

1 **Identification of differential hypothalamic DNA methylation and gene  
2 expression associated with sexual partner preferences in rams.**

3 Surajit Bhattacharya<sup>1</sup>, Rebecka Amodei<sup>2</sup>, Eric Vilain<sup>1</sup> and Charles E. Roselli<sup>2\*</sup>

4 <sup>1</sup>Center for Genetic Medicine Research, Children's Research Institute, Children's National  
5 Hospital, Washington, DC, United States of America

6 <sup>2</sup>Department of Chemical Physiology & Biochemistry, Oregon Health & Science University  
7 Portland, Oregon, United States of America

8 **Abbreviated title:** Methylation and Partner Preference in Rams

9 **Keywords:** Sexual partner preference, methylation, epigenetics, RNA-Seq, reduced  
10 representation bisulfite sequencing, DNA methylation, sheep, hypothalamus

11 \* Corresponding author

12 E-mail: [rosellic@ohsu.edu](mailto:rosellic@ohsu.edu) (CER)

13

14

15

16

17

18

19

20 **Abstract**

21 The sheep is a valuable model to test whether hormone mechanisms that sexually differentiate  
22 the brain underlie the expression of sexual partner preferences because as many as 8% of rams  
23 prefer same-sex partners. Epigenetic factors such as DNA methylation act as mediators in the  
24 interaction between steroid hormones and the genome. Variations in the epigenome could be  
25 important in determining morphological or behavior differences among individuals of the same  
26 species. In this study, we explored DNA methylation differences in the hypothalamus of male  
27 oriented rams (MORs) and female oriented rams (FORs). We employed reduced representation  
28 bisulfite sequencing (RRBS) to generate a genome-wide map of DNA methylation and RNA-Seq  
29 to profile the transcriptome. We found substantial DNA methylation and gene expression  
30 differences between FORs and MORs. Although none of the differentially methylated genes  
31 yielded significant functional terms directly associated with sex development, three  
32 differentially expressed genes were identified that have been associated previously with sexual  
33 behaviors. We hypothesize that these differences are involved in the phenotypic variation in  
34 ram sexual partner preferences, whereas future studies will have to find the specific  
35 mechanisms. Our results add an intriguing new dimension to sheep behavior that should be  
36 useful for further understanding epigenetic and transcriptomic involvement.

37 **Introduction**

38 The mechanisms underlying the development of sexual orientation remain unknown. A large  
39 amount of empirical data suggest that genes and prenatal hormones are important  
40 determinants [1]. Given that sexual orientation represents one of the largest sex differences in  
41 humans, the leading neurohormone theory posits that like other sexually dimorphic behaviors,  
42 sexual orientation reflects the sexual differentiation of the brain under the influence of  
43 androgens. Simply stated, exposure to high levels of androgens during a critical period of  
44 gestation (i.e., most males and a few females) programs attraction to females in adulthood.  
45 While exposure to low levels of androgens (i.e., most females and a few males) programs sexual  
46 attraction to males. There is also compelling evidence implicating the involvement of  
47 epigenetic mechanisms in mediating the long-term effects of hormones on the sexual  
48 differentiation of the brain in animal models [2–4]. Evidence in rodents suggests that perinatal  
49 androgen exposure reduces DNA methylation in male brains compared to female brains,  
50 releasing masculinizing genes from epigenetic repression and ultimately masculinizing sexual  
51 behavior [5] and brain anatomy [6]. It is not known currently whether epigenetic factors  
52 influence human sexual orientation although circumstantial evidence suggests that it could [4].  
53 Domestic rams have emerged as an important animal model for human sexual orientation.  
54 Approximately 8% of rams in natural populations of common western breeds can be reliably  
55 identified to show exclusive and enduring sexual partner preference for either the opposite sex  
56 (female-oriented) or same sex (male-oriented)[7]. Like men, rams have a sexually dimorphic  
57 nucleus (SDN) in the preoptic area/anterior hypothalamus [8,9]. The volume of the ovine SDN  
58 correlates with sexual partner preference and is larger in female-oriented rams than in male-  
59 oriented rams and ewes. The precise function of the ovine SDN remains unclear but its volume

60 has been shown to be a biomarker of prenatal androgen exposure [10]. Thus, the volume  
61 difference between sexes and between female- and male-oriented rams most likely results  
62 from a developmental difference in androgen exposure and may be reflected in differences in  
63 DNA methylation states in the brain [11]. The medial basal hypothalamus is another brain area  
64 that plays crucial roles in neuroendocrine control systems and sexual behaviors [12]. The  
65 ventromedial nucleus is a major anatomical component of the medial basal hypothalamus that  
66 is larger in males than in females, regulated by perinatal hormone exposure, and involved in  
67 facilitating male sexual behavior[13–16]. The present study evaluated the genome-wide  
68 epigenetic and transcriptomic levels of the medial basal hypothalamus in female- and male-  
69 oriented rams. We hypothesize that the DNA methylome and transcriptome of the  
70 hypothalamus differs between these rams as evidence of a legacy of differential androgen  
71 exposure during early fetal development.

## 72 **Materials and Methods**

### 73 **Animals and behavioral classifications**

74 Archival hypothalamic tissues were used in this study. The tissue was obtained from 4-5-year-  
75 old adult rams that were given behavioral tests at the USDA Sheep Experiment Station in  
76 Dubois, ID and classified as male-oriented rams (MORs) (n = 5) or female-oriented rams (FORs)  
77 (n = 4). The sheep were of mixed western breeds, including Rambollet, Targhee and Polypay.  
78 Rams were given sexual partner preference tests administered as described previously [17].  
79 Those that exclusively mounted other rams were classified as male oriented rams (MORs),

80 whereas rams that exclusively mounted females were classified as female oriented rams  
81 (FORs); (Table 1). All experimental animal protocols met the stipulations and guidelines of the  
82 NIH policy on the Care and Use of Laboratory Animals and were approved by the Institutional  
83 Animal Care and Use Committee of the Oregon Health and Science University.

84 **Table 1: Number of mounts on female and male stimulus animals in the last two of four**  
85 **sexual partner preference tests.**

Ram Number	Classification	F/M Mounts* SPP Test #3	F/M Mounts SPP Test #4
R2579	MOR/SSP	0/(36)†	0/38
R2810	MOR/SSP	0/(40)	0/3
A8337	MOR/SSP	0/97	0/6
T8384	MOR/SSP	0/(56)	0/(20)
R2423	MOR/SSP	0/(15)	0/0
A8736	FOR/OSP	(7)/0	(9)/0
R3139	FOR/OSP	(9)/0	(8)/0
R3362	FOR/OSP	(7)/0	(8)/0
A9707	FOR/OSP	(6)/0	(10)/0

86  
87 **Sample collection and preparation**  
88 The sheep were euthanized with an overdose (15 mg/kg) of sodium pentobarbital (Euthozol;  
89 Delmarval Laboratories Inc, Midlothian, VA). The medial basal hypothalamus was dissected as a  
90 block of tissue that extended from the caudal aspect of the optic chiasm to the rostral aspect of  
91 the mammillary bodies, bilaterally to the optic nerves and dorsally to the top of the third  
92 ventricle. The dissection was split through the ventricle into left and right halves that were  
93 frozen immediately on dry ice and stored in a -80° C freezer. Genomic DNA was extracted from  
94 one half of the hypothalamus using the DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD,  
95 USA) and concentrated using the Genomic DNA Clean & Concentrator Kit (Zymo Research,

96 Irvine, CA, USA) as directed by the manufacturer. The concentration and quality of genomic  
97 DNA was verified with absorbance spectroscopy and Qubit fluorimetry (ThermoFisher Scientific,  
98 Waltham, MA, USA). RNA was extracted from the remaining half of the hypothalamus using the  
99 RNeasy Mini kit (Qiagen). RNA was quantified with the Qubit RNA Broad range kit  
100 (Thermofisher) and integrity was verified on a 4200 Tape station (Agilent, Santa Clara, CA, USA).  
101 All RNA samples that were used in these studies had RIN values greater than 8.0.

102 **Reduced representation bisulfite sequencing**

103 To analyze DNA methylation, we used reduced representation bisulfite sequencing (RRBS) [17],  
104 a genome-wide approach that examines about 2 million CpGs (7-10% of all CpGs in genome)  
105 that are highly enriched key regulatory regions including promoters, CpG islands and CpG island  
106 shores.

107 To generate RRBS libraries, ~150ng of sheep genomic DNA was digested overnight with the  
108 restriction enzyme Mspl (New England Biolabs, Ipswich, MA, USA). The DNA was then purified  
109 with AMPure XP magnetic beads (Beckman Coulter, Pasadena, CA, USA) before use with the  
110 NEXTflex Bisulfite-Seq Kit (BioScientifica, Bristol, UK). The DNA was then end repaired, A-tailed  
111 and ligated with the NEBNext Methylated Adaptor (New England Biolabs). The ligated DNA was  
112 size-selected using AMPure XP magnetic beads to produce a final library size of 350 bp. Bisulfite  
113 conversion was performed with the EZ DNA Methylation-Gold Kit (Zymo Research) before  
114 carrying out PCR amplification with NEBNext Multiplex Oligos (New England Biolabs) to barcode  
115 each library. A final AMPure XP bead purification was performed, and the resulting libraries  
116 were quantified with the Qubit High Sensitivity double stranded (dsDNA) Assay (Life

117 Technologies, Carlsbad, CA, USA) and the Bioanalyzer High Sensitivity Analysis (Agilent).  
118 Libraries were multiplexed and sequenced on the Illumina NextSeq or HiSeq2500 to obtain ~30  
119 million single end, 75 bp reads. The sequence data was deposited under the gene expression  
120 omnibus (GEO) accession number *GSE158287*. Library names and associated phenotypes are in  
121 Table 2.

122 **Table 2: Sample names and associated Phenotypes**

Sample Name	Phenotype
LIB181217CR_ECL28_1_S1	MOR1
LIB181217CR_ECL28_2_S2	MOR2
LIB181217CR_ECL28_3_S3	MOR3
LIB181217CR_ECL28_4_S4	MOR4
LIB181217CR_ECL28_5_S5	MOR5
LIB181217CR_ECL28_6_S6	FOR1
LIB181217CR_ECL28_7_S7	FOR2
LIB181217CR_ECL28_8_S8	FOR3
LIB181217CR_ECL28_9_S9	FOR4

123  
124 **Bioinformatic analysis**  
125 Quality reports for all the nine sample sequences (five MORs and four FORs), were generated  
126 using FastQC [18] (generates per sample quality report) and MultiQC [19](generates a multi  
127 sample quality report, by aggregating the individual FastQC reports). All samples passed the per  
128 base sequence quality metrics, i.e. none of the bases have their lower quartile less than 10  
129 Phred score[20,21] or median less than 25 Phred score (FastQC and MultiQC are in  
130 <https://github.com/VilainLab/SheepMethylation/tree/master/FastQC> and  
131 <https://github.com/VilainLab/SheepMethylation/tree/master/MultiQC> respectively). From the  
132 MultiQC reports, it can be observed that for most of the samples the “per base sequence

133 “content” graph starts with a C or T followed by two Gs, excepting one sample

134 LIB181217CR\_ECL28\_7\_S7\_R1\_001 (FOR2) where the percentage of the T’s more than Cs or Gs

135 for the first three bases (S1 Fig). This discrepancy in the FOR2 sequence can be due to improper

136 MSPI digestion during library preparation, which in turn does not enrich reads that start with

137 CGG or TGG.

138 Next, trimming was performed using Trim Galore [22] to get high quality reads for better

139 methylation calls. It trims all reads having a Phred score less than 20 (i.e., 99% base call

140 accuracy), read length less than 20bp after quality trimming and adapter contamination and/or

141 when reads start with CAA or CGA (S1 File). For all but one sample, sequences removed for

142 quality score criterion were less than 15% of the total number of sequences for that sample and

143 for length criterion; it was less than 5%. For LIB181217CR\_ECL28\_7\_S7\_R1\_001 (FOR2) sample,

144 the sequences removed for quality score criterion were 16.4% and for lengths less than 20 bp

145 were 6.9%. Similarly, for RRBS trimming excluding FOR2, all the samples had RRBS sequences

146 trimmed due to adapter contamination was < 30% and RRBS sequences trimmed due to reads

147 starting with CAA and CGA at 0.1%. For FOR2 sample, the reads trimmed due to adapter

148 contamination were 37%, whereas for the other criteria trimmed reads were 0.2% of all the

149 sequences in the sample (S1 file).

150 Bismark [23] with the Bowtie 2 [24] alignment option was used to align the trimmed sequence

151 to the reference genome (Oar\_rambouillet\_v1.0) and extract the methylation pattern, in the

152 form of cytosine reports for 3 different contexts CpG, CHH and CHG (where H can be A, T or C).

153 Default parameters were used for Bismark and the mapping efficiency was between ~64-69%.

154 The percentage of methylated cytosine in CpG sites in the sample, calculated by dividing

155 number of methylated cytosines and total number of CpG sites, ranged from ~45-64% with  
156 FOR2 having the maximum methylated cytosines in the CpG context. The range of methylation  
157 on C was 0.6-0.8% for all CHH and CHG context (S2 File).

158 Further quality checks, normalizations and differential methylation analyses were performed  
159 using the R Bioconductor package, methylKit [25]. Normalization was performed using the  
160 normalizeCoverage function of methylKit, which normalizes the coverage between samples by  
161 using a scaling factor derived by the difference of median coverages between samples.

162 Differential methylation can be broadly classified into two parts, differentially methylated  
163 cytosines (DMC) and differentially methylated regions (DMR). While DMC looks at differences  
164 in methylated cytosines between two conditions (MOR and FOR in this case), the DMR looks at  
165 methylation differences in two regions (non-overlapping 1000 bases in this case) between the  
166 conditions. The TileMethylCounts function from the methylKit package was used to estimate  
167 the number of methylated Cs in 1000 bases of non-overlapping windows across the whole  
168 genome. To identify the number of CpG contexts in a sample, we used a coverage threshold  
169 between 10X (i.e., at least 10 reads cover that particular CpG context) and 99<sup>th</sup> percentile of the  
170 highest CpG coverage per sample. In addition, at least 3 out of 5 samples were required to pass  
171 the coverage criterion. All the samples were then merged by using the unite function in  
172 methylKit. A further filtration was applied in this step to remove samples that had at least three  
173 replicates having coverage for a CG position. We also merged both the strands to increase the  
174 coverage of CpG, using the destrand=TRUE option.

175 Differential methylation was calculated using calculateDiffMeth and getMethylDiff functions  
176 from the methylKit package. The calculateDiffMeth function calculates the differential

177 methylation using a logistic regression model based on a Chi-square test followed by an  
178 overdispersion correction using the McCullagh and Nelder method [26], and then adjusts the p  
179 value using a Sliding Linear Model (SLIM) [26] multiple test correction method. The  
180 getMethylDiff function was used to extract the significant hypo and hyper DMR and DMC from  
181 the result of calculateDiffMeth function. A false discovery rate (FDR) q value threshold of < 0.1  
182 and methylation difference of  $\pm 10\%$  was used to identify significant DMRs and DMCs.  
183 Annotation of the DMR and DMC was performed using genomation [27]. Codes used for gene  
184 alignment, methylation extraction and differential methylation analysis are in  
185 <https://github.com/VilainLab/SheepMethylation/tree/master/Codes>. The workflow of the  
186 bioinformatics pipeline for the transcriptomic analysis is illustrated in S2A Fig.

## 187 **RNA Library Preparation and Sequencing**

188 Sequencing libraries were prepared using fragmentation, end repair, ligation and PCR using the  
189 Illumina Stranded mRNA Ligation Prep (Illumina, San Diego, CA, USA). Briefly, 1 ug of total RNA  
190 was purified, fragmented and primed with random hexamers to generate first strand  
191 complementary DNA (cDNA) and the first stand cDNA was converted into second strand cDNA.  
192 The 3' ends of the second strand cDNA were subjected to blunt-end repair. In the next step,  
193 pre-index anchors (RNA index anchors) were ligated to the ends of the double-stranded  
194 cDNA fragments to prepare them for dual indexing. A subsequent PCR amplification step  
195 followed to add the index adapter sequences (IDT for Illumina RNA UD Indexes Set A, Ligation  
196 UDP0001-UDP0005). This step selectively amplified the anchor-ligated DNA fragments and  
197 adds indexes and primer sequences for cluster generation. For indexing PCR, initial

198 denaturation was carried out at 98° C for 30 sec, followed by 10 cycles of the following thermal-  
199 cycle profile: denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds, and  
200 extension at 72°C for 30 seconds. A final extension at 72°C for 5 min was followed by a 4°C  
201 hold. The resulting product was a dual-indexed library of DNA fragments with adapters at each  
202 end. The libraries were purified using Agencourt AmPureXP beads (Beckman Coulter) and  
203 eluted in 15 µl of resuspension buffer. Libraries were quantified using the Qubit broad range  
204 assay kit (Thermofisher) and sized using the DNA 1000 kit (Agilent Technologies). The final 300  
205 bp libraries were pooled in equimolar amounts and normalized. The pooled library (1.2 pM)  
206 was sequenced on the Nextseq 550 using the NextSeq 500/550 High Output Kit v2.5 (150  
207 cycles, 2 x 75 bp) and data captured in the Base space sequence Hub (Illumina).

## 208 **RNAseq Analysis**

209 Preprocessing of the fastq files were performed using the method mentioned above. Quality  
210 check was performed using fastQC [18] for single samples and MultiQC [19] for multi sample  
211 summary, followed by quality and adapter trimming by trimmomatic [22]. Next, the fastq was  
212 aligned to the Oar\_rambouillet\_v1.0 from Ensembl, using STAR [28] followed by read  
213 quantification using RSEM[29]. Differential expression analysis was performed using deseq2  
214 [30] with the significance threshold being log2 fold change > 0.58 (1.5 fold change) and log2  
215 fold change < -0.58 (-1.5 fold change), and p-value < 0.1. The workflow of the bioinformatics  
216 pipeline for the transcriptomic analysis is illustrated in S2B Fig.

## 217 **Functional annotation and visualization**

218 Functional annotation was performed using gProfileR [31]. Visualization was done using  
219 methylKit [25] , ViewBS [32], ggplot2 [33] and GOplot [34].

220 **Quantitative PCR method**

221 Total RNA (0.5 µg) was converted to cDNA using the SuperScript™ III First-Strand Synthesis  
222 System (Invitrogen, Waltham, MA, USA) according to the manufacturer's directions. Real time  
223 PCR reactions were run in triplicate using PowerSYBR Green Master Mix (Invitrogen). Primer  
224 sets (S3 Table) for ovine genes were designed specifically to cross exon junctions using Clone  
225 Manager software version 8 (Sci-Ed Software, Westminster, CO, USA). All reactions were run in  
226 a Quant Studio 7 Flex Thermal Cycler (Applied Biosystems, Life Technologies, Eugene, OR, USA).  
227 The primer efficiencies were  $\geq 95\%$  for all primer pairs, and all melting curves showed a single  
228 peak. Quantification of gene expression was performed by the delta delta Ct method, using  
229 cDNA from MBH dissections obtained from four adult Polypay rams as calibrators and  
230 normalized against the reference gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH).  
231 Data are reported as the fold difference relative to the mean for MORs.

232 **Results**

233 **Distinct differential methylation patterns are observed between  
234 MORs and FORs in all the three methylation contexts.**

235 A global methylation analysis of the three contexts (CpG, CHG and CHH) for all the samples,  
236 reveals higher average methylation levels for CpG context compared to the other 2 contexts  
237 (Fig 1 A, S5 – 11 Fig). For each of the three different methylation contexts investigation of

238 differential methylation patterns between MORs and FORs can be broadly classified into two  
239 types: evaluation of differentially methylated cytosines (DMC) and differentially methylated  
240 regions (DMR).

241 **Fig 1. Distinct global CpG methylation pattern observed between contexts and conditions. A)**  
242 Global Methylation levels show higher methylation of CG context: Average Global methylation  
243 levels of the eight samples (5 MORs, 3 FORs). CG methylation level is higher than the other  
244 contexts for all the samples. B) Heatmap of DMR methylation ratios for reads in the CpG  
245 context plotted against animal number shows differential methylation patterns between MORs  
246 and FORs. Higher value red, lower value blue.

247 The range of CpG context sites, that passes both the coverage, and the sample threshold  
248 criterion are between ~985,376 to ~1,201,527 for all samples excepting FOR2, for which the  
249 number of CpG context site is 250,828 (Table 3). Hierarchical clustering of the average  
250 methylation profile for CpG context revealed that FOR2 was not clustered with the other  
251 samples and is an outlier (S4A Fig). The same pattern was observed in the other contexts (CHG  
252 and CHH), with FOR2 being an outlier in both the scenarios (S4B and C Fig). This makes the  
253 FOR2 sample an outlier and it was removed from further downstream DMR and DMC analysis,  
254 for all the three contexts.

255 **Table 3: Methylation sites per context**

Sample Name	Phenotype	CpG Region	CHG region	CHH region
LIB181217CR_ECL28_1_S1	MOR1	1083643	1483439	2704253
LIB181217CR_ECL28_2_S2	MOR2	1184415	1618708	3039998
LIB181217CR_ECL28_3_S3	MOR3	1111216	1596032	3048524
LIB181217CR_ECL28_4_S4	MOR4	1201527	1652283	3058649

LIB181217CR_ECL28_5_S5	MOR5	1162673	1612379	3019239
LIB181217CR_ECL28_6_S6	FOR1	1197471	1697854	3275614
LIB181217CR_ECL28_7_S7	FOR2	250828	460853	644272
LIB181217CR_ECL28_8_S8	FOR3	1216055	1703738	3241296
LIB181217CR_ECL28_9_S9	FOR4	985376	1367634	2531456

256

257 DNA methylation has various functions, and methylation can occur in different locations. We  
258 evaluated 656,897 filtered CpG context sites and identified 1552 DMCs of which 803 were  
259 hypomethylated and 749 were hypermethylated in MORs compared to FORs. Of all the  
260 differentially methylated cytosines, 44% are located in the intergenic regions, followed by 37%  
261 in introns, 11% in exons and 8% in promoters (Fig 2A). A similar distribution pattern of  
262 methylated DMCs was observed for hypo- and hypermethylated CpG regions, with the majority  
263 of the DMC's located in intergenic regions and the least in the promoter region (Fig 2B and C, S4  
264 File DMC\_CpG tab, DMC\_hyper\_CpG tab and DMC\_hypo\_CpG tab).

265 **Fig 2. Distribution of total CpG DMC and DMR across the different genomic regions.** A) Distribution of  
266 DMC CpG; B) Distribution of hyper DMC CpG; C) Distribution of hypo DMC CpG; D) Distribution of DMR  
267 CpG; E) Distribution of hyper DMR CpG; F) Distribution of hypo DMR CpG; Distribution of hypo DMR  
268 CpG. Legend: black = promoter region; pink = exon; green = intron and blue = intergenic regions.

269 For evaluation of DMRs, we looked at non-overlapping 1000 base pair regions and identified  
270 805 DMRs of which 478 were hypomethylated and 327 were hypermethylated in MORs  
271 compared to FORs. Visualization of all DMRs shows distinct differential patterns between the  
272 two phenotypes i.e., MORs and FORs (Fig 1B). The distribution of DMRs is similar to that of  
273 DMCs across genomic regions with the maximum (46%) falling in the intergenic regions, 33% in  
274 introns, 16% in exons and 5% in the promoter regions (Fig 2D). Likewise, the distribution of

275 hypo- and hypermethylated DMRs and DMCs are similar across genomic regions (Fig 2E and  
276 F,S4 File DMR\_CpG tab, DMR\_hyper\_CpG tab and DMR\_hypo\_CpG tab).  
  
277 Although previous studies have not been conclusive about the function of non-CpG (CHG and  
278 CHH) methylations in mammals, they have been observed previously in developing mouse brain  
279 [29]. This enabled us to explore the methylation profile in these two contexts. For the CHG  
280 context there are 25 DMCs with 10 hypermethylated and 15 hypomethylated (S5 File  
281 DMC\_CHG tab, DMC\_hyper\_CHG tab and DMC\_hypo\_CHG tab), and 16 regions for DMR, with  
282 nine hypermethylated and seven hypomethylated. Out of the 25 DMCs, 60% are in the intronic  
283 and 40% in intergenic regions (Fig 3A). For the hypermethylated CHGs, 80% are in the intronic  
284 and 20% are intergenic regions (Fig 3B), whereas 47% of hypomethylated CHGs fall in intronic  
285 and 53% fall in intergenic regions (Fig 3C). For DMR CHGs, 62% fall in intergenic regions, 25% in  
286 introns and 12% in exons (Fig 3D). For hypermethylated CHGs, most DMRs (56%) fall in  
287 intergenic regions, whereas 33% fall in introns and 11% fall in the exons (Fig 3E). For  
288 hypomethylated CHGs, 71% of DMRs fall in intergenic regions, while equal distribution (14%)  
289 falls in exons and introns (Fig 3 F, S5 File DMR\_CHG tab, DMR\_hyper\_CHG tab and  
290 DMR\_hypo\_CHG tab).

291 **Fig 3. Distribution of CHG DMC and DMR across the different genomic regions.** A) Distribution of DMC  
292 CHG; B) Distribution of hyper DMC CHG; C) Distribution of hypo DMC CHG; D) Distribution of DMR CHG;  
293 E) Distribution of hyper DMR CHG; F) Distribution of hypo DMR CHG; Legend: black = promoter region;  
294 pink = exon; green = intron and blue = intergenic regions.

295 For CHH context, there are 15 DMC regions with 7 hypomethylated and 8 hypermethylated,  
296 and 56 DMR regions with 17 hypomethylated and 39 hypermethylated. In the case of CHHs

297 DMCs, the pattern is similar to DMC distribution in CpG context. Out of 15 DMCs, most (60%)  
298 fall in intergenic regions, 33% in introns and 7% in promoter regions (Fig 4A). For  
299 hypermethylated CHHs, most DMCs fall in intergenic regions (50%), 38% in introns and 12% in  
300 exons (Fig 4B). For the hypomethylated CHHs, 71% of the DMCs fall in the intergenic regions  
301 and 29% in introns (Fig 4C, S6 File DMC\_CHH tab, DMC\_hyper\_CHH tab and DMC\_hypo\_CHH  
302 tab). The distribution of DMR CHHs follows a similar pattern as DMC, with most falling in  
303 intergenic regions (46%), followed by 34% in introns, 14% in exons and 5% in the promoter  
304 regions (Fig 4D). Hypomethylated DMRs follow the same pattern as the distribution of all CHH  
305 DMRs (most are located in the intergenic regions and the fewest in the promoter region). For  
306 hypermethylated DMRs, the same percentage of DMRs fall in the intronic and intergenic  
307 regions (38%), followed by exon (16%) and promoter (8%) regions (Fig 4E and F, S6 File  
308 DMR\_CHH tab, DMR\_hyper\_CHH tab and DMR\_hypo\_CHH tab).

309 **Fig 4. Distribution of CHH DMC and DMR across the different genomic regions.** A) Distribution  
310 of DMC CHH; B) Distribution of hyper DMC CHH; C) Distribution of hypo DMC CHH; D)  
311 Distribution of DMR CHH; E) Distribution of hyper DMR CHH; F) Distribution of hypo DMR CHH;  
312 Legend: black = promoter region; pink = exon; green = intron and blue = intergenic regions.

313 **Functional annotation of the individual methylation contexts reveals  
314 distinct functional clusters.**

315 To identify functionally relevant genes overlapping DMC/DMRs, we performed functional  
316 annotation using gProfileR. We only chose genes that had DMC/DMRs in their gene body (i.e.,  
317 exons and introns) or promoters, and left out intergenic DMC/DMRs from further analysis.

318 Functional annotation for DMC CpG (hyper- and hypomethylated cytosines) context produced  
319 28 significantly enriched functional clusters, adjusted p-value < 0.1 (Fig 5A, S4 File  
320 DMC\_CpG\_GO tab). The significantly enriched functional terms are all gene ontology (GO)  
321 terms with 14 of them pertaining to biological processes (BP), eight to molecular functions (MF)  
322 and six to cellular components (CC). The BP GO clusters contain mainly developmental and  
323 biological regulation processes, whereas the MF terms pertain to protein binding and  
324 electrophysiological activities, while the CC pathways includes membrane and cytoplasm  
325 related terms. The hypermethylated DMCs yielded ten significant GO terms, with seven MF  
326 terms comprising electrophysiological and protein binding activities while three CC terms are  
327 related to membrane and cation channel complex terms. The hypomethylated DMCs yielded 12  
328 significant functionally relevant terms, with seven BP terms pertaining to biological regulation  
329 and response to wound, two MF terms associated with protein binding and three CC terms  
330 related to cytoplasm, cell periphery and Schaffer collateral - CA1 synapse (S4 File DMC\_  
331 hyper\_CpG\_GO and DMC\_hypo\_CpG\_GO tab).

332 Functional annotation of DMRs for CpG regions revealed nine significant enriched functional  
333 terms (Fig 5A, S4 File DMR\_CpG\_GO tab). Of the nine functional terms, five are CC functions  
334 related to synapse and cell periphery, three are BP functions related to central nervous system  
335 neuron development, activation of GTPase activity and movement of cell or subcellular  
336 component), and one is a MF function associated with calcium ion binding. The  
337 hypermethylated DMRs, yielded no significantly enriched terms, whereas the hypomethylated  
338 DMRs yielded five terms with three BP functions related to regulation of GTPase activity and  
339 chemorepulsion of axons and two MF functions associated with GTPase regulator activity and

340 nucleoside-triphosphatase regulator activity (S4 File DMR\_ hyper\_CpG\_GO and DMR\_ hypo\_  
341 CpG\_GO tab).

342 **Figure 5. Functional annotation for DMC and DMR for CpG context.** A) Representation of gene  
343 distribution across functional annotation terms for DMC CpG. The distribution of  
344 hypermethylated (red) and hypomethylated (blue) terms for each of the functional terms is  
345 represented in each quadrant. B) Representation of gene distribution across functional  
346 annotation terms for DMR CpG. The distribution of hypermethylated (red) and hypomethylated  
347 (blue) terms for each of the functional terms is represented in each quadrant. Enrichment of  
348 each term is reported as a z-score, where z-score is the ratio of difference between number of  
349 hyper methylated and hypomethylated DMC genes, and square root of total number of genes  
350 for that term.

351 The non-CpG methylation yielded fewer functionally relevant terms, compared to the CpG  
352 context. The functional annotation for DMCs in CHG context (both hyper- and hypomethylated  
353 combined) yielded no significantly enriched terms. Hypermethylated DMCs yielded two  
354 significant functionally enriched terms: one human phenotype (unilateral radial aplasia) and  
355 one CC function related to mitochondrial pyruvate dehydrogenase complex. Hypomethylated  
356 DMCs yielded only one MF function associated with phosphomevalonate kinase activity (S5 File  
357 DMC\_CHG\_GO tab, DMC\_hypo\_CHG\_GO tab and DMC\_hyper\_CHG\_GO tab). The CHG DMRs  
358 yielded no significantly enriched functional annotation terms. A similar pattern was observed  
359 for functional annotation of DMC/DMR for the CHH context. Only one enriched term was  
360 observed for DMCs in the CHH context and it was associated with the MF phosphomevalonate  
361 kinase activity. Hypermethylated DMCs yielded three BP functions linked with regulation of

362 clathrin coat assembly and gastric acid secretion. Hypermethylated DMCs yielded one MF term  
363 linked to phosphomevalonate kinase activity. Like DMRs in CHG context, none of the DMRs in  
364 CHH context yielded any relevant functional terms (S6 File DMC\_CHH\_GO tab,  
365 DMC\_hypo\_CHH\_GO tab, DMC\_hyper\_CHH\_GO tab).

366 **Overlap of regions across the three methylation contexts show  
367 distinct functional features.**

368 To understand the effect of the different methylation contexts (CpG, CHH and CHG), we  
369 investigated the DMRs that overlap for the multiple contexts. We identified three genes  
370 common between all three contexts, one common between CpG and CHG, five common  
371 between CpG and CHH and six common between CHH and CHG (Fig 6A; S7 File). The genes  
372 common between the three contexts are *ENSOARG00020023439*, *TPGS2* and *SCNN1B*. The  
373 DMR was in an intergenic region near *ENSOARG00020023439* (spindlin-2B homologue in sheep);  
374 DMR coordinates: chromosome X- 50,554,001-50,555,000; intergenic near the gene) and was  
375 hypomethylated (methylation difference MD = -35.6%, corrected P = 0.007) in MOR compared  
376 to FOR in the CpG context, whereas it was hypermethylated in the CHG (MD = 16.36%, P = 0.08)  
377 and CHH context (MD = 16.6%, P = 0.05). For *TPGS2* (tubulin polyglutamylase complex subunit;  
378 chromosome 23: 24,612,001-24,613,000), the DMR was in the intron regions and was  
379 hypermethylated in MOR compared to FOR, in all the three contexts (CpG: MD = 30.15%, P =  
380 0.07; CHG: MD = 10.63%, P = 0.06; CHH: MD = 11.14%, P = 0.005). A similar pattern was  
381 observed for the gene *SCNN1B* (sodium channel epithelial one subunit beta, chromosome 24:  
382 21,728,001- 21,729,000), with the DMR occurring in exon and intron regions, and

383 hypermethylated for all the three contexts (CpG: MD = 14.4%, P = 0.08; CHG: MD = 15.75%, P =  
384 0.03; CHH: MD = 15.86%, P = 0.0001; S7 File). The three genes also have distinct functional  
385 features. While spindlin-2B (human homologue of *ENSOARG00020023439*), is involved in  
386 regulation of cell cycle progression [35] and H3K4me3-binding activity [36], *TPGS2* codes for a  
387 protein component of neuronal polyglutamylase complex [37], whereas *SCNN1B* is responsible  
388 for sodium channel activity and mutation of the gene leads to autosomal disorders like Liddle  
389 syndrome [38].

390 **Fig 6. DMR genes common between the three contexts.** DMR genes common between the  
391 three contexts. Venn diagram depicting the genes that are shared among CHH (green),  
392 CpG(purple) and CHG (yellow), in the DMC context. There are three genes in common among  
393 the 3 contexts, one in common between CpG and CHG, five in common between CpG and CHH,  
394 and six in common between CHG and CHH.

395 The DMR for the unannotated gene *ENSOARG00020011386* (DMR coordinates: chromosome  
396 18: 66,982,001-66,983,000) common between CHG and CpG contexts was in the intergenic  
397 region near the gene and was hypomethylated in MORs for both the contexts (S7 File). For the  
398 six genes common between CHH and CHG, the DMR for four genes (*ENSOARG00020000663*,  
399 *EPCAM*, *ADAMTS15*, and *PLXND1*) were in the intergenic region, whereas for the other two  
400 genes (*MAGI1*, *TVP23A*) the DMR was in the intronic region. All the genes except one (*PLXND1*),  
401 was hypermethylated in MORs compared to FORs, in the two contexts. Functional annotation  
402 of the genes revealed three significantly enriched CC functional terms related to cellular  
403 junctions. DMRs for the genes common between CHH and CpG, overlap the gene body with  
404 four genes (*U6*, *GSE1*, *MIR153-2* and *AGPAT4*), having DMRs in the intron, whereas for *CARD11*,

405 DMR overlaps both exon and intron. There were only two significant functional annotation  
406 terms, one CC (CBM complex) and one HP (decreased specific antibody response to  
407 polysaccharide vaccine) associated with these genes (S6 File).

408 **Differential expression analysis reveals significantly expressed genes  
409 associated with sexual partner preference.**

410 To investigate the relationship between DNA methylation changes and gene expression, RNA-  
411 Seq analysis was performed to identify differences in gene expression between the two  
412 phenotypes. A total of 15 differentially expressed genes were detected between phenotypes,  
413 with only one gene overlapping with the DMR gene lists and none with the DMC gene lists (Fig  
414 7A). The gene *BFSP1* ( $\log_2 \text{FC} = 1.15$ ,  $\text{qvalue} = 0.002$ ), was hypomethylated in CpG DMR context,  
415 with a methylation difference of -11.6% MOR vs. FOR ( $\text{q value} = 0.02$ ; S8 Table or File).

416 **Fig 7. Differential gene expression associated with sexual partner preference phenotype. A)**  
417 Heatmap of differentially expressed genes plotted against animal number and grouped by phenotype,  
418 i.e., FOR or MOR. RNAseq analysis identified 15 genes that were differentially expressed between FORs  
419 and MORs (adjusted p-value < 0.1,  $\log_2 \text{Fold change} \geq \text{absolute (0.58)}$ ). Heatmap colors are represented  
420 by Z-score and annotation of ram phenotype has blue for FORs and red for MORs. Go pathway analysis  
421 identified enrichment of three differentially expressed genes involved in hormone activity: prolactin  
422 (*PRL*); MOR vs. FOR  $\log_2$  fold difference ( $\log_2 \text{FD} = -4.5$ ,  $P = 1.8\text{E-07}$ ), cholecystokinin (*CCK*);  $\log_2 \text{FD} = -$   
423 1.2,  $P = 5.09\text{E-05}$  and neurotensin (*NTS*);  $\log_2 \text{FD} = 1.4$ ,  $P = 8.40\text{E-06}$ . Differential gene expression was  
424 confirmed using qPCR for: (B) *PRL* ( $\log_2 \text{FD} = -4.2$ ,  $P = 1.8\text{E-02}$ ), (C) *CCK* ( $\log_2 \text{FD} = -1.37$ ,  $P = 0.13$ ) and (D)  
425 *NTS* ( $\log_2 \text{FD} = 0.67$ ,  $P = 0.24$ ). Data (mean  $\pm$  SEM) were analyzed by a Student's t test.

426 To explore further the effect of the differentially expressed genes, we performed functional  
427 annotation. GO pathway analysis identified enrichment of three genes involved in hormone  
428 activity (MF): prolactin (*PRL*); MOR vs. FOR log<sub>2</sub> fold difference (log<sub>2</sub> FD) = -4.5, Fold change = -  
429 0.04, P = 1.8E-07), cholecystokinin (*CCK*); log<sub>2</sub> FD = -1.2, Fold change = -0.43, P = 5.09E-05 and  
430 neurotensin (*NTS*); log<sub>2</sub> FD = 1.4, Fold Change = 2.639016, P = 8.40E-06 (S8 File). To confirm the  
431 differences in gene expression between MORs and FORs identified with RNAseq, we performed  
432 quantitative real-time PCR. We observed down regulation of *PRL* in MORs vs. FORs (log<sub>2</sub> FD = -  
433 4.2, Fold Change = - 0.054, P = 0.001) and *CCK* (log<sub>2</sub> FD = -1.3, Fold Change = -0.406, P = 0.08)  
434 and up regulation of NTS (log<sub>2</sub> FD = 0.67, Fold Change =1.59, P = 0.24) in MORs compared to  
435 FORS, which is in accordance with what was seen in the RNA-Seq analysis (Fig 7B and C).

436 **Discussion**

437 In the present study, genome-wide DNA methylation in hypothalami of rams exhibiting  
438 exclusive male versus female sexual partner preferences were analyzed for the first time. Out  
439 of the three methylation contexts, CpG, CHG and CHH, the most significant differences were  
440 observed in the CpG context with 1552 DMC and 805 DMRs being significantly methylated.  
441 There were more hypomethylated CpGs in MORs compared to FORs for both the DMC and  
442 DMR groups. The distribution in the case for DMCs was ~52% hypomethylated and ~48%  
443 hypermethylated, whereas for DMRs the distribution was 60% hypomethylated compared to  
444 40% hypermethylated. Functional annotation of the differentially methylated genes that fall in  
445 the DMC or DMR regions revealed that most of the significant functional terms were related to  
446 developmental processes, regulatory and electrophysiological activities that may be associated

447 with the many homeostatic functions of the hypothalamus. Functional terms associated with  
448 development of sexual characteristics and sex development were also identified but none of  
449 them was differentially enriched.

450 CpG is considered the most relevant context because 80% of methylation events in humans  
451 occur at CpG sites [42]. However, we also evaluated the CHG and CHH contexts because they  
452 have been previously associated with brain development [43]. Moreover, CHH methylation is  
453 highly conserved in the brain across vertebrate species and requires active maintenance in  
454 postmitotic neurons [44]. We observed a pattern similar to previous studies [42,44], with fewer  
455 significantly methylated DMCs and DMRs in the non-CpG (CHH and CHG) context, compared to  
456 the CpG context. The DMCs for the CHG context followed the same pattern as for CpG, with  
457 more hypomethylated than hypermethylated genes. In contrast, the DMRs for the CHG context,  
458 and both DMC and DMR for the CHH contexts, exhibited more hypermethylated than  
459 hypomethylated genes. There were only a few significant functional terms in both the  
460 contexts, and most of them were related to molecular functions such as phosphomevalonate  
461 kinase activity or biological processes pertaining to regulation of clathrin coat assembly and  
462 regulation of gastric acid secretion. There were genes in common among all three different  
463 contexts. Most of them were associated with molecular functions and cellular component  
464 functionalities and none was associated with sexual behaviors, neuroendocrine functions or  
465 development.

466 Transcriptomic analysis revealed 15 differentially expressed genes between the two  
467 phenotypes with only one overlapping with the methylated list. The gene, Beaded Filament  
468 Structural Protein 1 (*BFSP1*) was hypomethylated in CpG DMR context and overexpressed in

469 MORs compared to FORs. This gene shows broad expression in a number of tissues including  
470 brain, and has been previously associated with cataracts in humans [39–41]. Additionally,  
471 functional annotation of the differentially expressed genes reveal one significant term  
472 associated with hormone activity (MF) and consisting of three genes PRL, CCK and NTS.  
473 Prolactin (PRL) is a hormone produced mainly by the pituitary gland, however, in some species  
474 it is synthesized in other tissues including brain [45,46] . PRL is best known for its role in the  
475 development of the mammary gland and milk production, but is also involved in the regulation  
476 of parental and sexual behaviors in both males and females [47,48]. The neuropeptide  
477 cholecystokinin (CCK) has been associated with mate preference in mice. CCK-expressing  
478 neurons in the bed nucleus of the stria terminalis of males are activated by the scent of female  
479 urine in association with the male's preference for estrus females [49]. Finally, neuropeptins  
480 (NTS) neurons in the medial preoptic area were shown to encode attractive male cues and  
481 direct behavior toward opposite-sex conspecifics in both sexes to drive social attraction toward  
482 a potential mate [50]. Quantitative PCR validations show, that PRL is the only gene that was  
483 significantly downregulated in MORs compared to FORs, which agrees with the RNASeq results.  
484 Although, neither CCK nor NTS showed significant fold differences with quantitative PCR, they  
485 show similar trends with RNA-Seq results, i.e., CCK downregulated and NTS upregulated in  
486 MORs compared to FORs.

487 To our knowledge, our study presents the first genome-wide analysis of DNA methylation  
488 profiles and gene expression of the adult sheep hypothalamus. We show that the epigenome  
489 of the hypothalamus, in the form of DNA methylation pattern, differs substantially between  
490 rams with different sexual partner preferences. This tentatively suggests that epigenetic factors

491 may be important mechanisms involved in sexual attraction. Specifically, we highlight  
492 expression differences in genes related to sexual behaviors. These data will be informative in  
493 providing a basis for better understanding of the epigenetic regulation of sexual behavior in  
494 sheep and help ascertain mechanisms that shape sexual partner preferences. However, further  
495 studies will be required to determine whether differences in DNA methylation and consequent  
496 gene expression are the cause or consequence of altered behavior. In addition, experiments  
497 should be conducted at earlier developmental landmarks are needed to capture the effects  
498 more efficiently. Finally, sample size is often a challenge with large animal models such as the  
499 sheep. This study is no exception and would benefit from a replication with more animals. Thus,  
500 further transcriptomic and epigenetic studies need to be performed with a larger sample size to  
501 ascertain the developmental effect that the epigenome/transcriptome has on the expression of  
502 sexual partner preferences in rams.

503 **Acknowledgements:** The reduced-representation bisulfite sequencing libraries were generated  
504 by the Knight Cardiovascular Research Institute Epigenetics Consortium at OHSU. The authors  
505 wish to thank Dr. Lucia Carbone for her oversight of the bisulfite sequencing and helpful  
506 comments on the manuscript. The authors would like to thank Dr. Susan Knoblach and Karuna  
507 Panchapakesan, for their help with the RNA sequencing. This work was supported by National  
508 Institutes Health grants R01OD011047 and OHSU SOM Innovation Award to C.E.R.

509 **References**

510 1. Bailey JM, Vasey PL, Diamond LM, Breedlove SM, Vilain E, Epprecht M. Sexual orientation,  
511 controversy, and science. *Psychol Sci Public Interest*. 2016;17: 45–101.  
512 doi:10.1177/1529100616637616

513 2. Forger NG. Past, present and future of epigenetics in brain sexual differentiation. *J*  
514 *Neuroendocrinol.* 2018;30. doi:10.1111/jne.12492

515 3. Ngun TC, Vilain E. The biological basis of human sexual orientation: Is there a role for  
516 epigenetics? 86th ed. In: Yamamoto D, editor. *Advances in Genetics:Epigenetic Shaping of*  
517 *Sociosexual InteractionsFrom Plants to Humans.* 86th ed. New York: Academic Press; 2014. pp.  
518 167–184. doi:<http://dx.doi.org/10.1016/B978-0-12-800222-3.00008-5>

519 4. Ngun TC, Ghahramani N, Sánchez FJ, Bocklandt S, Vilain E. The genetics of sex differences in  
520 brain and behavior. *Front Neuroendocrinol.* 2011;32: 227–46. doi:10.1016/j.yfrne.2010.10.001

521 5. Nugent BM, Wright CL, Shetty AC, Hodes GE, Lenz KM, Mahurkar A, et al. Brain feminization  
522 requires active repression of masculinization via DNA methylation. *Nat Neurosci.* 2015;18: 690–  
523 7. doi:10.1038/nn.3988

524 6. Cisternas CD, Cortes LR, Golyntsev I, Castillo-Ruiz A, Forger NG. Neonatal inhibition of DNA  
525 methylation disrupts testosterone-dependent masculinization of neurochemical phenotype.  
526 *Endocrinology.* 2020;161. doi:10.1210/ENDOCR/BQZ022

527 7. Roselli C. Programmed for Preference: The biology of same-sex attraction in rams. *Neurosci*  
528 *Biobehav Rev.* 2020;114. doi:10.1016/J.NEUBIOREV.2020.03.032

529 8. LeVay S. A difference in hypothalamic structure between heterosexual and homosexual men.  
530 *Science.* 1991;253: 1034–7. doi:10.1126/science.1887219

531 9. Roselli CE, Larkin K, Resko JA, Stellflug JN, Stormshak F. The volume of a sexually dimorphic  
532 nucleus in the ovine medial preoptic area/anterior hypothalamus varies with sexual partner  
533 preference. *Endocrinology.* 2004;145: 478–83. doi:10.1210/en.2003-1098

534 10. Roselli CE, Estill CT, Stadelman HL, Meaker M, Stormshak F. Separate critical periods exist  
535 for testosterone-induced differentiation of the brain and genitals in sheep. *Endocrinology.*  
536 2011;152: 2409–15. doi:10.1210/en.2010-1445

537 11. Ghahramani NM, Ngun TC, Chen P-Y, Tian Y, Krishnan S, Muir S, et al. The effects of  
538 perinatal testosterone exposure on the DNA methylome of the mouse brain are late-emerging.  
539 *Biol Sex Differ.* 2014;5: 8. doi:10.1186/2042-6410-5-8

540 12. Saper CB, Lowell BB. The hypothalamus. *Curr Biol.* 2014;24: R1111–R1116.  
541 doi:10.1016/J.CUB.2014.10.023

542 13. Matsumoto A, Arai Y. Sex difference in volume of the ventromedial nucleus of the  
543 hypothalamus in the rat. *Endocrinol Jpn.* 1983;30: 277–80. doi:10.1507/endocrj1954.30.277

544 14. Schwarz JM, Liang S-L, Thompson SM, McCarthy MM. Estradiol induces hypothalamic  
545 dendritic spines by enhancing glutamate release: a mechanism for organizational sex  
546 differences. *Neuron.* 2008;58: 584–98. doi:10.1016/j.neuron.2008.03.008

547 15. Yang CF, Chiang MC, Gray DC, Prabhakaran M, Alvarado M, Juntti SA, et al. Sexually  
548 dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and  
549 aggression in males. *Cell.* 2013;153: 896–909. doi:10.1016/J.CELL.2013.04.017

550 16. Lee H, Kim DW, Remedios R, Anthony TE, Chang A, Madisen L, et al. Scalable control of  
551 mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. *Nat* 2014;509:5097502.  
552 2014;509: 627–632. doi:10.1038/nature13169

553 17. Perkins A, Roselli CE. The Ram as a Model for Behavioral Neuroendocrinology. *Horm Behav.*  
554 2007;52: 70. doi:10.1016/j.yhbeh.2007.03.016

555 18. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].  
556 2010. Available: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

557 19. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple  
558 tools and samples in a single report. *Bioinformatics*. 2016;32: 3047–3048.  
559 doi:10.1093/bioinformatics/btw354

560 20. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using  
561 phred. I. Accuracy assessment. *Genome Res.* 1998;8: 175–85. doi:10.1101/gr.8.3.175

562 21. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error  
563 probabilities. *Genome Res.* 1998;8: 186–94. doi:10.1101/GR.8.3.186

564 22. Kruger F. Trimgalore. 2021. doi:10.5281/zenodo.5127899

565 23. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq  
566 applications. *Bioinformatics*. 2011;27: 1571–1572. doi:10.1093/bioinformatics/btr167

567 24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:  
568 357–9. doi:10.1038/nmeth.1923

569 25. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al.  
570 methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation  
571 profiles. *Genome Biol.* 2012;13: R87. doi:10.1186/gb-2012-13-10-r87

572 26. McCullagh P (Peter), Nelder JA. Generalized linear models. Chapman and Hall; 1989.

573 27. Akalin A, Franke V, Vlahovićek K, Mason CE, Schubeler D. genomation: a toolkit to  
574 summarize, annotate and visualize genomic intervals. *Bioinformatics*. 2015;31: 1127–1129.  
575 doi:10.1093/bioinformatics/btu775

576 28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal  
577 RNA-seq aligner. *Bioinformatics*. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635

578 30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-  
579 seq data with DESeq2. *Genome Biol.* 2014;15: 1–21. doi:10.1186/S13059-014-0550-  
580 8/FIGURES/9

581 31. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server  
582 for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids  
583 Res.* 2019;47: W191–W198. doi:10.1093/nar/gkz369

584 32. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for  
585 visualization of high-throughput bisulfite sequencing data. *Bioinformatics*. 2018;34: 708–709.  
586 doi:10.1093/bioinformatics/btx633

587 33. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016.  
588 Available: <https://ggplot2.tidyverse.org>

589 34. Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining  
590 expression data with functional analysis. *Bioinformatics*. 2015;31: 2912–4.  
591 doi:10.1093/bioinformatics/btv300

592 35. Fletcher BS, Dragstedt C, Notterpek L, Nolan GP. Functional cloning of SPIN-2, a nuclear anti-  
593 apoptotic protein with roles in cell cycle progression. *Leukemia*. 2002;16: 1507–1518.  
594 doi:10.1038/SJ.LEU.2402557

595 36. Bae N, Gao M, Li X, Premkumar T, Sbardella G, Chen J, et al. A transcriptional coregulator,  
596 SPIN-DOC, attenuates the coactivator activity of Spindlin1. *J Biol Chem*. 2017;292: 20808–  
597 20817. doi:10.1074/JBC.M117.814913

598 37. NCBI. TPGS2 tubulin polyglutamylase complex subunit 2 [Homo sapiens (human)]. Available:  
599 <https://www.ncbi.nlm.nih.gov/gene/25941>

600 38. Liddle GW, Bledsoe T, Coppage WS. A familial renal disorder simulating primary  
601 aldosteronism but with negligible aldosterone secretion. *Trans Assoc Am Physe*. 1963;76: 199–  
602 213.

603 39. Zhai Y, Li J, Yu W, Zhu S, Yu Y, Wu M, et al. Targeted exome sequencing of congenital  
604 cataracts related genes: broadening the mutation spectrum and genotype-phenotype  
605 correlations in 27 Chinese Han Families. *Sci Rep*. 2017;7. doi:10.1038/S41598-017-01182-9

606 40. Wang H, Zhang T, Wu D, Zhang J. A novel beaded filament structural protein 1 (BFSP1) gene  
607 mutation associated with autosomal dominant congenital cataract in a Chinese family. *Mol Vis*.  
608 2013;19: 2590. Available: [/pmc/articles/PMC3874047/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3874047/)

609 41. Ramachandran RD, Perumalsamy V, Hejtmancik JF. Autosomal recessive juvenile onset  
610 cataract associated with mutation in BFSP1. *Hum Genet*. 2007;121: 475–482.  
611 doi:10.1007/S00439-006-0319-6

612 42. Tirado-Magallanes R, Rebbani K, Lim R, Pradhan S, Benoukraf T. Whole genome DNA  
613 methylation: Beyond genes silencing. *Oncotarget*. 2017. doi:10.18632/oncotarget.13562

614 43. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic  
615 reconfiguration during mammalian brain development. *Science*. 2013;341.  
616 doi:10.1126/SCIENCE.1237905

617 44. Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. Distribution, recognition and regulation of  
618 non-CpG methylation in the adult mammalian brain. *Nat Neurosci*. 2014;17: 215.  
619 doi:10.1038/NN.3607

620 45. Cabrera-Reyes EA, Limón-Morales O, Rivero-Segura NA, Camacho-Arroyo I, Cerbón M.  
621 Prolactin function and putative expression in the brain. *Endocr* 2017 572. 2017;57: 199–213.  
622 doi:10.1007/S12020-017-1346-X

623 46. Roselli CE, Bocklandt S, Stadelman HL, Wadsworth T, Vilain E, Stormshak F. Prolactin  
624 Expression in the Sheep Brain. *Neuroendocrinology*. 2008;87: 206–215. doi:10.1159/000114643

625 47. Larsen CM, Grattan DR. Prolactin, neurogenesis, and maternal behaviors. In: *Brain,*  
626 *Behavior, and Immunity [Internet]*. 2012 [cited 17 Dec 2021] pp. 201–209.  
627 doi:10.1016/j.bbi.2011.07.233

628 48. Salais-López H, Agustín-Pavón C, Lanuza E, Martínez-García F. The maternal hormone in the  
629 male brain: Sexually dimorphic distribution of prolactin signalling in the mouse brain. *PLoS One*.  
630 2018;13: e0208960. doi:10.1371/JOURNAL.PONE.0208960

631 49. Giardino WJ, Eban-Rothschild A, Christoffel DJ, Li S Bin, Malenka RC, de Lecea L. Parallel  
632 circuits from the bed nuclei of stria terminalis to the lateral hypothalamus drive opposing  
633 emotional states *Nat Neurosci*. 2018;21: 1084. doi:10.1038/S41593-018-0198-X

634 50. McHenry JA, Otis JM, Rossi MA, Robinson JE, Kosyk O, Miller NW, et al. Hormonal gain  
635 control of a medial preoptic area social reward circuit. *Nat Neurosci*. 2017;20: 449.  
636 doi:10.1038/NN.4487

637 **Supporting information**

638 **S1 Fig. Per base sequence content graph shows discrepancy in starting bases of the sequence**  
639 **in FOR7 compared to other samples.** A) Sequence content across all bases for FOR1: Graph  
640 showing the representation of the nucleotides across all base position in sample FOR1.  
641 Percentage of C (marked in blue) is highest in the first base followed by G (in black) in the next  
642 two positions. Similar pattern was observed in all other samples, excepting FOR2. A) Sequence  
643 content across all bases for FOR2: Graph showing the representation of the nucleotides across  
644 all base position in sample FOR2. Percentage of T (marked in red) is highest in the first three  
645 base. Color code Thymine (T) =red, Adenosine = Green, Cytosine = Dark Blue, Guanine = Black.

646 **S2 Fig. Workflow of Methylation analysis Pipeline.** Quality Check using fastqc and trimgalore  
647 was used to trim reads less than 20 Phred score. Alignment and methylation count was  
648 calculated using Bismark, followed by methylKit to estimate the differentially methylated  
649 regions (DMRs) and differentially methylated cytosines (DMCs). Methylation fold change  
650 greater than 10; and q value < 0.01 was used for determining the most significant Genes.  
651 Annotation of the DMR and DMC was done using genomation. Functional annotation was  
652 performed using gProfiler functional annotation tool; followed by visualization using ViewBS for  
653 heatmaps, methylKit for dendrogram and distribution of genomic regions for DMRs and GOPlot  
654 for gene ontology visualization.

655 **S3 Fig. Quality of the sample reveals, Sample 2810 has more reads than the other samples. A)**  
656 Raw read counts from fastq: The plot of the sequence counts shows that for the trimmed  
657 sample 2810, the number of reads is greater than 175 million reads, whereas for the other  
658 samples has 30 to 75 million reads. We can also observe that the number of duplicate reads  
659 (black) in this sample is also greater than the other samples. C) Align read counts from STAR:  
660 Aligned read counts from STAR show that sample 2810 has more unmapped reads (red), and  
661 least uniquely mapped reads (dark blue) than any of the other samples.

662 **S4 Fig. Hierarchical clustering of sample methylation patterns across the 3 contexts. A)** CpG  
663 hierarchical Clustering: Hierarchical clustering of the methylation pattern of replicates of all the  
664 samples, in CpG context) CHG hierarchical Clustering: Hierarchical clustering of the methylation  
665 pattern of replicates of all the samples, in CHG context. C) CpG hierarchical Clustering:  
666 Hierarchical clustering of the methylation pattern of replicates of all the samples, in CHH  
667 context.

668 **S5 Fig. Methylation pattern of Chromosome 1 and chromosome 2 in CG, CHG and CHH**  
669 **context:** Average methylation levels of the different context between the 2 different samples  
670 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
671 mega base (Mb).

672 **S6 Fig. Methylation pattern of Chromosome 3 and chromosome 4 in CG, CHG and CHH**  
673 **context:** Average methylation levels of the different context between the 2 different samples  
674 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
675 mega base (Mb).

676 **S7 Fig. Methylation pattern of Chromosome 5 and chromosome 6 in CG, CHG and CHH**  
677 **context:** Average methylation levels of the different context between the 2 different samples  
678 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
679 mega base (Mb).

680 **S8 Fig. Methylation pattern of Chromosome 7 and chromosome 8 in CG, CHG and CHH**  
681 **context:** Average methylation levels of the different context between the 2 different samples  
682 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
683 mega base (Mb).

684 **S9 Fig. Methylation pattern of Chromosome 9 and chromosome 10 in CG, CHG and CHH**  
685 **context:** Average methylation levels of the different context between the 2 different samples  
686 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
687 mega base (Mb).

688 **S10 Fig. Methylation pattern of Chromosome 11 and chromosome 12 in CG, CHG and CHH**  
689 **context:** Average methylation levels of the different context between the 2 different samples  
690 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
691 mega base (Mb).

692 **S11 Fig. Methylation pattern of Chromosome 13 and chromosome 14 in CG, CHG and CHH**  
693 **context:** Average methylation levels of the different context between the 2 different samples  
694 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
695 mega base (Mb).

696 **S12 Fig. Methylation pattern of Chromosome 15 and chromosome 16 in CG, CHG and CHH**  
697 **context:** Average methylation levels of the different context between the 2 different samples  
698 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
699 mega base (Mb).

700 **S13 Fig. Methylation pattern of Chromosome 17 and chromosome 18 in CG, CHG and CHH**  
701 **context:** Average methylation levels of the different context between the 2 different samples  
702 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
703 mega base (Mb).

704 **S14 Fig. Methylation pattern of Chromosome 19 and chromosome 20 in CG, CHG and CHH**  
705 **context:** Average methylation levels of the different context between the 2 different samples  
706 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
707 mega base (Mb).

708 **S15 Fig. Methylation pattern of Chromosome 21 and chromosome 22 in CG, CHG and CHH**  
709 **context:** Average methylation levels of the different context between the 2 different samples  
710 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
711 mega base (Mb).

712 **S16 Fig. Methylation pattern of Chromosome 23 and chromosome 24 in CG, CHG and CHH**  
713 **context:** Average methylation levels of the different context between the 2 different samples  
714 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
715 mega base (Mb).

716 **S17 Fig. Methylation pattern of Chromosome 25 and chromosome 26 in CG, CHG and CHH**  
717 **context:** Average methylation levels of the different context between the 2 different samples  
718 MOR (red) and FOR (blue). Y-axis average methylation levels, x-axis chromosome coordinates in  
719 mega base (Mb).

720 **S18 Fig. Methylation pattern of Chromosome XX in CG, CHG and CHH context:** Average  
721 methylation levels of the different context between the 2 different samples MOR (red) and FOR  
722 (blue). Y-axis average methylation levels, x-axis chromosome coordinates in mega base (Mb).

723 **S1 File. Results of quality trimming step by TrimGalore. (XLXS)**

724 **S2 File. Results from the alignment and methylation sites determination steps. (XLXS)**

725 **S3 File. Oligonucleotide primers used for real-time polymerase chain reaction. (XLXS)**

726 **S4 File. Differentially methylation and Functional annotation of CpG context. (XLXS)**

727 **S5 File. Differentially methylation and Functional annotation of CHG context. (XLXS)**

728 **S6 File. Differentially methylation and Functional annotation of CHH context. (XLXS)**

729 **S7 File. DMR Genes Overlapping between CpG, CHG and CHH context. (XLSX)**

730 **S8 File Differentially expressed genes between MORs and FORs. (XLSX)**

731

732 **Author Contributions**

733 **Conceptualization:** Eric Vilain, Charles Roselli

734 **Data curation:** Surajit Bhattacharya.

735 **Formal analysis:** Surajit Bhattacharya.

736 **Investigation:** Rebecka Amodei.

737 **Methodology:** Rebecka Amodei.

738 **Project administration:** Charles Roselli.

739 **Resources:** Eric Vilain, Charles Roselli.

740 **Visualization:** Surajit Bhattacharya.

741 **Writing – Original Draft Preparation:** Surajit Bhattacharya, Eric Vilain, Charles Roselli.

742 **Writing – Review & Editing:** Surajit Bhattacharya, Rebecka Amodei, Eric Vilain, Charles Roselli.

743

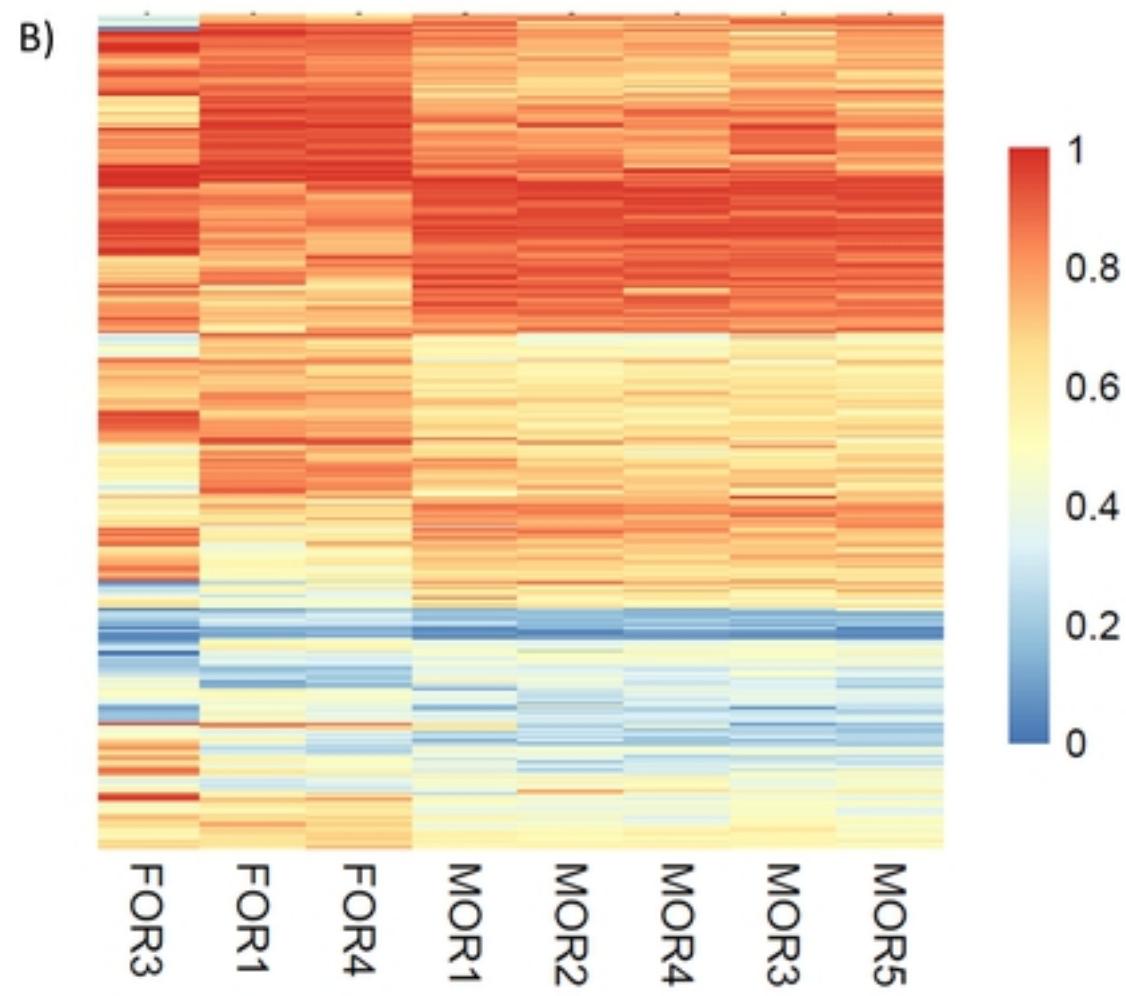
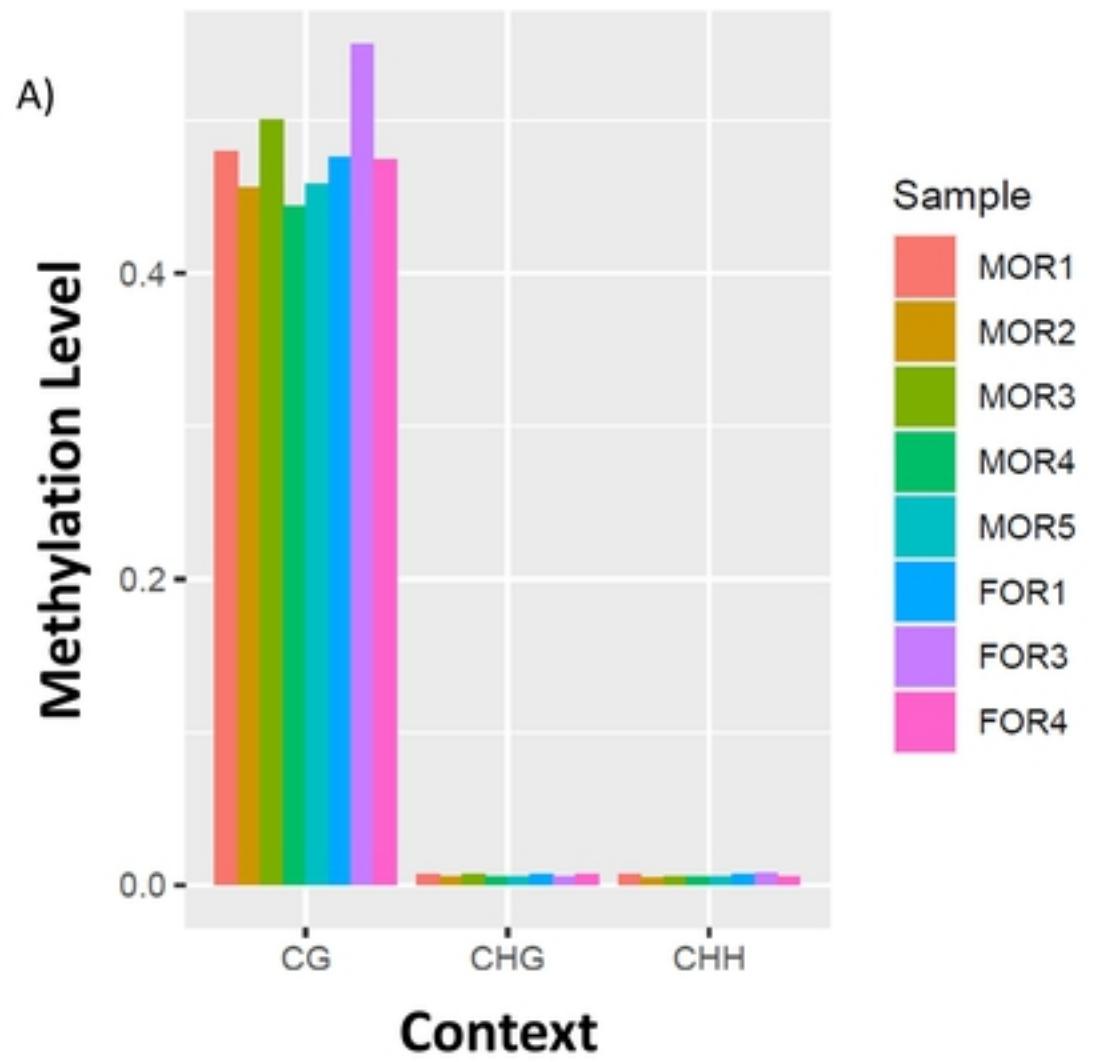


Figure 1

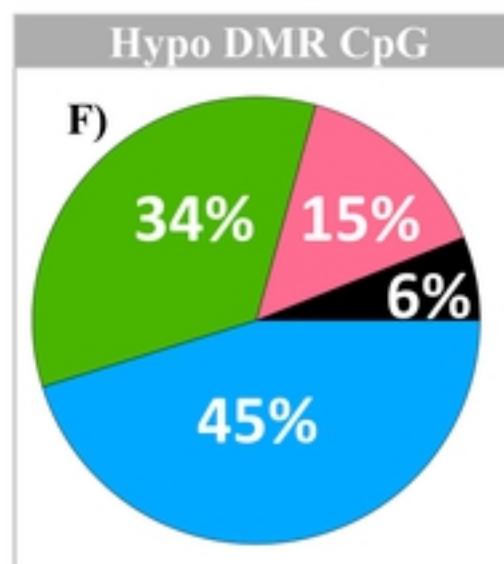
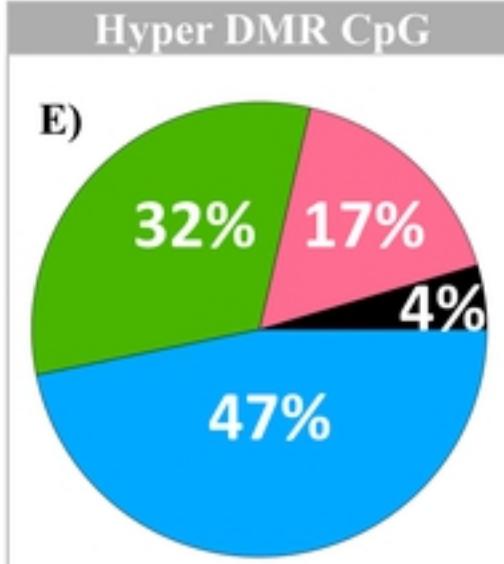
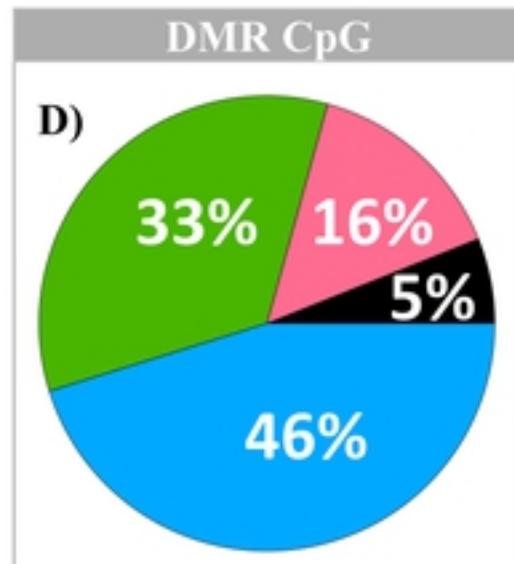
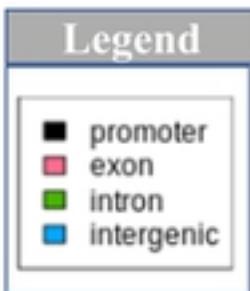
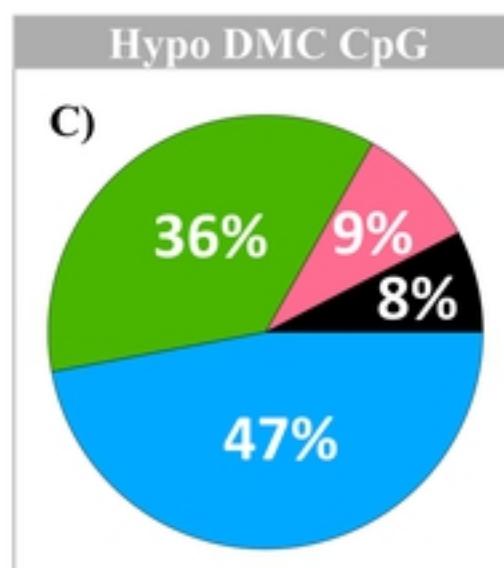
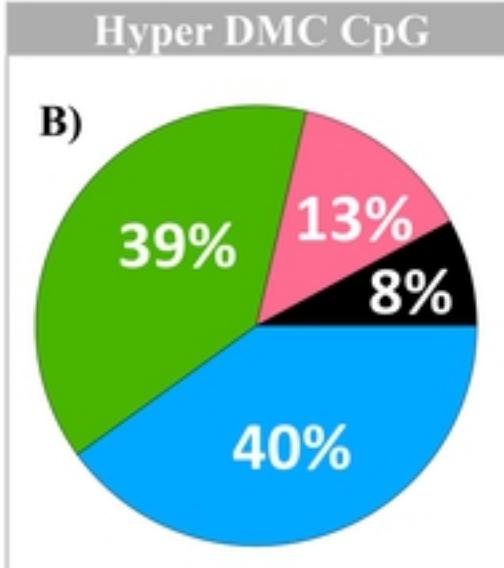
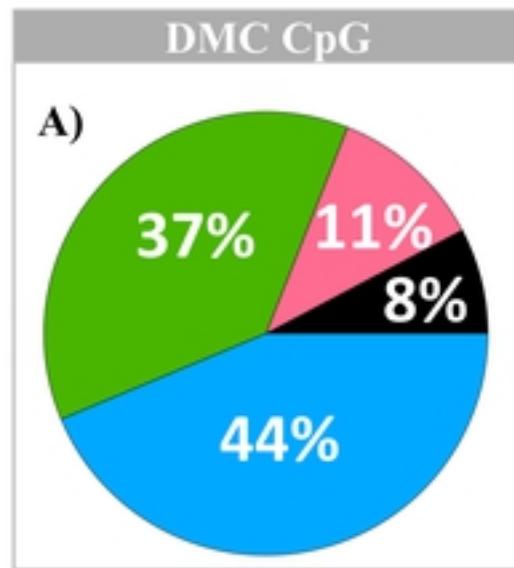
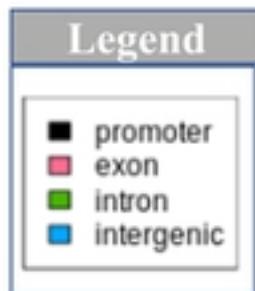
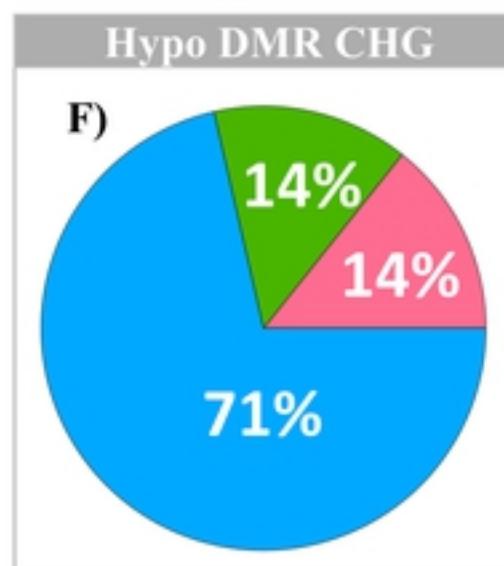
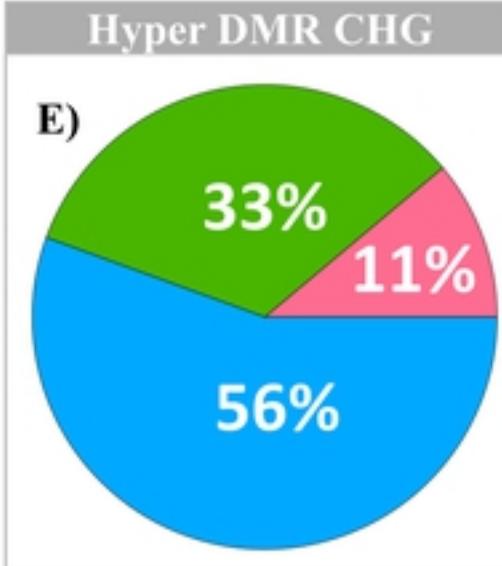
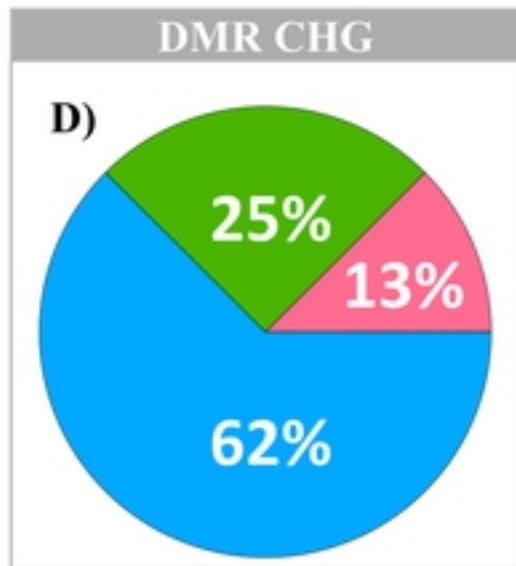
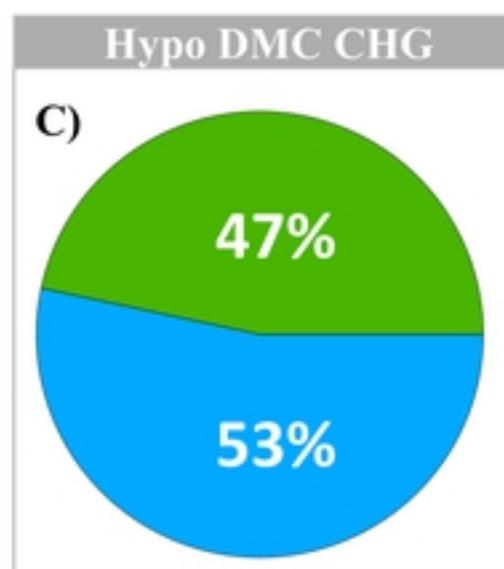
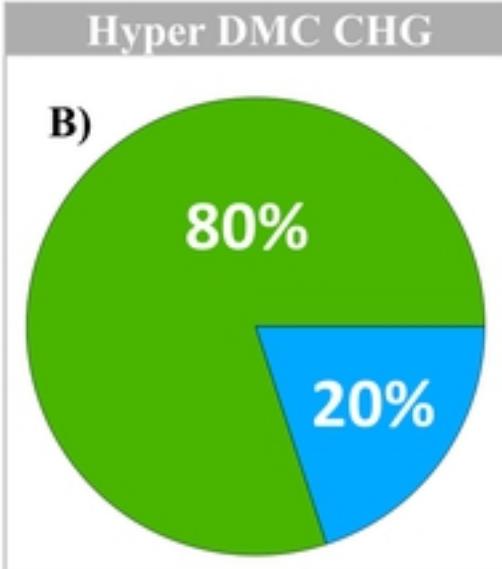
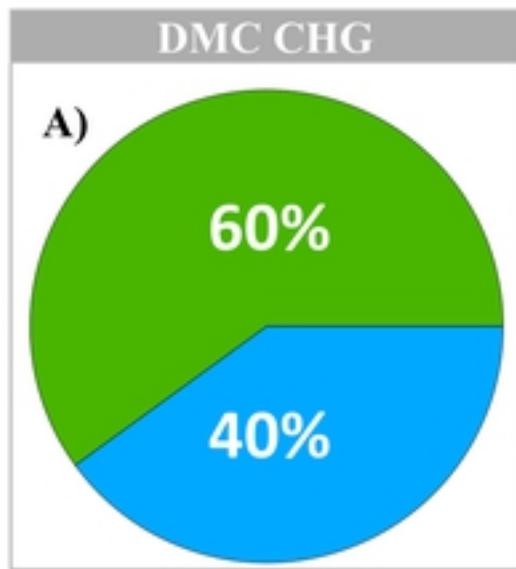


Figure 2



**Figure 3**

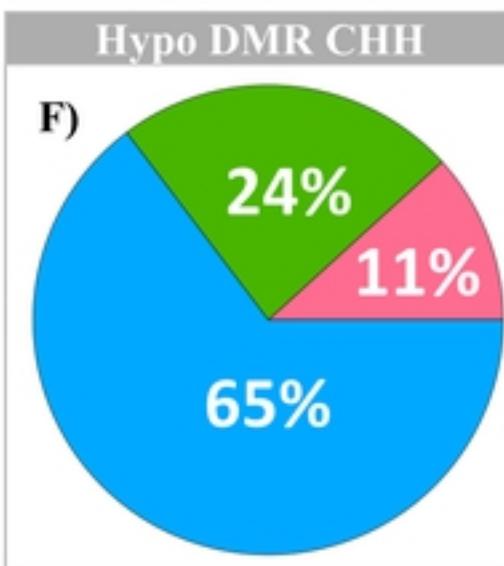
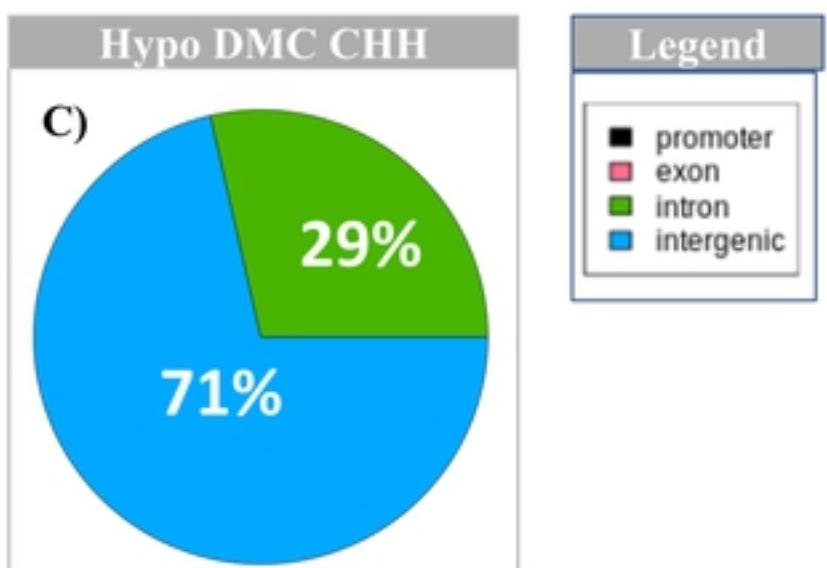
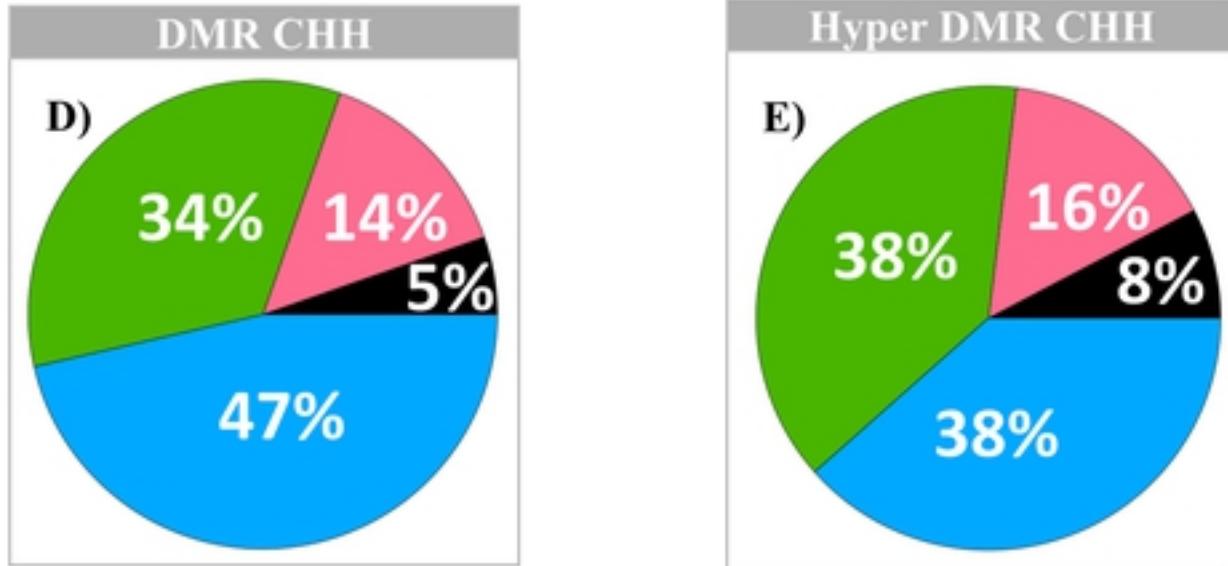
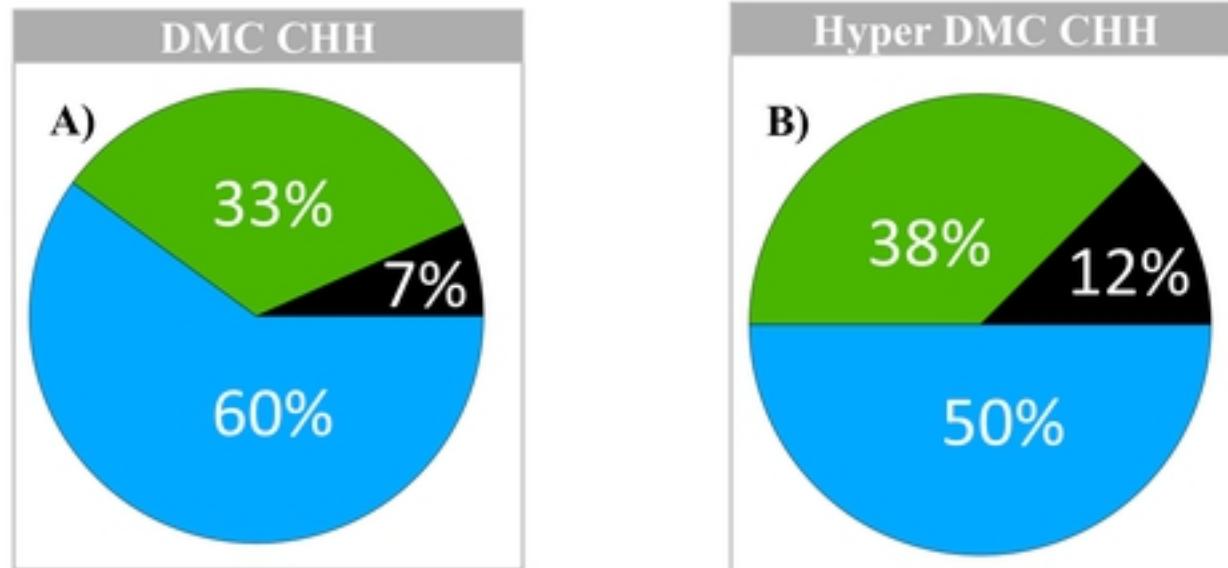
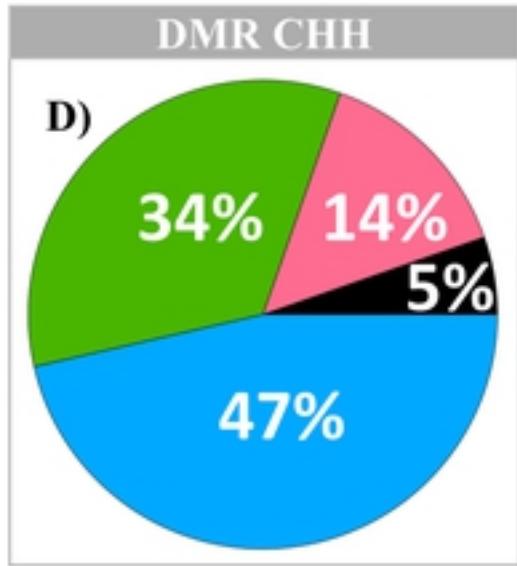
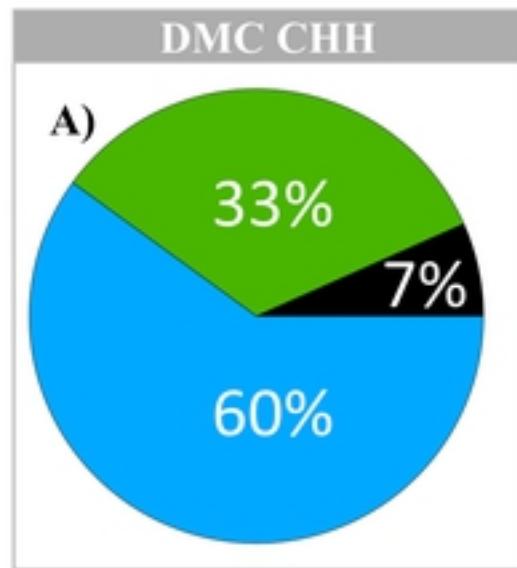
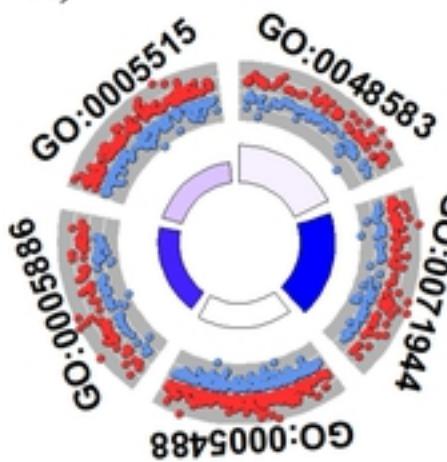


Figure 4

### Gene distribution for DMC CpG functional annotation

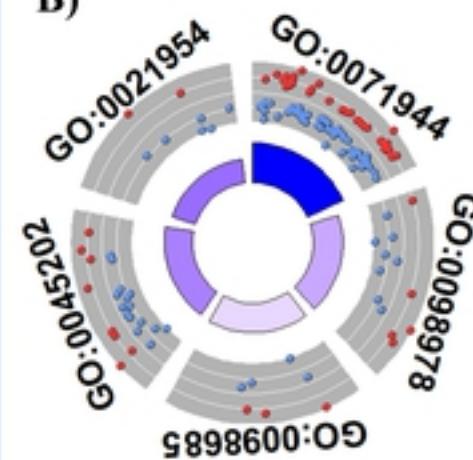
A)



ID	Description
GO:0048583	regulation of response to stimulus
GO:0071944	cell periphery
GO:0005488	binding
GO:0005886	plasma membrane
GO:0005515	protein binding

### Gene distribution for DMR CpG functional annotation

B)



ID	Description
GO:0071944	cell periphery
GO:0098978	glutamatergic synapse
GO:0098685	Schaffer collateral - CA1 synapse
GO:0045202	synapse
GO:0021954	central nervous system neuron development

z-score



logFC

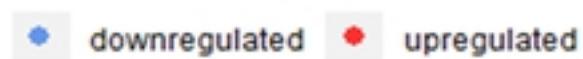


Figure 5

Common Genes Between CpG, CHG and CHH

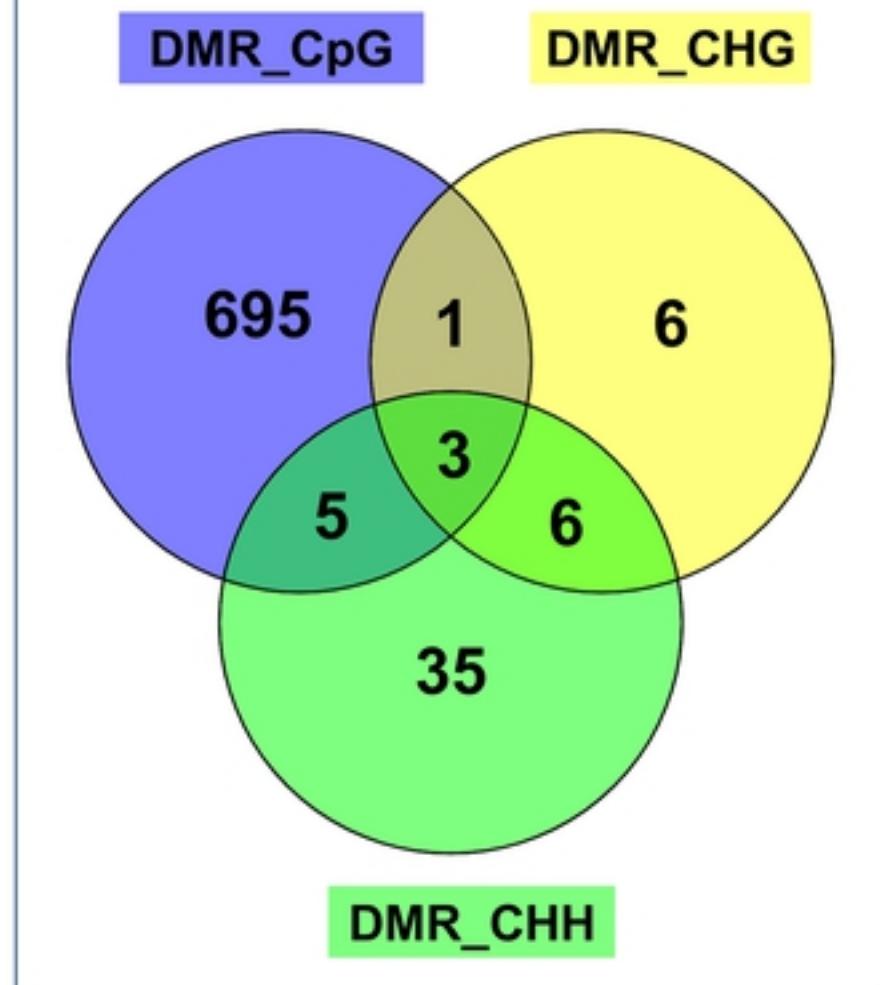


Figure 6

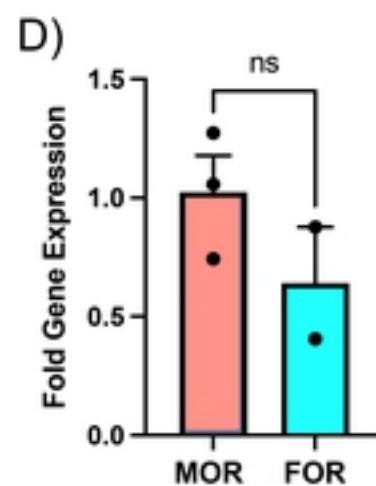
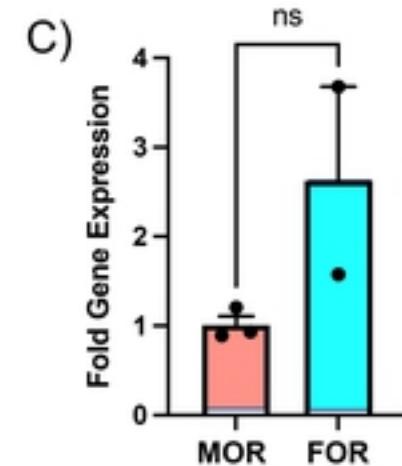
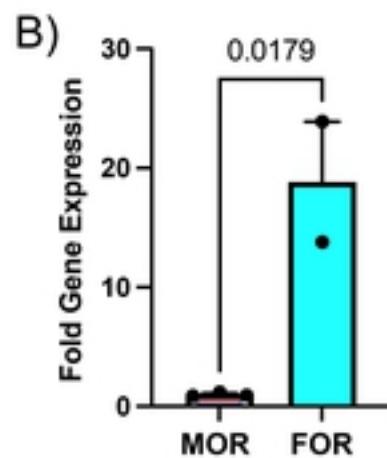
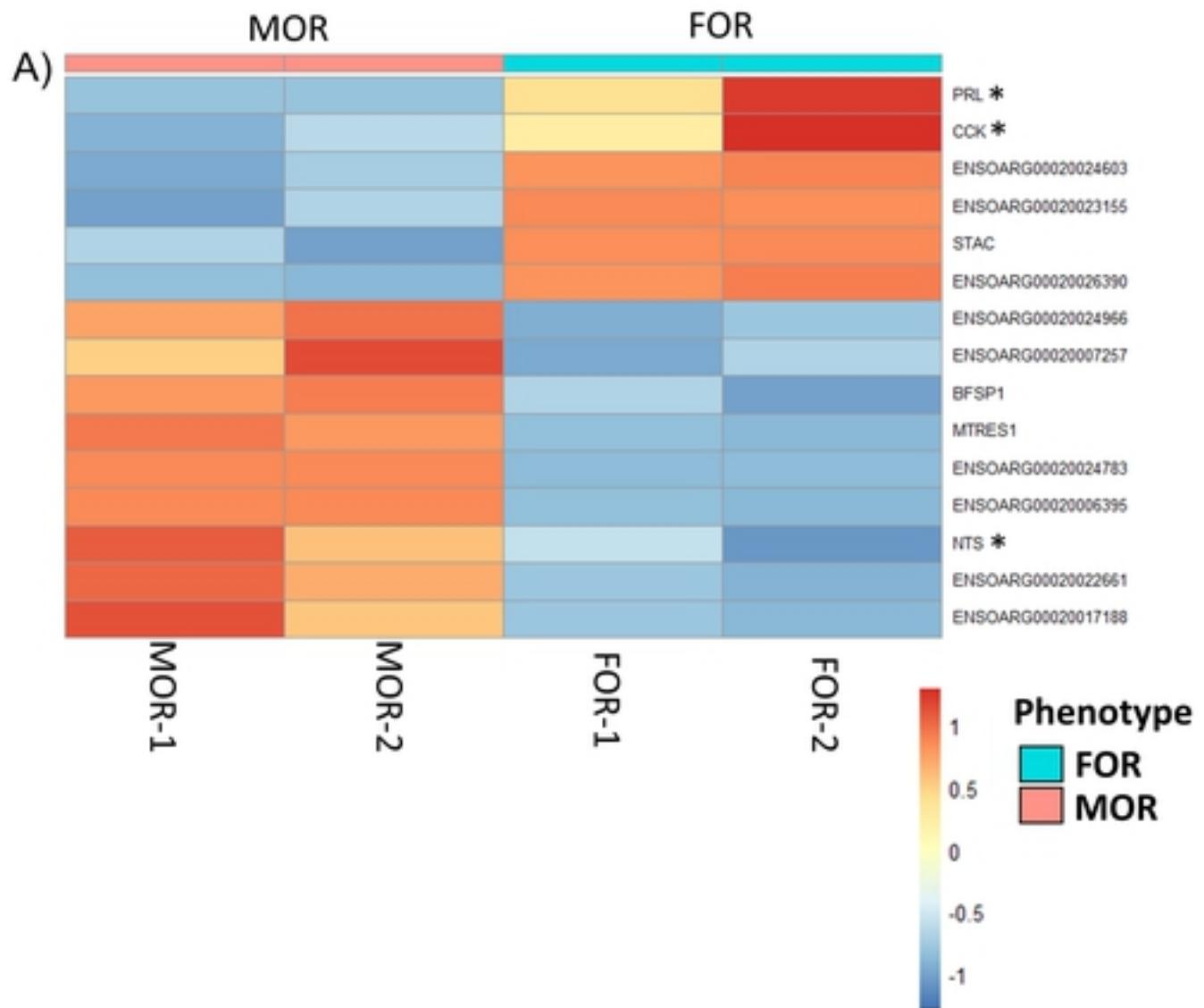


Figure 7