Methylation pattern of polymorphically imprinted *nc886* is not conserved across mammalia

Daria Kostiniuk[1¶], Hely Tamminen[1¶], Pashupati P. Mishra[2,3,4], Saara Marttila[1,5&], Emma Raitoharju[1,3,6&*]

[1]Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[2]Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[3]Finnish Cardiovascular Research Centre, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[4]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland

[5]Gerontology Research Center, Tampere University, Finland

[6]Tampere University Hospital, Tampere, Finland

*Corresponding author

Email: emma.raitoharju@tuni.fi (ER)

¶DK and HT are joint first authors.

& SM and ER are joint senior authors.

Abstract

Background: In humans, the *nc886* locus is a polymorphically imprinted metastable epiallele. Periconceptional conditions have an effect on the methylation status of *nc886*, and further, this methylation status is associated with health outcomes in later life, in line with the Developmental Origins of Health and Disease (DOHaD) hypothesis. Animal models would offer opportunities to study the associations between periconceptional conditions, *nc886* methylation status and metabolic phenotypes further. Thus, we set out to investigate the methylation pattern of the *nc886* locus in non-human mammals.

Data: We obtained DNA methylation data from the data repository GEO for mammals, whose *nc886* gene included all three major parts of *nc886* and had sequency similarity of over 80% with the human *nc886*. Our final sample set consisted of DNA methylation data from humans, chimpanzees, bonobos, gorillas, orangutangs, baboons, macaques, vervets, marmosets and guinea pigs.

Results: In human data sets the methylation pattern of *nc886* locus followed the expected bimodal distribution, indicative of polymorphic imprinting. In great apes, we identified a unimodal DNA methylation pattern with 50% methylation level in all individuals and in all subspecies. In Old World monkeys, the between individual variation was greater and methylation on average was close to 60%. In guinea pigs the region around the *nc886* homologue was non-methylated. Results obtained from the sequence comparison of the CTCF binding sites flanking the *nc886* gene support the results on the DNA methylation data.

Conclusions: Our results indicate that unlike in humans, *nc886* is not a polymorphically imprinted metastable epiallele in non-human primates or in guinea pigs, thus implying that animal models are not applicable for *nc886* research. The obtained data suggests that the *nc886*

region may be classically imprinted in great apes, and potentially also in Old World monkeys, but not in guinea pigs.

Key words: nc886, vtRNA2-1, primates, imprinting, DNA methylation

Background

In mammalian genomic imprinting, only one parental allele is expressed, while gene expression from the other allele is suppressed in a parent-of-origin-dependent manner. Imprinting has been hypothesized to have evolved in response to a parent-offspring conflict, where the maternal and paternal genome differ in their interests regarding the supply of resources [1]. Paternal genomes favor the opportunistic strategy to enhance the growth of the developing offspring through the expression of growth-enhancing genes, while maternal genomes aim to conserve maternal resources over multiple pregnancies [2]. On the other hand, according to the trophoblast defense theory, suppression of genes that promote placental development or to activate genes that limit this process in oocytes are thought to protect the females from parthenogenetically activated oocytes and consequential malignant trophoblast [3,4]. The expression of imprinted genes in general has been associated with fetal and placental growth and suggested to a have a role in the development of cardiometabolic diseases in adulthood [2,4,5]. Imprinting arose relatively recently at most loci— while only a few imprinted genes in Eutherians are also imprinted in marsupials, to date no imprinting has been reported in the egg-laying monotreme mammals [6]. This supports the conflict-of-interest theory, as only in non-egg laying mammals the growing fetus directly consumes the maternal resources [4]. Genetic imprinting is best described in mice, and while the mouse is an informative proxy for human imprinted gene regulation, less than half of the 100 human imprinted genes have been shown to be similarly imprinted in mice (https://www.geneimprint.com/site/home). Distinct differences in placental evolution, physiology, and reproductive biology of the primate and murine groups may be responsible for these differences.

During gametogenesis and fertilization, the original DNA methylation pattern of imprinted genes is erased, and parent of origin-based methylation pattern is established. While most

72 imprinted genes are located in clusters that are regulated by insulators or long noncoding RNAs
73 [3], some unclustered imprinted genes can be regulated by differential promoter methylation
74 [4]. Parental imprints are maintained after fertilization through these mechanisms despite
75 extensive reprogramming of the mammalian genome [4]. A common feature of imprinted genes
76 is insulators, such as CCCTC binding factor (CTCF) binding sites, which block the enhancers
77 from interacting with gene promoters and/or act as barrier to the spread of transcriptionally
78 repressive condensed chromatin [7].

79 The locus harbouring non coding 886 (*nc886*, also known as VTRNA2-1) in chromosome
80 5q31.1 is a unique example of imprinting, as it is the only known polymorphically imprinted
81 locus across tissues in adult human population, where the polymorphism is not due to genetic
82 variation [8–12]. However, in the placenta polymorphic imprinting is more common [13,14].
83 The *nc886* differentially methylated region (DMR) is 1.9-kb long, and its boundaries are
84 marked by two CTCF binding sites [10,15]. This DMR has been shown to present maternal
85 imprinting in ~75% of individuals in several populations [10,12,16]. This means that while in
86 all individuals the paternal allele is unmethylated, in approximately 75% of individuals the
87 maternal allele is methylated (individuals present a 50% methylation level at *nc886* locus) and
88 in the remaining 25% of individuals the maternal allele is unmethylated (individuals present a
89 0% methylation level at the *nc886* locus).

90 The *nc886* gene codes for a 102nt long, non-coding RNA, which is then cleaved into two short
91 RNAs (hsa-miR-886-3p/nc886-3p [23 nt] and hsa-miR-886-5p/nc886-5p [24–25 nt]) [17–19].
92 There is no consensus on whether the effects of *nc886* expression is mediated by the 102 nt
93 long hairpin structure or the nc886-3p and -5p molecules, as the short molecules have been
94 indicated to function as miRNAs, while the hairpin loop has been shown to inhibit protein
95 kinase R (PKR) [17,20]. Expression of nc886 RNAs is strongly associated with the methylation
96 status of the *nc886* locus. Individuals with non-methylated *nc886* present approximately two-

97 fold levels of nc886 RNAs in blood, as compared to individuals with monoallelic methylation
98 [16], thus implying allele-specific expression. However, allele-specific expression has not been
99 experimentally shown, as the 102 nt transcript does not harbour SNPs [10].

100 The periconceptional environment has been suggested to affect DNA methylation patterns in
101 maternal alleles[21], including the *nc886* epiallele [10,15,16]. Season of conception, maternal
102 age and socioeconomic status have been linked to changes in the proportion of offspring with
103 unmethylated maternal allele [10,15,16]. On the other hand, lower levels of *nc886* methylation
104 have been linked to cleft palate [22], and a non-methylated *nc886* epiallele has been associated
105 with an elevated childhood BMI [23]. The methylation status of this epiallele has also been
106 associated with allergies [24], asthma [25], infections [26], and inflammation [27]. We and
107 others have also shown that both the *nc886* methylation status and RNA expression are
108 associated with indicators of glucose metabolism [16,28].

109 These results indicate that *nc886* could mediate the association between periconceptional
110 conditions and later metabolic health, in line with the Developmental Origins of Health and
111 Disease (DOHaD) hypothesis (aka the Barker hypothesis) [29]. More detailed analysis on
112 periconceptional conditions and *nc886* methylation status and investigations between *nc886*
113 and metabolic phenotypes, with less cofounding factors, would be needed to confirm this
114 hypothesis. As carcinogenesis [9] and pluripotency induction [16] affect the DNA methylation
115 pattern in *nc886* locus, *in vitro* work has its limitations. Animal models could be a feasible
116 option for this research. Unfortunately, rodents do not harbor the *nc886* gene, limiting the use
117 of traditional model organisms [10]. Thus, this study was set up to investigate 1) which animals
118 have *nc886* gene, 2) whether this gene is surrounded by similar CTFC elements as the human
119 homolog and 3) whether the methylation status of the *nc886* region suggest polymorphic
120 imprinting in non-human mammals.

121

122 # Materials and methods

123 The presence of *nc886* gene was investigated in ensemble, in 65 amniota vertebrates Mercator-
124 Pecan collection and 24 primates EPO-extended collection [30]. To select species for further
125 investigation, we required the *nc886* gene have 80% sequence similarity with the human
126 homolog and to present the sequences for nc886-3p and nc886-5p RNAs, as well as the loop
127 structure, previously shown to mediate the binding of PKR [17,31] (S1 and S2 Figs). The
128 existence and sequence similarity of the centromeric (chr5:135415115-135415544) and
129 telomeric CTCF (chr5:135418124-135418523) binding site flanking *nc886* gene were also
130 investigated in species shown to harbor intact *nc886* gene. If homologous CTCF-binding sites
131 were not discovered, CTCFBSDB 2.0 [32] was utilized to predict possible non-homologous
132 sites. Interactions of the *nc886* flanking CTCF-sites were also investigated using K562 CTCF
133 ChIA-PET Interactions data and Hi-C data in genome browser [33] and Hi-C data in 3D
134 Genome Browser from HUVEC and K562 cell lines [34].

135 For species harboring the *nc886* gene, we investigated the Gene Expression Omnibus (GEO)
136 repository [35] for available DNA methylation data, with both general and binomial name of
137 the species. DNA methylation data was available in apes from chimpanzees (*Pan troglodytes*,
138 n=83; GSE136296 [36] and n=5; GSE41782 [37]), bonobos (*Pan paniscus*, n=6; GSE41782
139 [37]), gorillas (*Troglodytes gorilla*, n=6; GSE41782 [37]) and orangutangs (*Pongo spp*., n=6;
140 GSE41782[37]). In Old World monkeys, data was obtained from baboons (*Papio spp.* n=28;
141 GSE103287 [38]), rhesus macaques (*Macaca mulatta,* n=10; GSE103287 [38]), vervets
142 (*Chlorocebus aethiops*, n=10; GSE103287 [38]) and in New World monkeys, from marmosets
143 (*Callithrix jacchus*, n=6; GSE103287 [38]). From primates, only data from blood or femur was
144 utilized. In addition to primates, we obtained DNA methylation data from guinea pig

145 hippocampus (*Cavia porcellus,* n=36; GSE109765 [39]). As reference, we utilized data from
146 human (*Homo sapiens*) blood (n=1658; GSE105018 [40]), femur (n=48; GSE64490 [41]) and
147 hippocampus (n=33; GSE72778 [42]).

148 Methylation profiling data obtained by high throughput sequencing (guinea pigs, GSE109765)
149 was processed as follows. Quality of the paired-end reads in all the samples was assessed using
150 FastQC [43] and MultiQC [44]. Paired-end fastq files were trimmed using Trimmomatic-0.39
151 with a sliding window of size 4 set to remove bases with phred score lower than 20 [45]. The
152 trimmed samples were analyzed using Bismark-0.23.0 tools [46]. The reads were aligned to
153 the guinea pig genome (cavPor3). Duplicate alignments, which can arise for example by PCR
154 amplification, were removed. Methylation information was extracted from the alignment result
155 files using Bismark's methylation extractor. DNA methylation values for CpGs inside the gene
156 were first inspected and then a wider region (±2000nt) around the gene was investigated.

157 Primate DNA methylation data from GSE41782, GSE105018, GSE64490, GSE72778
158 (profiled with Illumina 450K) and GSE136296 (profiled with Illumina EPIC) were available
159 as processed data and was used as such. Primate DNA methylation data from GSE103271,
160 GSE103280, GSE103286, which are subseries of GSE103332 (profiled with Illumina EPIC),
161 were available as raw data, and were normalized by using minfi quantile normalization for each
162 species separately.

163 From primate data, the 14 CpGs in the *nc886* DMR previously reported to show bimodal
164 methylation pattern in humans were retrieved [10,16]. In all the primate species, for which
165 methylation data was available, the sequence on the binding site of the Illumina probes was
166 investigated. Only data from sites with the CG-sequence intact in the species in question were
167 further utilized. (S1 File). Similar process was repeated for 50 probes in paternally expressed

168 10 (*PEG10*) previously shown to be imprinted [47]. *PEG10* was used as evolutionally
169 conserved reference for a classically imprinted gene [6].

170

171 Results and Discussion

172 *nc886* gene in non-human mammals

173 Human *nc886* has been suggested to be an evolutionally young gene, producing a 102 nt long
174 RNA, which is then ineffectively cleaved to two miRNA-like RNAs [17]. In line with previous
175 reports [10], *nc886* gene, with intact short RNA coding sequences and the hairpin loop, can be
176 found in primates, in guinea pig (*Cavia porcellus*), Eurasian red squirrel (*Sciurus vulgaris*),
177 and Alpine marmot (*Marmota marmota*), with two of the latter having insertions in the
178 centromeric end of the gene (Fig 1, Fig 2, S1 Fig, S1 Table). In humans, and all the other
179 species where *nc886* was found, *nc886* is located between *TGFB1* and *SMAD5*.

180

181 **Fig 1. Schematic presentation of the similarity of the *nc886* gene and flanking CTCF**
182 **binding sites.** Guinea pigs lack the whole centromeric flanking CTCF binding site, whilst the
183 marmosets have a region with sequence similarity, but lack the binding sequence of the CTCF.
184 Mouse genome does not contain either the *nc886* gene or the CTCF binding sites flanking it.
185 For detailed sequence comparisons, see S1 and S2 Figs.

186

187 **Fig 2. Schematic presentation of the *nc886* locus.** In red, the *nc886* gene (chr5:135416180-
188 135416287), in green the centromeric CTCF binding site (chr5:135415115-135415544) and in
189 blue the telomeric CTFC binding site (chr5:135418124-135418523). The binding motif of the

190 CTCF sites are presented in yellow and the 14 CpGs with bimodal methylation pattern in
191 human cohorts are presented according to their genomic location (GRCh37/hg19).

192

193 Upon further inspection of primate *nc886* gene, almost 100% sequence similarity was
194 identified in apes (*hominoidea*). The sequence similarity was high (over 97%) between humans
195 and the investigated Old World monkeys (*Cercopithecidae*), while less (over 90%) similarity
196 can be seen between humans and New World monkeys (*Ceboidea*) and even less (84-89%)
197 between humans and tarsiers (*Tarsiidae*) or lemurs (*Lemuroidea*). Sequences coding for the
198 short nc886 RNAs are identical within Old World anthropoids (*Catarrhini).* Differences
199 between human and New World monkey and prosimian *nc886* sequences can be found both in
200 the short nc886 RNA and the hairpin coding regions, but the most significant differences can
201 be found in the centromeric end of the gene (Fig 1 and S2 Fig).

202

203 CTCF binding site sequence in mammals with *nc886* gene

204 In humans, the *nc886* locus is flanked by two CTCF binding sites (Fig 2) [10]. The telomeric
205 CTCF binding site can be found in most vertebrates, with the binding motif being identical in
206 all species showed to have the *nc886* gene (S1 Table). The telomeric CTCF binding site was
207 shown to interact with a CTCF binding site (chr5:135222814-135223707) locating near the *IL9*
208 gene in the CTCF ChIA-PET data, the interaction is also supported by the Hi-C data from cell
209 lines (S3 Fig). This prediction is in line with our previous finding indicating that genetic
210 polymorphisms near *IL9* gene are associated with the expression of nc886 RNAs [16]. This
211 CTCF binding site locating near *IL9* can be found in all primates (S1 Table). Together these
212 results suggest that the evolutionally conserved telomeric CTCF binding site of *nc886* interacts
213 with CTCF binding sites near *IL9* gene, possibly forming a topologically associating domain

214 (TAD), or an interaction within one (sub-TAD), and bringing the suggested enhancer area near
215 the *nc886* gene [48]. On the contrary, the centromeric CTCF binding site is present only in
216 primates and even in primates, the binding sequence cannot be identified in marmosets. As
217 CTCF binding sites can act as barriers to the spread of transcriptionally repressive condensed
218 chromatin [7], presence, or absence, of the CTCF binding site may be associated with the
219 methylation status of *nc886* locus in different species. For the centromeric CTCF binding site
220 no interactions were detected according to CTCF ChIA-PET data. There are also changes in
221 the CTFC binding motif in gorillas (position 9), in all Old World monkeys (position 4) and also
222 in New World monkeys (position 14) (S1 Table). According to the CTCFBSDB 2.0, the guinea
223 pig genome does not harbor any non-homologous predicted CTCF binding sites in the
224 centromeric side of the *nc886* gene.

225

226 Guinea pigs present a non-methylated *nc886* locus

227 One of the aims of this study was to investigate whether the information gathered from the
228 *nc886* gene and DMR from cell culture and population studies could be supplemented with
229 research on animal models. In line with a previous report [10] we identified this gene only in
230 primates, guinea pigs and few members of the squirrel family, of which guinea pig was the
231 most promising candidate as a model organism. In data from Constantinof et al. [39], in guinea
232 pig hippocampi (n=36) the whole *nc886/vtRNA2-1* gene was non-methylated (S4 Fig). The
233 surrounding *nc886* region (Scaffold DS562872.1: 24,622,179-24,622,280) +/- 2000 nt was
234 mostly unmethylated, with only 2% of the reads in the region being methylated. It should be
235 noted that the number of reads in the region in the data utilized was low (max number of
236 reads=17, average number of reads=7). *nc886* methylation pattern in human hippocampi
237 presented the expected bimodal distribution, and thus the discovered methylation pattern in the

238 guinea pig hippocampi was most likely not due to the selection of tissue (S4 Fig). This
239 identified lack of methylation in the *nc886* locus is compatible with the absence of the telomeric
240 CTCF binding site, as CTCF binding sites can delineate the boundaries of an imprinted region
241 [49]. These results thus suggest that guinea pigs are not suitable model organisms for the
242 investigation of establishment of *nc886* methylation status.

243

244 Imprinted *nc886* region in great apes

245 The blood of chimpanzees, gorillas, bonobos and orangutans presented beta-values close to 0.5
246 with a unimodal distribution in the *nc886* region (Fig 3, S2 Table). This methylation pattern
247 closely resembles the methylation pattern in the known maternally imprinted gene *PEG10* (S5
248 Fig). The methylation levels are also very similar to those presented in humans with
249 monoallelic methylation (Fig 3). In these data sets, that include more than 110 apes, we did not
250 identify any individuals with methylation level close to 0 in the *nc886* locus, whereas in
251 humans 25% of the population present a methylation level close to 0 at this locus [10,16]. If
252 the prevalence on non-methylated individuals in apes was similar to humans, already 11
253 individuals would present at least one non-methylated individual with 95% probability. Of
254 individual species, we had the largest dataset for chimpanzees (n=83 in GSE136296 and n=5
255 in GSE41782). Again, assuming the same proportion of non-methylated individuals as in
256 humans (25%), probability of not identifying any non-methylated chimpanzees in a population
257 of 88 individuals is extremely low, $1.01*10^{-11}$. These results imply that the *nc886* locus is not
258 polymorphically imprinted in apes.

259

260 **Figure 3. Blood DNA methylation beta values in *nc886* region in great apes (Hominidae).**
261 In each graph, one dot represents one individual. In humans a bimodal methylation pattern can

262 be detected, in line with the expected population distribution of individuals with monoallelic
263 methylation (75% of the population, methylation level ~0.5) and non-methylated individuals
264 (25% of the population, methylation level close to 0) [10,16]. All of the other species present
265 a unimodal methylation pattern, with all individuals having methylation beta-values near 0.5
266 across the *nc886* locus. Number of individuals is 1658 for humans (GSE105018), 6 for bonobos
267 (GSE41782), 83 and 5 for chimpanzees (GSE136296 and GSE41782, respectively), 6 for
268 gorillas (GSE41782) and 6 for orangutans (GSE41782). For schematic representation of the
269 CpG sites in the *nc886* locus, see Fig 2.

270

271 Within species, median of standard deviation of the probe methylation of *nc886* is around 0.04,
272 which is comparable to that seen in *PEG10* in apes (median of SD 0.03) and in human blood
273 (0.02), indicating good data quality (S2 Table). In all apes the methylation levels of *nc886*
274 region resembled those seen in *PEG10,* which is an evolutionally conserved maternally
275 imprinted gene [6]. It is thus reasonable to suggest that *nc886* could be classically imprinted in
276 other great apes, excluding humans.

277

278 *nc886* region methylation patterns in Old World monkeys

279 The patterns of *nc886* region methylation are very similar in all of the Old World monkeys.
280 The median methylation level is close to 0.60, higher as compared to apes. The interindividual
281 variation is larger, especially in baboons, than in apes, with the median SD within a probe being
282 0.11. The SD within a probe is also higher in probes locating in *PEG10* in baboons, where the
283 median of probe SD is 0.05 (S2 Table). As the between individual variation in methylation
284 levels of a known evolutionally conserved imprinted gene is also higher, this suggests a
285 technical bias in the data, potentially due to the use of Illumina Infinium 450K and EPIC

286 methylation assays, that are designed for humans. All methylation data available for Old World
287 monkeys was from femur, but as the methylation pattern of *nc886* has been reported to be
288 constant across tissues in humans [12] and in the human reference data set form femur samples
289 both *nc886* and *PEG10* present similar methylation patterns as in blood (Fig 4 and S6 Fig), this
290 phenomenon most likely is not caused by the tissue of origin. Regardless of the precise
291 methylation levels of the Old World monkeys, in the 48 individuals we did not identify any
292 presenting a non-methylated methylation pattern in *nc886* region, probability of which is
293 $1.0*10^{-6}$, when assuming similar distribution as in humans. This implies that similar to non-
294 human great apes, the *nc886* locus is not polymorphically imprinted in Old World monkeys.

295

296 **Figure 4. Femur DNA methylation beta values in the *nc886* region in humans and in**
297 **monkeys.** In humans a bimodal methylation pattern can be detected, similar to blood and in
298 line with the expected population distribution of individuals with monoallelic methylation
299 (75% of the population, methylation level ~0.5) and non-methylated individuals (25% of the
300 population, methylation level close to 0) [10,16]. In Old World monkeys the between
301 individual variation is bigger than in apes. No individuals presenting a non-methylated *nc886*
302 region are detected. In marmosets only 3 probes were considered to provide reliable
303 methylation values and none of them present methylation levels near 0.5. Number of
304 individuals is 48 for humans (GSE64490), 28 for baboons (GSE103287), 10 for macaques
305 (GSE103287), 10 for vervets (GSE103287) and 6 for marmosets (GSE103287). For schematic
306 representation of the CpG sites in the *nc886* locus, see Fig 2.

307

308 nc886 region methylation patterns in New World monkeys

309 Data from the six marmosets is not conclusive, as we found that only 3 of the probes in *nc886*
310 region bound to areas with no great sequence differences. The methylation levels of two of
311 these probes were around 0.8 and one 0.32, showing no indications of imprinting, while the
312 median of methylation beta values in the probes locating in the *PEG10* is 0.46 in marmosets
313 (Fig 4 and S6 Fig, S2 Table). The lack of imprinting of any kind in marmosets is further
314 supported by the finding that they lack the centromeric CTCF binding motif, which is thought
315 to have an important role in insulating the DMR [10]. To draw conclusions on the methylation
316 status of *nc886* in New World monkeys in general, methylation data from more species would
317 be needed.

318

319 Limitations of the study

320 Our study is purely descriptive in nature. In guinea pigs, the shallowness of the sequencing
321 data limits the ability to make conclusions of the methylation pattern. In non-human primates,
322 utilizing methylation arrays that have been designed for humans raises questions on data
323 quality, especially in marmosets. Concerns over data quality are however mitigated by the
324 observed methylation pattern in well-established imprinted gene, *PEG10*, as well as the
325 consistency of observed results across different species and data sets. In addition, while a
326 methylation level of 0.50 implies allele specific methylation, we were not able to confirm this
327 due to lack of suitable data.

328

329 Conclusions

330 We describe here an analysis on the methylation status of *nc886* region in non-human
331 mammals. A genetic locus, with more than 80% similarity to human *nc886* gene can be found

332 in primates, guinea pigs and some members of the squirrel family. We obtained DNA
333 methylation data for 8 different non-human primate species and for guinea pigs, and in none
334 of these species could we observe methylation pattern indicative of similar polymorphic
335 imprinting, as could be observed, and has been reported [10,12,16], in humans. The observed
336 methylation pattern in apes and in Old World monkeys implies that the *nc886* region might be
337 classically imprinted, although these findings have to be interpreted with caution, as apart from
338 chimpanzees the sample number were low, and in Old World monkeys the variation between
339 individuals was notable. In guinea pigs, the most feasible potential model organism of those
340 harboring the *nc886* locus, the data indicated that the locus is completely unmethylated. It is
341 noteworthy that only primates, whose genome also contained the centromeric CTCF binding
342 sequence flanking the *nc886* gene, had methylation levels indicative of genetic imprinting.

343 In conclusion, we were unable to identify an animal model suited to study the establishment of
344 the methylation pattern of the polymorphically imprinted metastable epiallele *nc886*. Further
345 studies on how this kind of unusual metastable developed, and how it links to the
346 periconceptional conditions and later life health traits, are thus restricted to *in vitro* and
347 population studies.

348

349 Declarations

350 Ethics approval

351 Datasets used in this study were retrieved from Gene Expression Omnibus (GEO,
352 www.ncbi.nlm.nih.gov/geo/) repository. For human data sets, informed consent or "Consent
353 for Autopsy" were given by all participants. For animal studies all protocols were approved by

354 the local ethical committees and/or the samples were collected during standard veterinary
355 checks or routine necropsies. Details can be found from the original publications [36–42].

356 Consent for publication

357 Not applicable

358 Availability of data and materials

359 All methylation data utilized are available in GEO, under accession numbers GSE136296,
360 GSE41782, GSE103287, GSE109765, GSE105018, GSE64490 and GSE72778.

361 Competing interests

362 The authors declare that they have no competing interests

363 Funding

364 This study has been supported by the Academy of Finland (https://www.aka.fi/en/) (grants
365 330809 and 338395 (E.R.)); the Tampere University Hospital Medical Funds
366 (shorturl.at/gsyV0) (grants 9AC077, 9X047, 9S054, and 9AB059 for E.R.); the Yrjö Jahnsson
367 Foundation (https://www.yjs.fi/en/) (S.M. and E.R); Orion Research Foundation
368 (https://www.orion.fi/en/rd/orion-research-foundation/) (E.R.); the Foundation of Clinical
369 Chemistry (https://www.juhaniahonlaaketieteensaatio.fi/) (E.R), Laboratorioläätietieteen
370 edistämissäätiö sr. (https://www.labes.fi/) (E.R.) and the Paulo Foundation
371 (https://www.paulo.fi/in-english) (S.M). The funders had no role in study design, data
372 collection and analysis, decision to publish, or preparation of the manuscript.

373 Author Contributions

S.M. and E.R. designed the research; D.K., H.T., P.P.M., S.M. and E.R. performed research; S.M. and E.R. provided resources; D.K., H.T., P.P.M., S.M. and E.R. curated data; S.M. and E.R wrote the original draft manuscript; D.K., H.T., P.P.M., S.M. and E.R. reviewed and edited the manuscript. Funding was acquired by S.M and E.R. All authors have read and agreed upon the published version of the manuscript.

Acknowledgements

The authors wish to thank Jooseppi Järvinen (MSc) for his contribution to the probability calculations, Nina Mononen (PhD) and Binisha Mishra (MSc) for their valuable input on the manuscript and to acknowledge all the researchers who contributed towards the collection of datasets utilized in this study.

References:

1. Moore T, Haig D. Genomic imprinting in mammalian development: a parental tug-of-war. Trends in Genetics. 1991. doi:10.1016/0168-9525(91)90230-N

2. Smith FM, Garfield AS, Ward A. Regulation of growth and metabolism by imprinted genes. Cytogenetic and Genome Research. 2006. doi:10.1159/000090843

3. Varmuza S, Mann M. Genomic imprinting - defusing the ovarian time bomb. Trends in Genetics. 1994;10. doi:10.1016/0168-9525(94)90212-7

4. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. Cold Spring Harbor Perspectives in Biology. 2014;6. doi:10.1101/cshperspect.a018382

5. Millership SJ, Van de Pette M, Withers DJ. Genomic imprinting and its effects on postnatal growth and adult metabolism. Cellular and Molecular Life Sciences. 2019. doi:10.1007/s00018-019-03197-z

396 6. Renfree MB, Suzuki S, Kaneko-Ishino T. The origin and evolution of genomic
397 imprinting and viviparity in mammals. Philosophical Transactions of the Royal Society
398 B: Biological Sciences. 2013. doi:10.1098/rstb.2012.0151

399 7. Lewis A, Reik W. How imprinting centres work. Cytogenetic and Genome Research.
400 2006. doi:10.1159/000090818

401 8. Treppendahl MB, Qiu X, Søgaard A, Yang X, Nandrup-Bus C, Hother C, et al. Allelic
402 methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict
403 outcome in AML. Blood. 2012. doi:10.1182/blood-2011-06-362541

404 9. Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, et al.
405 Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked
406 between hypermethylated repeats and is frequently altered in cancer. Epigenetics. 2014.
407 doi:10.4161/epi.28323

408 10. Carpenter BL, Zhou W, Madaj Z, DeWitt AK, Ross JP, Grønbæk K, et al. Mother–child
409 transmission of epigenetic information by tunable polymorphic imprinting. Proceedings
410 of the National Academy of Sciences of the United States of America. 2018.
411 doi:10.1073/pnas.1815005115

412 11. Zink F, Magnusdottir DN, Magnusson OT, Walker NJ, Morris TJ, Sigurdsson A, et al.
413 Insights into imprinting from parent-of-origin phased methylomes and transcriptomes.
414 Nature Genetics. 2018. doi:10.1038/s41588-018-0232-7

415 12. Carpenter BL, Remba TK, Thomas SL, Madaj Z, Brink L, Tiedemann RL, et al. Oocyte
416 age and preconceptual alcohol use are highly correlated with epigenetic imprinting of a
417 noncoding RNA (nc886). Proceedings of the National Academy of Sciences of the
418 United States of America. 2021;118. doi:10.1073/pnas.2026580118

419 13. Hanna CW, Peñaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G, et al.
420 Pervasive polymorphic imprinted methylation in the human placenta. Genome
421 Research. 2016;26. doi:10.1101/gr.196139.115

422 14. Sanchez-Delgado M, Court F, Vidal E, Medrano J, Monteagudo-Sánchez A, Martin-
423 Trujillo A, et al. Human Oocyte-Derived Methylation Differences Persist in the Placenta
424 Revealing Widespread Transient Imprinting. PLoS Genetics. 2016;12.
425 doi:10.1371/journal.pgen.1006427

426 15. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al.
427 Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human
428 epiallele responsive to periconceptional environment. Genome Biology. 2015.
429 doi:10.1186/s13059-015-0660-y

430 16. Marttila S, Viiri LE, Mishra PP, Kühnel B, Matias-Garcia PR, Lyytikäinen L-P, et al.
431 Methylation status of nc886 epiallele reflects periconceptional conditions and is
432 associated with glucose metabolism through nc886 RNAs. Clinical Epigenetics.
433 2021;13. doi:10.1186/s13148-021-01132-3

434 17. Fort RS, Garat B, Sotelo-Silveira JR, Duhagon MA. vtRNA2-1/nc886 produces a small
435 RNA that contributes to its tumor suppression action through the microRNA pathway in
436 prostate cancer. Non-coding RNA. 2020. doi:10.3390/ncrna6010007

437 18. Kong L, Hao Q, Wang Y, Zhou P, Zou B, Zhang Y xiang. Regulation of p53 expression
438 and apoptosis by vault RNA2-1-5p in cervical cancer cells. Oncotarget. 2015.
439 doi:10.18632/oncotarget.4948

440 19. Lee K, Kunkeaw N, Jeon SH, Lee I, Johnson BH, Kang GY, et al. Precursor miR-886,
441 a novel noncoding RNA repressed in cancer, associates with PKR and modulates its
442 activity. RNA. 2011. doi:10.1261/rna.2701111

443 20. Lee YS. A Novel Type of Non-coding RNA, nc886, Implicated in Tumor Sensing and
444 Suppression. Genomics & Informatics. 2015. doi:10.5808/gi.2015.13.2.26

445 21. Steegers-Theunissen RPM, Twigt J, Pestinger V, Sinclair KD. The periconceptional
446 period, reproduction and long-term health of offspring: The importance of one-carbon
447 metabolism. Human Reproduction Update. 2013. doi:10.1093/humupd/dmt041

448 22. Gonseth S, Shaw GM, Roy R, Segal MR, Asrani K, Rine J, et al. Epigenomic profiling
449 of newborns with isolated orofacial clefts reveals widespread DNA methylation changes
450 and implicates metastable epiallele regions in disease risk. Epigenetics. 2019.
451 doi:10.1080/15592294.2019.1581591

452 23. van Dijk SJ, Peters TJ, Buckley M, Zhou J, Jones PA, Gibson RA, et al. DNA
453 methylation in blood from neonatal screening cards and the association with BMI and
454 insulin sensitivity in early childhood. International Journal of Obesity. 2018.
455 doi:10.1038/ijo.2017.228

456 24. Yu S, Zhang R, Liu G, Yan Z, Hu H, Yu S, et al. Microarray analysis of differentially
457 expressed microRNAs in allergic rhinitis. American Journal of Rhinology and Allergy.
458 2011. doi:10.2500/ajra.2011.25.3682

459 25. Suojalehto H, Lindström I, Majuri ML, Mitts C, Karjalainen J, Wolff H, et al. Altered
460 microRNA expression of nasal mucosa in long-term asthma and allergic rhinitis.
461 International Archives of Allergy and Immunology. 2014. doi:10.1159/000358486

462 26. Sharbati J, Lewin A, Kutz-Lohroff B, Kamal E, Einspanier R, Sharbati S. Integrated
463 microrna-mrna-analysis of human monocyte derived macrophages upon mycobacterium
464 avium subsp. hominissuis infection. PLoS ONE. 2011.
465 doi:10.1371/journal.pone.0020258

466 27. Asaoka T, Sotolongo B, Island ER, Tryphonopoulos P, Selvaggi G, Moon J, et al.
467 MicroRNA signature of intestinal acute cellular rejection in formalin-fixed paraffin-
468 embedded mucosal biopsies. American Journal of Transplantation. 2012.
469 doi:10.1111/j.1600-6143.2011.03807.x

470 28. Lin CH, Lee YS, Huang YY, Tsai CN. Methylation status of vault rna 2-1 promoter is a
471 predictor of glycemic response to glucagon-like peptide-1 analog therapy in type 2
472 diabetes mellitus. BMJ Open Diabetes Research and Care. 2021;9. doi:10.1136/bmjdrc-
473 2020-001416

474 29. Barker DJP, Osmond C. INFANT MORTALITY, CHILDHOOD NUTRITION, AND
475 ISCHAEMIC HEART DISEASE IN ENGLAND AND WALES. The Lancet. 1986.
476 doi:10.1016/S0140-6736(86)91340-1

477 30. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human
478 Genome Browser at UCSC. Genome Research. 2002;12. doi:10.1101/gr.229102

479 31. Bou-Nader C, Gordon JM, Henderson FE, Zhang J. The search for a PKR code—
480 differential regulation of protein kinase R activity by diverse RNA and protein
481 regulators. RNA. 2019. doi:10.1261/rna.070169.118

482 32. Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: A database for CTCF-binding
483 sites and genome organization. Nucleic Acids Research. 2013;41.
484 doi:10.1093/nar/gks1165

485 33. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al.
486 ENCODE Data in the UCSC Genome Browser: Year 5 update. Nucleic Acids Research.
487 2013;41. doi:10.1093/nar/gks1172

488 34. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser:
489 A web-based browser for visualizing 3D genome organization and long-range chromatin
490 interactions. Genome Biology. 2018;19. doi:10.1186/s13059-018-1519-9

491 35. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI
492 GEO: Archive for functional genomics data sets - Update. Nucleic Acids Research.
493 2013;41. doi:10.1093/nar/gks1193

494 36. Guevara EE, Lawler RR, Staes N, White CM, Sherwood CC, Ely JJ, et al. Age-
495 associated epigenetic change in chimpanzees and humans. Philosophical Transactions
496 of the Royal Society B: Biological Sciences. 2020;375. doi:10.1098/rstb.2019.0616

497 37. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom
498 C, et al. Dynamics of DNA Methylation in Recent Human and Great Ape Evolution.
499 PLoS Genetics. 2013;9. doi:10.1371/journal.pgen.1003763

500 38. Housman G, Quillen EE, Stone AC. Intraspecific and interspecific investigations of
501 skeletal DNA methylation and femur morphology in primates. American Journal of
502 Physical Anthropology. 2020;173. doi:10.1002/ajpa.24041

503 39. Constantinof A, Boureau L, Moisiadis VG, Kostaki A, Szyf M, Matthews SG. Prenatal
504 Glucocorticoid Exposure Results in Changes in Gene Transcription and DNA
505 Methylation in the Female Juvenile Guinea Pig Hippocampus Across Three
506 Generations. Scientific Reports. 2019;9. doi:10.1038/s41598-019-54456-9

507 40. Hannon E, Knox O, Sugden K, Burrage J, Wong CCY, Belsky DW, et al. Characterizing
508 genetic and environmental influences on variable DNA methylation using monozygotic
509 and dizygotic twins. PLoS Genetics. 2018. doi:10.1371/journal.pgen.1007544

510 41. Horvath S, Mah V, Lu AT, Woo JS, Choi OW, Jasinska AJ, et al. The cerebellum ages
511 slowly according to the epigenetic clock. Aging. 2015;7. doi:10.18632/aging.100742

512 42. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, et al. Huntington's
513 disease accelerates epigenetic aging of human brain and disrupts DNA methylation
514 levels. Aging. 2016;8. doi:10.18632/aging.101005

515 43. Andrews S. FastQC - A quality control tool for high throughput sequence data.
516 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Babraham Bioinformatics.
517 2010.

518 44. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for
519 multiple tools and samples in a single report. Bioinformatics. 2016;32.
520 doi:10.1093/bioinformatics/btw354

521 45. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
522 sequence data. Bioinformatics. 2014;30. doi:10.1093/bioinformatics/btu170

523 46. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for
524 Bisulfite-Seq applications. Bioinformatics. 2011;27. doi:10.1093/bioinformatics/btr167

525 47. Hernandez Mora JR, Tayama C, Sánchez-Delgado M, Monteagudo-Sánchez A, Hata K,
526 Ogata T, et al. Characterization of parent-of-origin methylation using the Illumina
527 Infinium MethylationEPIC array platform. Epigenomics. 2018;10. doi:10.2217/epi-
528 2017-0172

529 48. Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, et al.
530 Clustered CTCF binding is an evolutionary mechanism to maintain topologically
531 associating domains. Genome Biology. 2020;21. doi:10.1186/s13059-019-1894-x

532 49. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell
533 identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell.
534 2014;159. doi:10.1016/j.cell.2014.09.030

535

536 Supporting information

537 **S1 Fig. Sequence alignment of the *nc886* gene in the 65 amniota vertebrates Mercator-**
538 **Pecan.** To state that a species presents the *nc886* we required 80% sequence similarity with the
539 human *nc886* and presence of both nc886-5p and nc886-3p short RNAs and the hairpin loop
540 present in the 102nt long nc886 RNA. Species above the dashed line were considered to
541 harbour the full *nc886* gene. Note: Sequence alignment figures are presented in the direction
542 of the gene. Species with no identified alignment in this region were excluded from the figure
543 (Sauropsids, opossum (*Monodelphis domestica*) and platypus (*Ornithorhynchus anatinus*)).

544 **S2 Fig. Sequence alignment of the *nc886* gene in the 24 primates in EPO-extended**
545 **collection.** Sequence similarity to human *nc886* gene decreases as evolutionary distance
546 increases and greatest diverge is seen in the centromeric end of the gene. Note: Sequence
547 alignment figures are presented in the direction of the gene.

548 **S3 Fig. Predicted interaction between *nc886* telomeric CTCF binding site and CTCF**
549 **binding site located near IL9.** A) ChIA-PET data and in HI-C data in Genome browser cell
550 lines and in 3D genome browser. B) HUVEC and C) K562 cell lines. The suggested sub-TAD
551 has been indicated with a black square and interactions in HI-C data have been circumscribed.
552 The telomeric CTCF binding site is located at chr5:135418124-135418523 and the CTCF
553 binding site near *IL9* at chr5:135223050-135223420.

554 **S4 Fig. Methylation of *nc886* region in hippocampi of A) guinea pigs (GSE109765) and**
555 **B) humans (GSE64509).** In (A) each circle presents a CpG site with measured methylation
556 level and each line one individual guinea pig (n=36). The number of reads per site is low (on
557 average 7) but none of the samples provide data that would indicate anything but non-
558 methylated DNA methylation status around the guinea pig *nc886* gene. In human hippocampi
559 (B) the methylation pattern of *nc886* presents the expected binomial methylation pattern, which
560 is in line with reported frequency of 25% non-methylated individuals and 75% individuals with
561 monoallelic methylation in a population.

562 **S5 Fig. Methylation pattern of Paternally expressed 10 (*PEG10*) in blood of great apes.**
563 Only probes locating in sites with no clear sequence differences as compared to the human
564 *PEG10* sequence are shown.

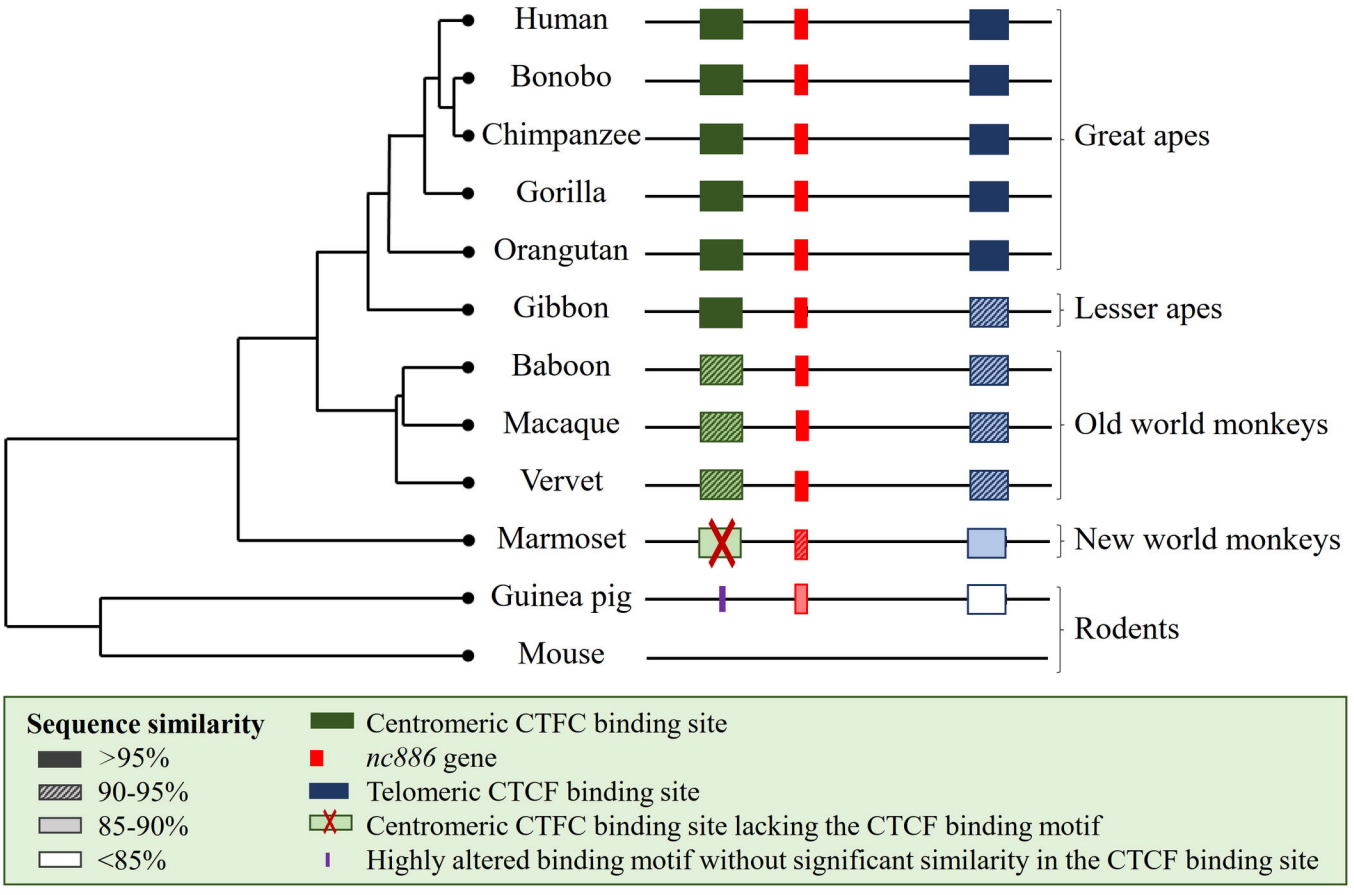
565 **S6 Fig. Methylation pattern of Paternally expressed 10 (*PEG10*) in femur bone of humans**
566 **and Old and New World monkeys.** Only probes locating in sites with no clear sequence
567 differences as compared to the human *PEG10* sequence are shown.

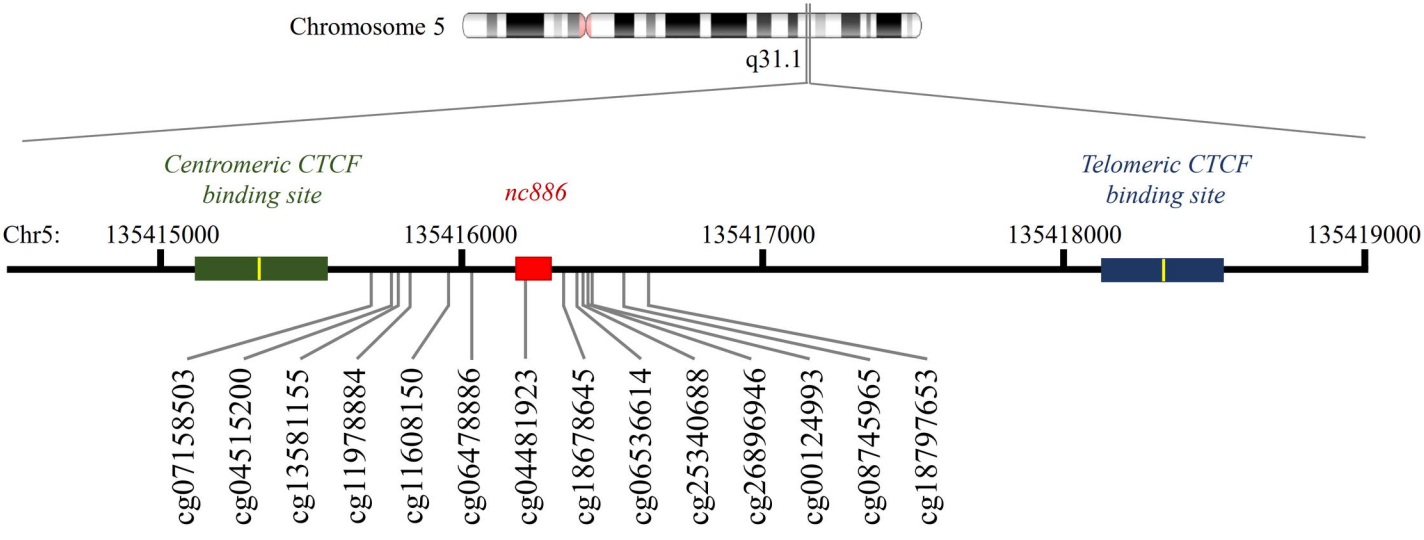
568 **S1 File. Sequence alignment of the *nc886* region between human, bonobo, chimpanzee,**
569 **gorilla, orangutang, gibbon, baboon, macaque, vervet and marmoset.** The sequence
570 comparison covers the region containing the 14 CpG sites presenting bimodal methylation
571 pattern in humans (chr5:135415593-135416666). These CpG sites are highlighted in yellow.

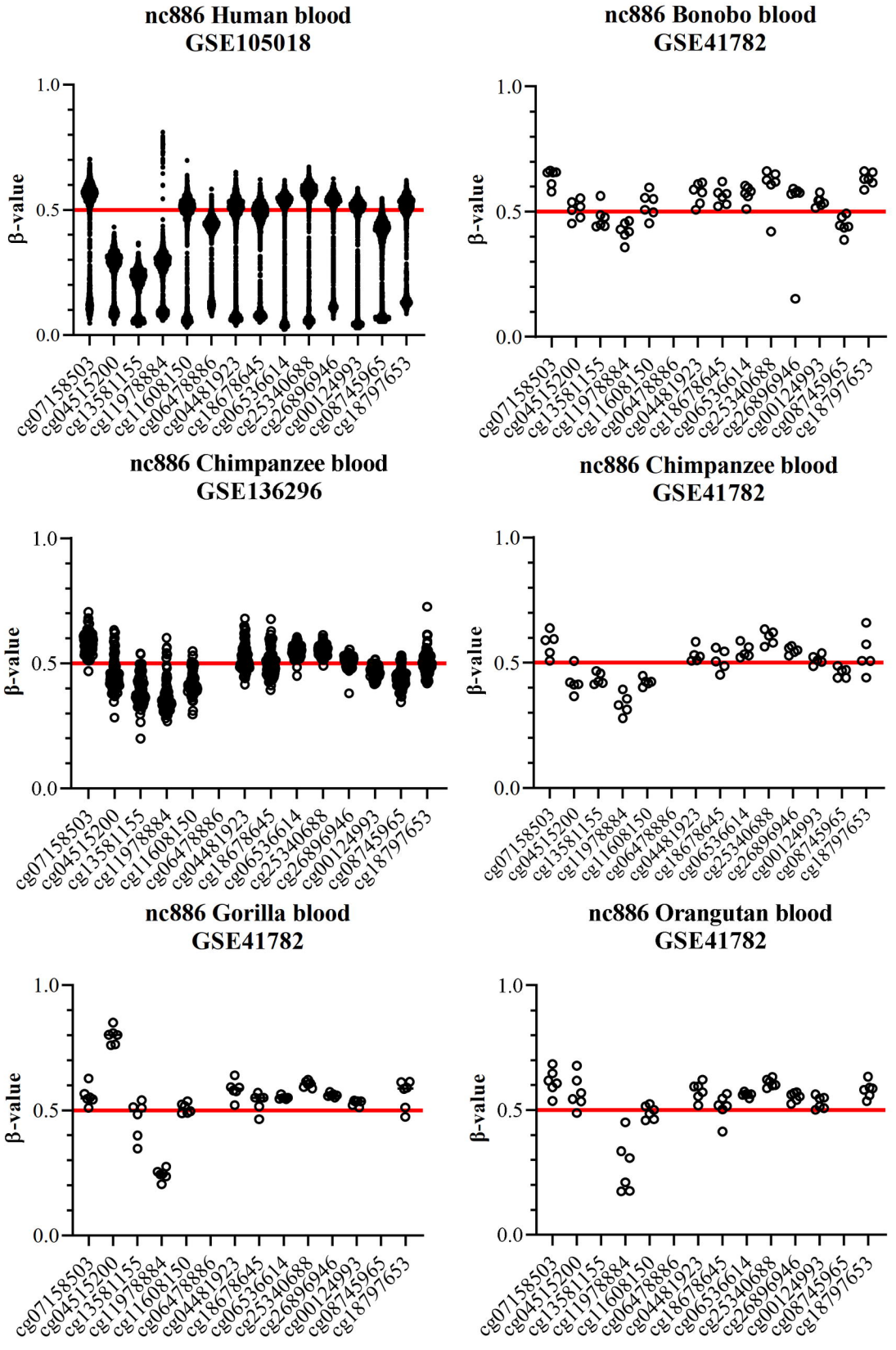
572 **S1 Table. Sequence similarities of *nc886* gene and its centromeric and telomeric CTCF**
573 **binding sites, and the CTCF binding site located near Il9 gene in primates, marmot,**
574 **squirrel, guinea pig and in mouse investigated from NCBI BLAST and ensemble**.

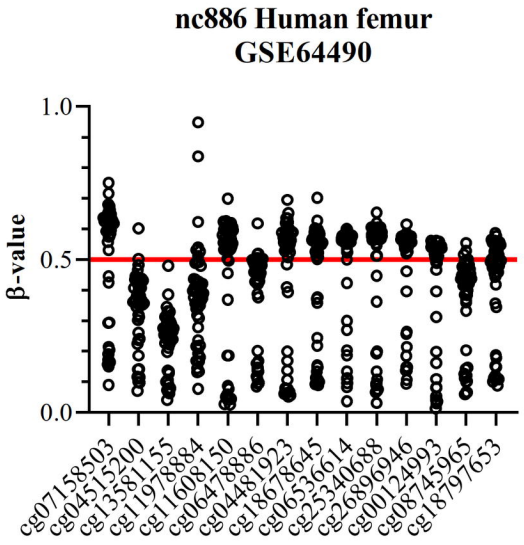
575 **S2 Table. Methylation medians and standard deviations of the PEG10 and *nc886* regions**
576 **and individual probes in the *nc886* region in primates.** Medians and standard deviations in

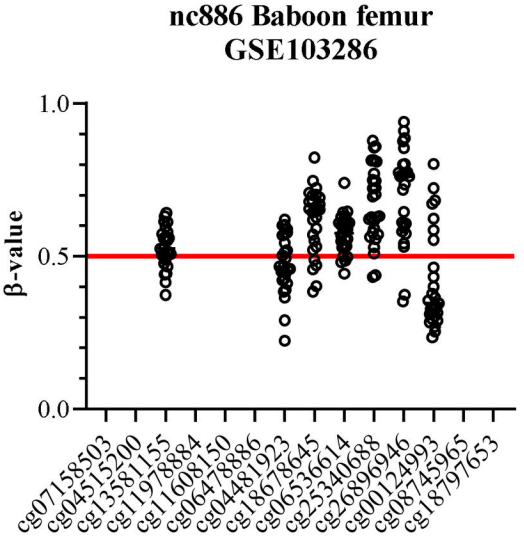
577 human *nc886* are not comparable to other species, as it is known to demonstrate bimodal
578 methylation pattern in humans.

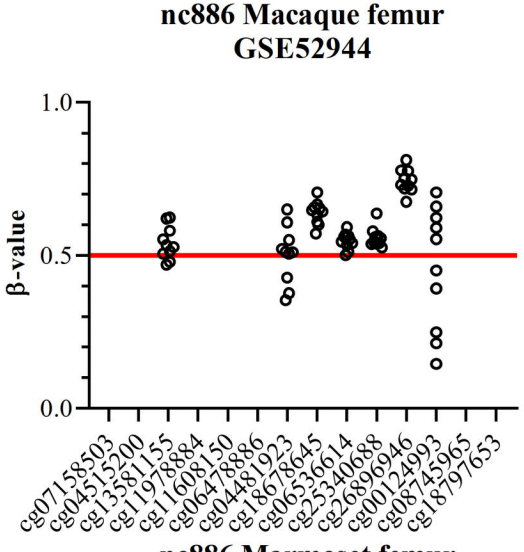

Human
Bonobo
Chimpanzee
Gorilla
Orangutan
Gibbon
Baboon
Macaque
Vervet
Marmoset
Guinea pig
Mouse
Great apes
Lesser apes
Old world monkeys
New world monkeys
Rodents
Sequence similarity
>95%
90-95%
85-90%
<85%
Centromeric CTFC binding site
nc886 gene
Telomeric CTCF binding site
Centromeric CTFC binding site lacking the CTCF binding motif
Highly altered binding motif without significant similarity in the CTCF binding site

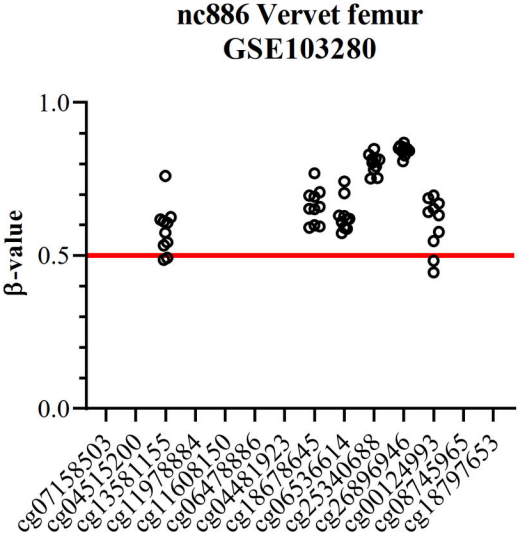

Chromosome 5
q31.1
Centromeric CTCF binding site
nc886
Telomeric CTCF binding site
Chr5:
135415000
135416000
135417000
135418000
135419000
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653

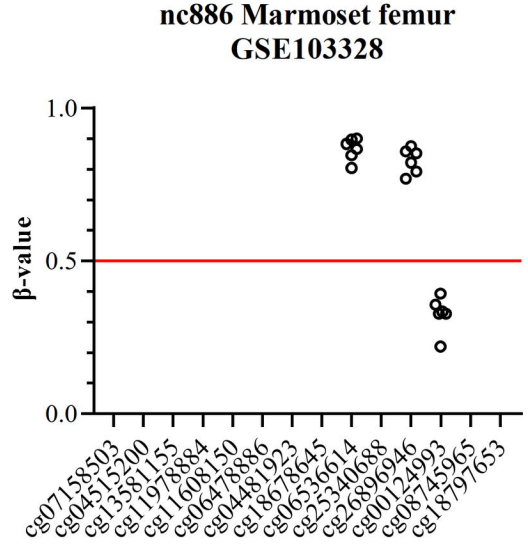

nc886 Human blood
GSE105018
nc886 Bonobo blood
GSE41782
nc886 Chimpanzee blood
GSE136296
nc886 Chimpanzee blood
GSE41782
nc886 Gorilla blood
GSE41782
nc886 Orangutan blood
GSE41782
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653


nc886 Human femur
GSE64490
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653


nc886 Baboon femur
GSE103286
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653


nc886 Macaque femur
GSE52944
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653


nc886 Vervet femur
GSE103280
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653


nc886 Marmoset femur
GSE103328
β-value
1.0
0.5
0.0
cg07158503
cg04515200
cg13581155
cg11978884
cg11608150
cg06478886
cg04481923
cg18678645
cg06536614
cg25340688
cg26896946
cg00124993
cg08745965
cg18797653