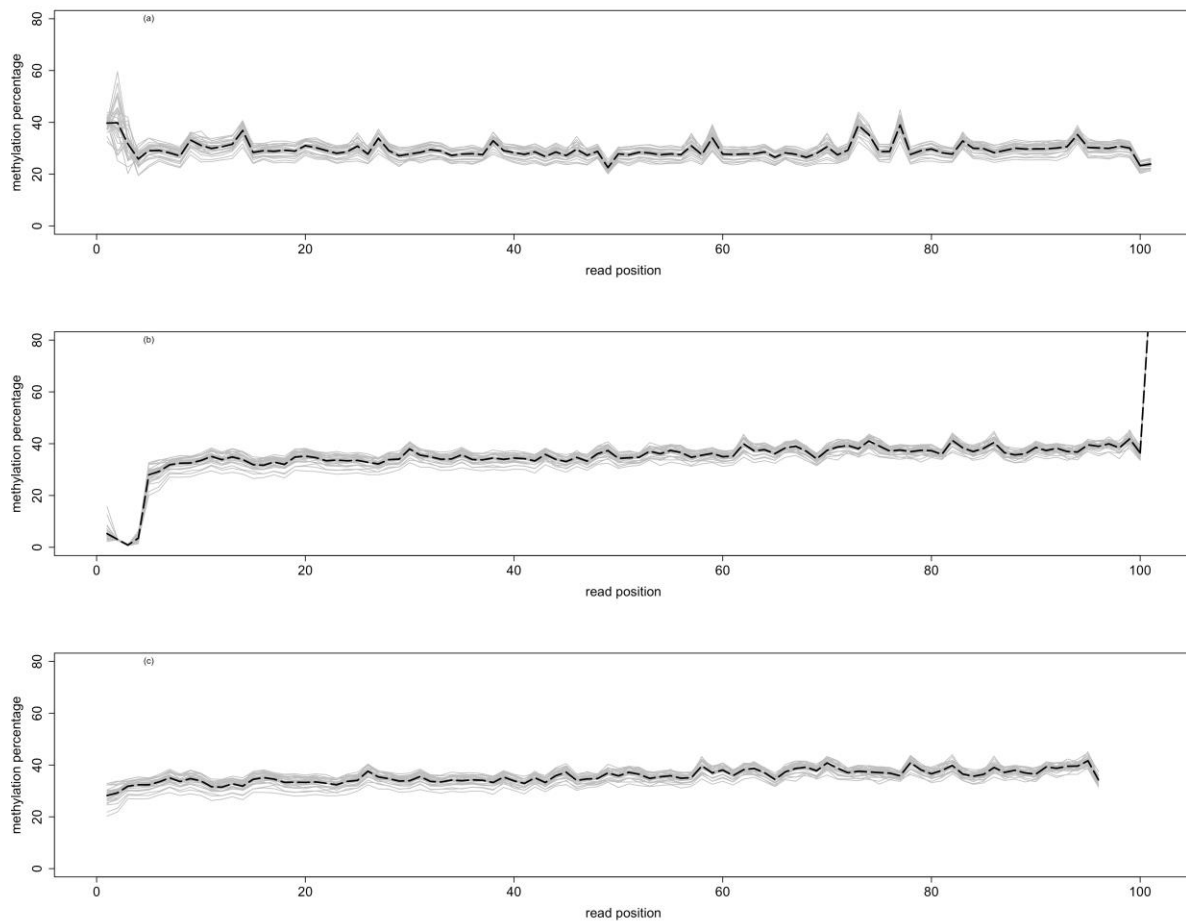
920

921 Figure 4. Results of the gene enrichment test for DMRs identified in the binomial regression
 922 in the CONTROL-SMELTER comparison after merging semantically similar gene ontology
 923 categories. Circle size indicates the frequency of the GO term in the underlying GO database
 924 (bubbles of more general terms are larger; <http://revigo.irb.hr/>) and color scale shows the fdr
 925 (\log_{10} scale) of the representative GO term for each merged category.

926

927

928

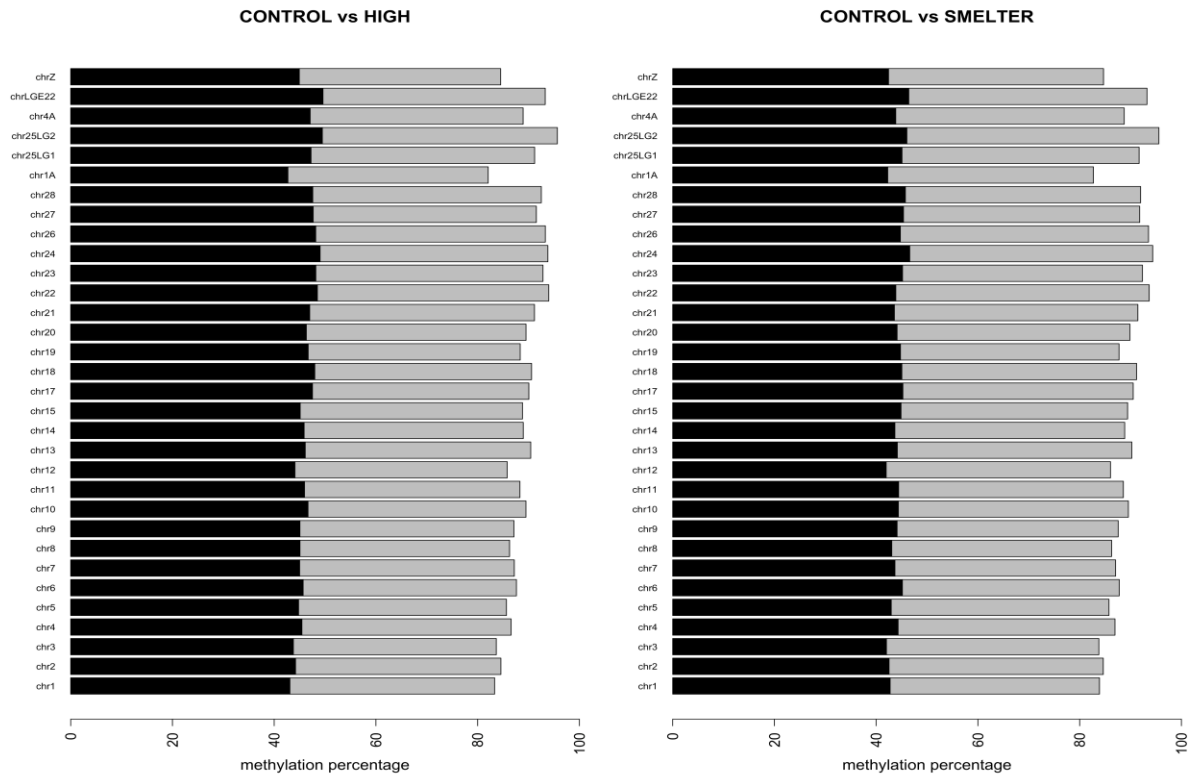
929 **Supplementary material**

930

931 **Supplementary figure 1**

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

938

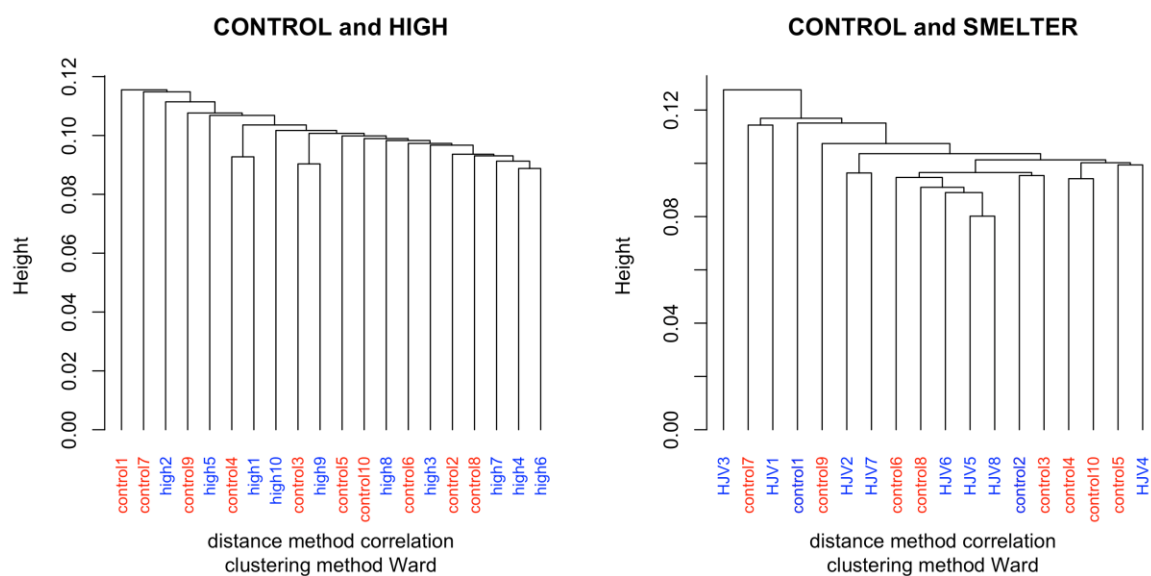


939

940 Supplementary figure 2

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

943

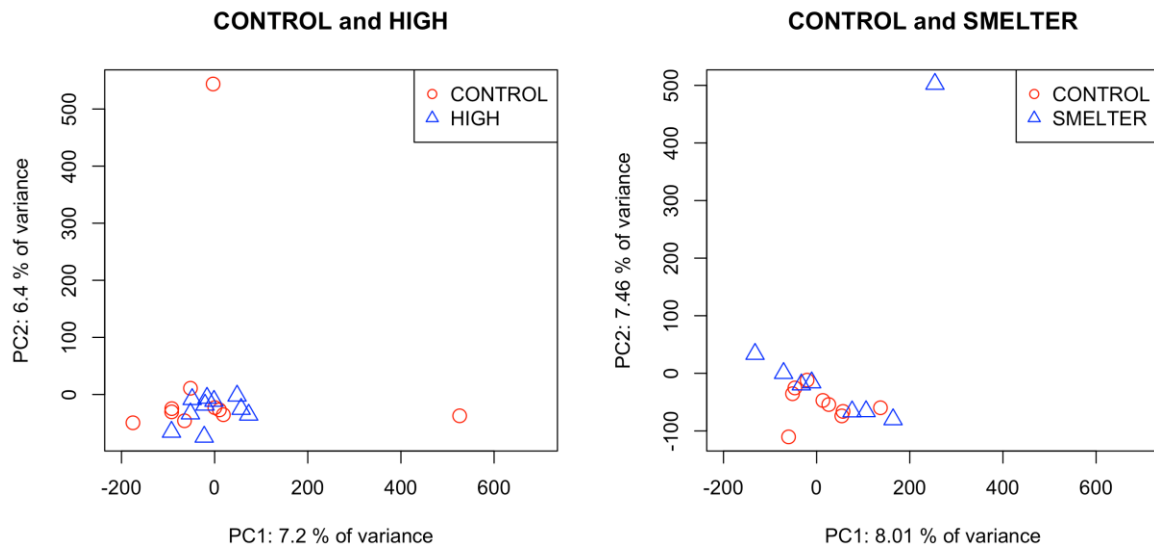


944

945 Supplementary figure 3

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

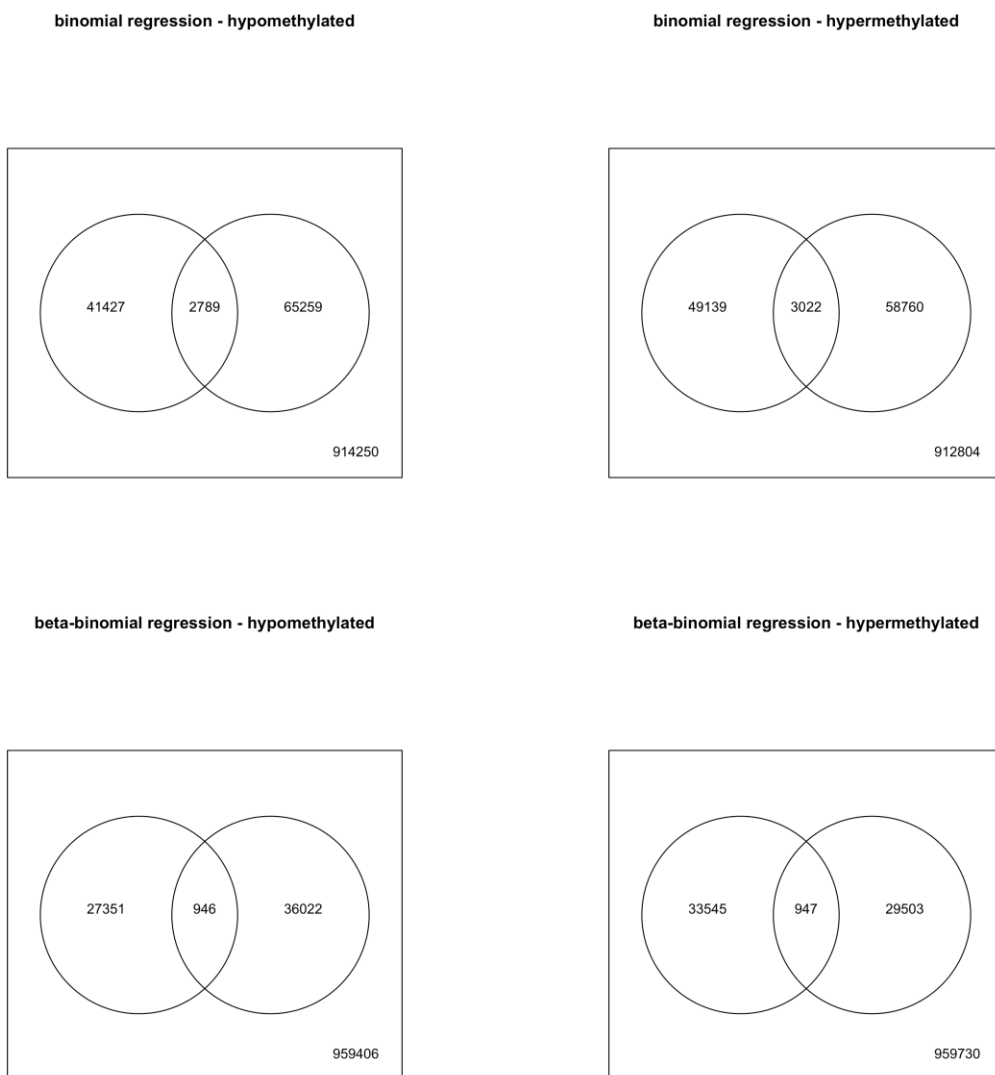
948



949

950 Supplementary figure 4

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

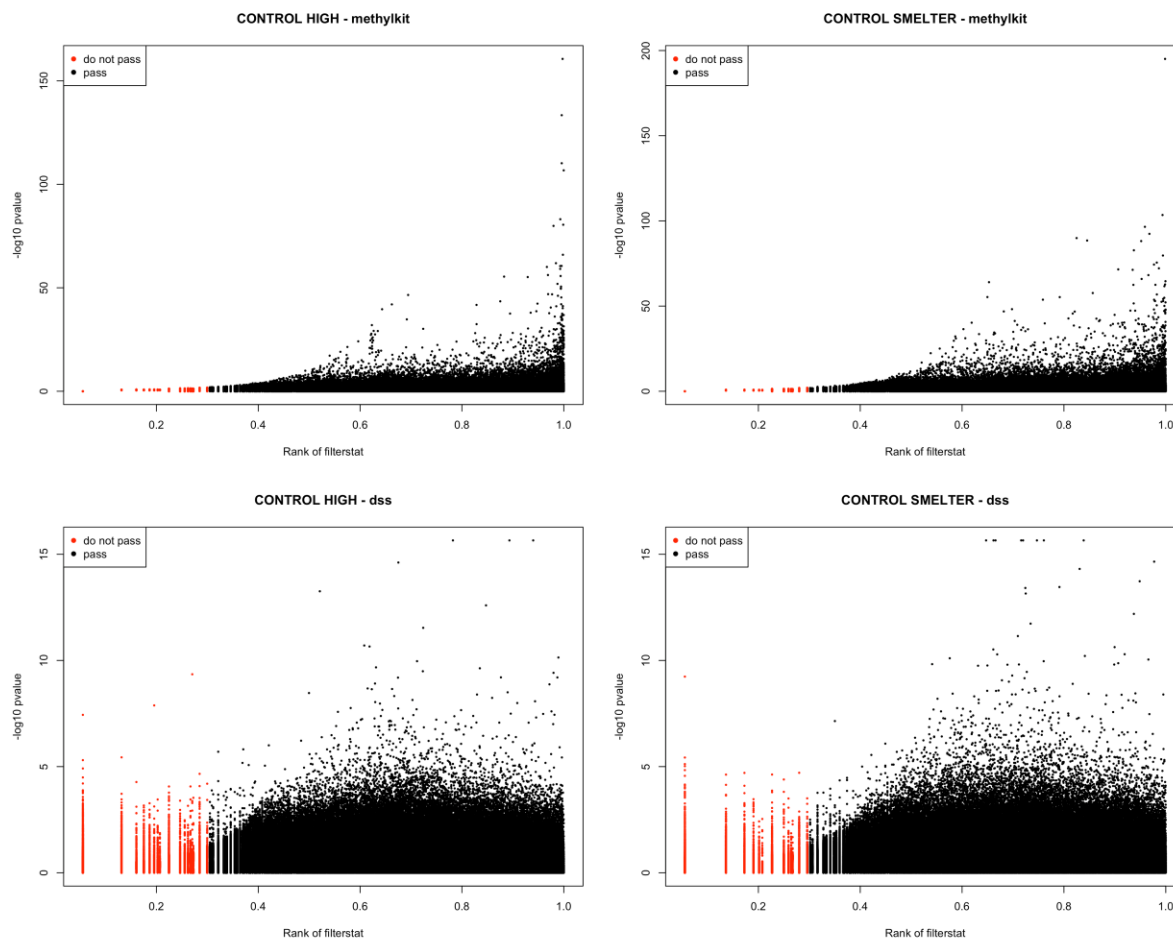


954

955 Supplementary figure 5.

956 Venn diagrams showing the overlap of the significant CpGs between CONTROL-HIGH (left
 957 side of each diagram) and CONTROL-SMELTER (right side of each diagram) comparisons.

958 In the upper panel is the overlap in binomial regression and in the lower panel the overlap in
 959 beta-binomial regression. Numbers at the lower right in each box show total number of sites.

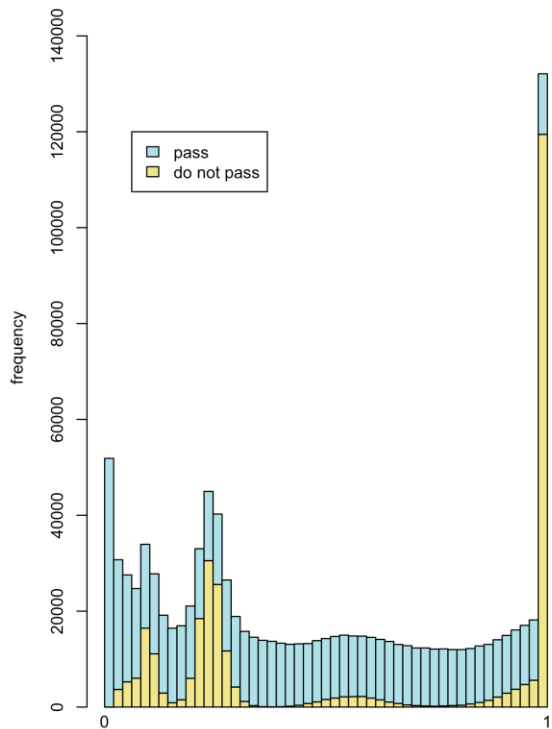


960

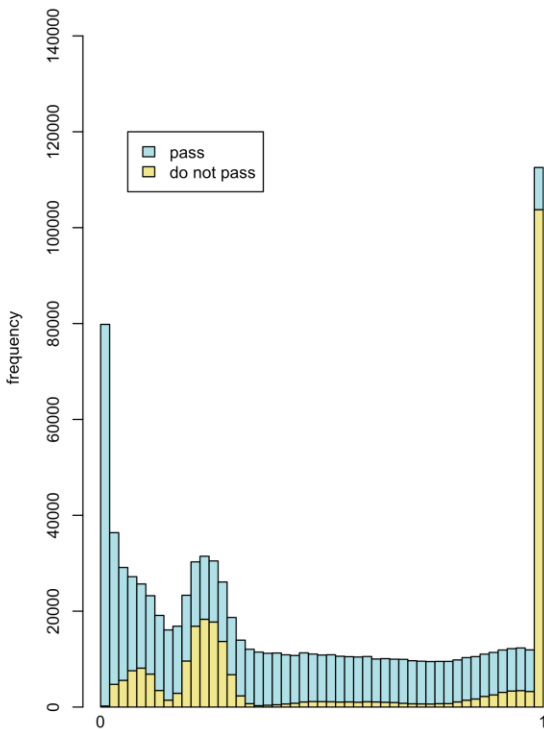
961 Supplementary figure 6. The principle of filtering uninformative CpGs. On the x-axis is the
 962 rank of filter statistics i.e. the mean coverage of Cs in each analyzed CpG. On the y-axis is
 963 the $-\log_{10}$ p-value obtained either from binomial regression (methylkit) or from beta
 964 binomial regression (dss). The red filled circles are CpGs not passing the filtering threshold
 965 (30%) while the filled black circles are CpGs passing the filtering threshold.

966

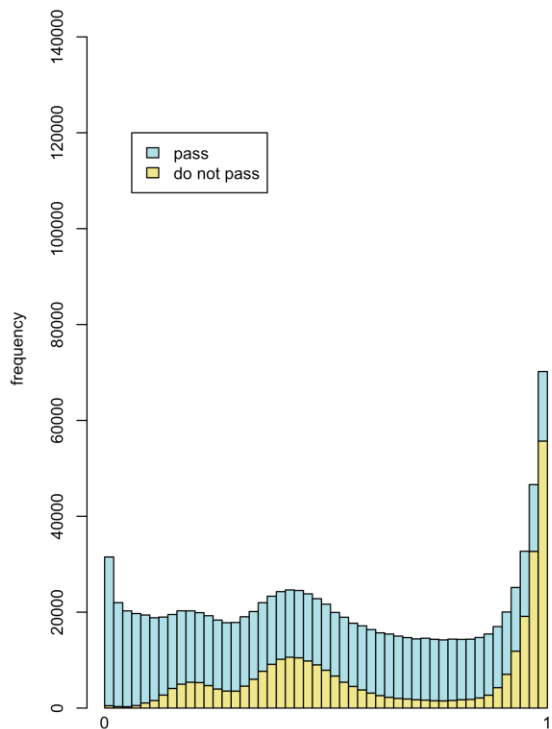
CONTROL HIGH - methylkit



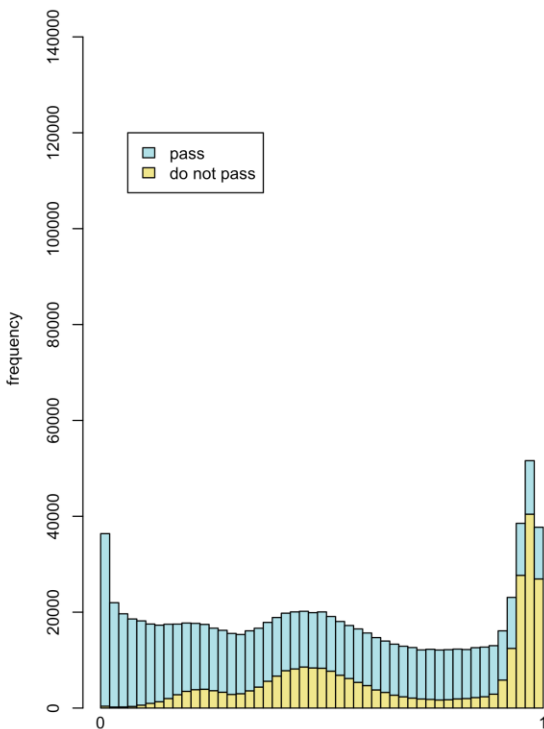
CONTROL SMELTER - methylkit



CONTROL HIGH - dss



CONTROL SMELTER - dss



968 Supplementary figure 7. The filtering approach shown as p-value histogram. The yellow
969 color shows the effect of removing low coverage CpG sites along the p-value distribution.
970 The blue colour shows the p-values remaining after the low coverage filtering.

971

972 ***DNA isolation***

973 DNA isolation was performed at the Center of Evolutionary Applications (University of
974 Turku, Finland). We used whole blood samples, which can be acquired without sacrificing
975 the individuals. In birds erythrocytes have nuclei and therefore >95% of the gained DNA is
976 from erythrocytes. DNA was extracted from c. 10 – 20 µl whole blood using the salt
977 extraction method modified from Aljnabi & Martinez (1997). Extracted DNA was treated
978 with RNase-I according to the manufacturer's protocol. DNA concentrations were measured
979 fluorometrically with a Qubit High Sensitivity kit (ThermoFisher Scientific) and we assessed
980 DNA integrity by running each DNA sample on an agarose gel.

981

982 ***RRBS library preparation***

983 We used a reduced representation bisulfite sequencing (RRBS) approach, which enriches the
984 regions of the genome that have a high CpG content (Meissner et al. 2005). It was previously
985 shown in the study species that the vast majority of the methylated cytosines (97%) were
986 derived from CpG context in blood (Derks et al. 2016). Sequencing was conducted at the
987 Finnish Microarray and Sequencing Center in Turku, Finland. The library preparation was
988 started from 200 ng of genomic DNA and was carried out according to a protocol adapted
989 from Boyle et al. (2012). The first step in the workflow involved the fragmentation of
990 genomic DNA with MspI where the cutting pattern of the enzyme (C[^]CGG) was used to
991 systematically digest DNA to enrich for CpG dinucleotides. After a fragmentation step, a

992 single reaction was carried out to end repair and A-tail (required for the adapter ligation) the
993 MspI digested fragments using Klenow fragment (3' => 5' exo), following the purification of
994 A-tailed DNA with bead SPRI clean-up method (AMPure magnetic beads). A unique
995 Illumina TruSeq indexing adapter was then ligated to each sample during adapter ligation
996 step to be able to identify pooled samples of one flow cell lane. To reduce the occurrence of
997 adapter dimers, a lower concentration of adapters (1:10 dilution) was used than recommended
998 by the manufacturer. These ligated DNA fragments were purified with the bead SPRI clean-
999 up method before putting samples through bisulfite conversion to achieve C-to-U conversion
1000 of unmethylated cytosines, whereas methylated cytosines remain intact. Bisulfite conversion
1001 and sample purification were done according to the Invitrogen MethylCode Bisulfite
1002 Conversion Kit. Aliquots of converted DNA were amplified by PCR (16 cycles) with
1003 Taq/Pfu Turbo Cx Polymerase, a proofreading PCR enzyme that does not stall when it
1004 encounters uracil, the product of the bisulfite reaction, in the template. PCR-amplified RRBS
1005 libraries were purified using two subsequent rounds of SPRI bead clean-ups to minimize
1006 primer dimers in the final libraries. The high quality of the libraries was confirmed with
1007 Advanced Analytical Fragment Analyzer and the concentrations of the libraries were
1008 quantified with Qubit® Fluorometric Quantitation, Life Technologies. We selected fragment
1009 sizes ranging between 150 – 1000 bp (average sizes were 250-350 bp) for sequencing.

1010

1011 *Sequencing*

1012 The samples were normalized and pooled for the automated cluster preparation, which was
1013 carried out with an Illumina cBot station. The 28 libraries were randomly combined in three
1014 pools, 10 or 8 samples in each pool and sequenced in 3 lanes. The samples were sequenced
1015 with an Illumina HiSeq 2500 instrument using TruSeq v3 sequencing chemistry. Paired-end
1016 sequencing with 2 x 100 bp read length was used with 6 bp index run.

1017

1018 Supplementary Table 1. Top 10 DMRs for CONTROL-HIGH and CONTROL-SMELTER comparisons, separately for binomial and
 1019 betabinomial models. Mean Met% = average methylation percentage. Met% difference = difference, in percentages, in methylation between two
 1020 groups. CO = control, HI = HIGH, SME = SMELTER

CHROMOSOME	POSITION (bp)	ANNOTATION	DMR length	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 binomial								
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 length	No. CpGs	Mean met% CO	Mean met% HI/SME	Met% difference
<i>C) CO-HIGH betabinomial</i>						
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
<i>D) CO-SMELTER betabinomial</i>						
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

1021

1022 *Genes at differentially methylated regions*

1023 Among the set of 10 most differentially methylated regions across the treatment groups, we
1024 identified 12 genes (*POMC*, *ITGA11*, *LEKRI*, *USH2A*, *ZPR1*, *JMJD1C*, *ADAMTS3*, *PDE1C*,
1025 *TBP*, *PAPD4*, *GCC1* and *UTRN*). Of these 12, *POMC*, showed lower methylation (ca. 20%,
1026 thus theoretically higher expression) in methylation *both* in SMELTER and HIGH treatments
1027 compared to CONTROL. *POMC* (pro-opiomelanocortin) is a neuronal hormone, which is
1028 cleaved to multiple key by-products, including (i) corticotropin (ACTH), controlling the
1029 stress response, (ii) appetite control and (iii) b-endorphin (Marco et al. 2016). Methylation of
1030 *POMC* has been associated with nutritional state (in rats, Ramamoorthy et al. 2018), maternal
1031 under nutrition (in ovine Stevens et al. 2010) and offspring early-life stress (in mice, Wu et
1032 al. 2014). Here, we report for the first time that *POMC* methylation may mediate early-life
1033 stress (nutritional and/or metal exposure) also in a wild vertebrate population. Furthermore,
1034 methylation of another stress related gene, *PDEC1* (phosphodiesterase 1C) was also
1035 decreased in SMELTER compared to control. Expression of *PDEC1* gene has been found to
1036 be associated with aldosterone stress hormone in chicken (e.g. Fallahsharoudi et al. 2017).

1037 The other differentially methylated genes in relation to metal exposure were related to
1038 (i) DNA damage: *JMJD1C* is a candidate histone demethylase and also plays a role in the
1039 pathway DNA-damage response (e.g. Watanabe et al. 2013). Our data suggests that its
1040 methylation was decreased (theoretical expression increased) in high Pb exposure compared
1041 to control, which is logical given that Pb exposure is likely to cause more oxidative stress and
1042 DNA damage (e.g. Wu et al. 2008, Rainio et al. 2015a). Other studies on Pb exposure also
1043 found differences in methylation in detoxification pathways (Sen et al. 2015); (ii) growth and
1044 development: *LEKRI* (Freathy et al. 2010), *ADAMTS3* (Janssen et al. 2016), *UTNR* (in
1045 mammals, e.g. Schofield et al. 1993). Furthermore, Pb has specifically been found to impair

1046 neurodevelopment (e.g. Morgan et al. 2000, Burger and Gochfeld 2005). Our data shows that
1047 methylation of *ZPRI* (zinc finger protein gene), an important protein in neural development
1048 (e.g. Doran et al. 2006) was increased by ca. 20% (theoretical expression decreased) in Pb
1049 exposure compared to controls, which may warrant further studies on *ZPRI*, Pb exposure and
1050 neurodevelopment. Previous studies have reported alternation in methylation of other genes
1051 related to neurodevelopment, such as another zinc finger protein gene, *Zfp974* and *Zfp787*,
1052 *ARTN*, *C5aRI* (Dou et al. 2019), *Syt2*, *Prkg1*, *Pcdhb20*, *Slc2a3*, *Klhl1* and *Snap29* (Singh et
1053 al. 2018) and *PAX1* and *MSH1* (Senut et al. 2014) (iii) transcription and intracellular
1054 processes: *TBP* is universal transcription factor required for all of the eukaryotic RNA
1055 polymerases (Shimada et al. 2003), *PAPD4* is a poly(A) RNA polymerase (Burroughs et al.
1056 2010), while *GCCI* is associated with Golgi apparatus structure (Gosavi et al. 2018). Thus,
1057 these genes may serve as potential candidates for further studies on the effects of pollution on
1058 organisms via DNA methylation.

1059