An equine Endothelin 3 cis-regulatory variant links blood pressure modulation to elite racing performance

Endothelin 3 affects blood pressure regulation and performance in horses

Kim Fegraeus[1]*, Maria K Rosengren[2]*, Rakan Naboulsi[2]*, Ludovic Orlando[3], Magnus Åbrink[4], Annika Thorsell[5], Ahmad Jouni[2], Brandon D Velie[6], Amanda Raine[1], Beate Egner[7], C Mikael Mattsson[8], Göran Andersson[2], Jennifer R.S Meadows[9] & Gabriella Lindgren[2,10]

[1]Department of Medical Sciences, Science for life laboratory, Uppsala University, Sweden

[2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences Uppsala, Sweden

[3]Centre d'Anthropobiologie et de Génomique de Toulouse (CNRS UMR 5288), Université Paul Sabatier, Toulouse, France

[4]Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden

[5]Proteomics Core Facility, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

[6]School of Life & Environmental Sciences, University of Sydney, Sydney, Australia

[7]Department of Cardio-Vascular Research, Veterinary Academy of Higher Learning, Babenhausen, Germany

[8]Silicon Valley Exercise Analytics (svexa), MenloPark, CA, USA

[9]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

[10]Center for Animal Breeding and Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium

Kim.fegraeus@medsci.uu.se, Maria.rosengren@slu.se, Rakan.naboulsi@slu.se, ludovic.orlando@univ-tlse3.fr, Magnus.abrink@slu.se, annika.thorsell@gu.se,

28 ahmad.mjouni@gmail.com, brandon.velie@sydney.edu.au, amanda.raine@medsci.uu.se,
29 beate.egner@vbsgroup.eu, mikael@svexa.com, Goran.Andersson@slu.se,
30 jennifer.meadows@imbim.uu.se, Gabriella.Lindgren@slu.se
31 *Equal contribution
32 Corresponding authors kim.fegraeus@medsci.uu.se, gabriella.lindgren@slu.se

33 Abstract

34 A previous selective sweep analysis of horse racing performance revealed a 19.6 kb candidate
35 region approximately 50 kb downstream of the Endothelin 3 (*EDN3*) gene. EDN3 and other
36 endothelin family members are associated with blood pressure regulation in humans and other
37 species, but similar association studies in horses are lacking. We hypothesized that the sweep
38 region includes a regulatory element acting on *EDN3* transcription, ultimately affecting blood
39 pressure regulation and athletic performance in horses. Selective sweep fine-mapping
40 identified a 5.5 kb haplotype of 14 SNPs shared within Coldblooded trotters (CBT) and
41 Standardbreds (SB). Most SNPs overlapped potential transcription factor binding sites, and
42 haplotype analysis showed significant association with all tested performance traits in CBTs
43 and earnings in SBs. From those, two haplotypes were defined: an elite performing haplotype
44 (EPH) and a sub-elite performing haplotype (SPH). While the majority of SNPs in the
45 haplotype were part of the standing variation already found in pre-domestication horses, there
46 has been an increase in the frequencies of the alternative alleles during the whole history of
47 horse domestication. Horses homozygous for EPH had significantly higher plasma levels of
48 EDN3, lower levels of EDN1, and lower exercise-related blood pressure compared to SPH
49 homozygous horses. Additionally, a global proteomic analysis of plasma from EPH or SPH
50 homozygous horses revealed higher levels of proteins involved in pathways related to immune
51 response and complement activation in the SPH horses. This is the first study to demonstrate
52 an association between the *EDN3* gene, blood pressure regulation, and athletic performance in

horses. The results advance our understanding of the molecular genetics of athletic performance, exercise-related blood pressure regulation, and biological processes activated by intense exercise.

Keywords: Athletic performance, Blood pressure, Endothelin, *Equus caballus*, Exercise, Genetics, Harness racing, Horse, Trotting

Author summary

The horse is one of the most common species used for studying athletic performance. For centuries, horses have been used by humans for transportation, agriculture and entertainment and this has resulted in selection for various traits related to athletic performance. A previous study discovered that a genetic region close to the *Endothelin3* gene was associated with harness racing performance. Endothelin3 is known to be involved in blood pressure regulation and therefore we hypothesized that this region influences blood pressure and racing performance in horses. In this study we have used additional horses and fine-mapped the candidate region and we also measured blood pressure in Coldblooded trotters during exercise. Horses with two copies of the elite-performing haplotype had higher levels of Endothelin3 in plasma, lower blood pressure and better racing performance results, compared to horses with two copies of the sub-elite performing haplotype. We also discovered that horses with the sub-elite performing haplotype had higher levels of proteins related to the immune system in plasma. This study is the first to link Endothelin3 to blood pressure regulation and performance in horses. It broadens the understanding of the biological mechanisms behind blood pressure regulation as well as inflammation and coagulation system in relation to racing performance.

Background

The use of domestic animals as models for genomic research has provided not only basic knowledge concerning gene function and biological mechanisms but also a complementary view on genotype–phenotype relationships (1). The horse is one of the most popular species used to study athletic performance. They have been intensively selected for centuries, to engender the optimal physical capacity for strength, speed, and endurance (2,3). Additionally, their recent population history, involving closed populations with similar phenotypic traits within breeds and large variations across breeds, has created a favorable genome structure for genetic mapping. These factors combined make the horse an optimal model for studying the molecular genetics underlying athletic performance and the complex biological processes activated by exercise (1,2,4,5). Previous work in horses has started to reveal the effect of genes linked to complex traits such as muscle mass and locomotion patterns on athletic performance, and many potential candidate genetic variants have been identified (6–12). However, understanding the mechanisms by which the identified variants exert their functions is much more challenging.

A selective sweep study on harness racing breeds revealed a 19.6 kb genomic region on chromosome 22, located about 50 kb downstream of the *Endothelin 3* (*EDN3*) gene, under selection in Coldblooded trotters (CBTs) (13). Five SNPs in high linkage disequilibrium (LD r^2 = 0.92–0.94) were significantly associated with racing performance, including the number of victories, earnings, and racing times. SNP NC_009165.3:g.46717860C>T (EquCab3.0) showed the strongest association within the sweep and was further genotyped in 18 additional horse breeds. The favorable T allele was found in high frequency in breeds used for racing, while it generally remained at low frequency in ponies and draught horses. It was hypothesized

100 that the identified region might contain a regulatory element influencing either the expression
101 of *EDN3* or other genes nearby (13).
102
103 The endothelin system encompasses the EDN ligands (e.g., EDN-1 to -3) and their G protein-
104 coupled receptors (endothelin receptor type A and B). Together, they are involved in a broad
105 range of biological functions, including cell proliferation and vasoconstriction (14). EDN2 and
106 -3 can be considered as natural antagonists for EDN1 (15). The endothelin receptor type A
107 (EDNRA) has the highest affinity for EDN1 and EDN2 but a low affinity for EDN3 (16). The
108 endothelin receptor type B (EDNRB) has equal affinity for all three isoforms of endothelin
109 (17). During exercise, endothelins are involved in the dynamic redistribution of blood flow to
110 the lung and active muscles, limiting the flow to non-exercise-essential tissues, such as kidneys
111 and intestines (15). Generally, EDNRA activation causes vasoconstriction and increased blood
112 pressure, while EDNRB activation results in nitric oxide release, vasodilatation, and decreased
113 blood pressure (18).
114
115 To date, most studies of the EDN system have focused on the role of EDN1. In human health
116 studies, *EDN1* genetic variants were found to be implicated in hypertension (19). In
117 performance trials, increased EDN1 plasma levels were observed in relation to exercise
118 duration (20), with elevated blood pressure leading to impaired exercise capacity in elite
119 athletes (20,21). There have also been performance studies in horses, where EDN1 plasma
120 levels largely mirrored the results found in humans, showing that the EDN1 response to
121 exercise challenge is conserved across both species (22). There is a clear role for
122 vasoconstriction in exercise performance, but the action of the EDN3 ligand is unclear. In this
123 study, we aimed to fill this gap of knowledge. First through the fine-mapping of the *EDN3*
124 region flanking the selective sweep, then the characterization of the spatial and temporal

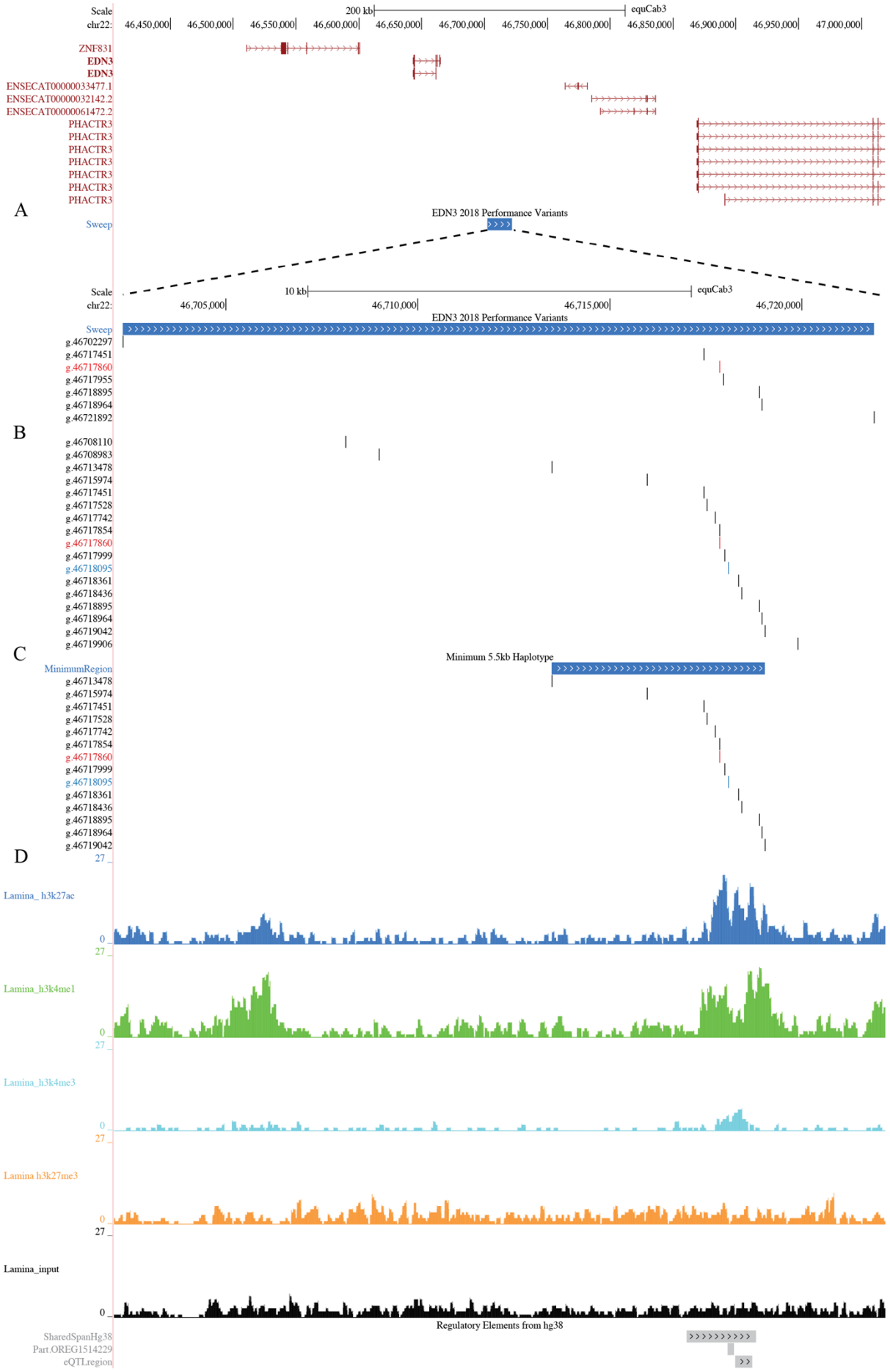
125 distribution of key genetic variants within horse populations, and finally by evaluating the
126 effect of opposite haplotypes on blood pressure, as well as EDN1 and EDN3 plasma levels,
127 before and during exercise. We also performed a global proteomic study on plasma to identify
128 differentially expressed proteins amongst the various *EDN3* haplotypes. While the key breed
129 of the study was the CBT, a popular harness racing breed in Sweden and Norway, the results
130 are placed in the context of additional racing breeds. The results presented here on the
131 endothelin pathway not only give new insights into genetic regulation of its established
132 vasomodulatory properties, but also its emerging role in inflammation.
133

134 ## Results

135 ### Fine-mapping and comparative analyses suggest a regulatory role of the selected
136 region

137 A 2018 study using the sparse Illumina SNP50 Genotyping BeadChip, identified a 19.6 kb
138 selective sweep shared between SBs and CBTs (ECA22:46,702,297-46,721,892; Figure 1A).
139 The sweep was located between *EDN3* and ENSECAG00000039543 (lncRNA), and five SNPs
140 within this region were found to be associated with harness racing performance (13). From
141 those five SNPs, g.46717860 C>T was the most significant. With a four-step process, we
142 reduced this 19.6 kb sweep region to a minimal 5.5 kb region (Figure 1).
143
144 In step one, we performed a sweep-performance association analysis using additional horse
145 material (n=661) and SNP genotypes extracted from the 670K Axiom Equine Genotyping
146 Array (Figure 1A). However, supplementing the 400 CBTs used originally in the 2018
147 study(13) with 221 additional CBTs did not provide increased resolution into the 19.6 kb
148 sweep. Seven SNPs within the original sweep span were extracted from the array, including
149 g.46717860. However, pairwise LD ($r^2 > 0.6$) to g.46717860 resulted in the same five SNP set

150 as found in (13), with all SNPs remaining significantly associated with harness racing
151 performance traits in linear models (i.e., number of victories, earnings, and race times, $P \leq$
152 0.05) (Supplementary Table S1).


A
Scale
chr22:
200 kb
equCab3
46,450,000
46,500,000
46,550,000
46,600,000
46,650,000
46,700,000
46,750,000
46,800,000
46,850,000
46,900,000
46,950,000
47,000,000
ZNF831
EDN3
EDN3
ENSECAT00000033477.1
ENSECAT00000032142.2
ENSECAT00000061472.2
PHACTR3
PHACTR3
PHACTR3
PHACTR3
PHACTR3
PHACTR3
PHACTR3
Sweep
EDN3 2018 Performance Variants
B
Scale
chr22:
10 kb
equCab3
46,705,000
46,710,000
46,715,000
46,720,000
EDN3 2018 Performance Variants
Sweep
g.46702297
g.46717451
g.46717860
g.46717955
g.46718895
g.46718964
g.46721892
C
g.46708110
g.46708983
g.46713478
g.46715974
g.46717451
g.46717528
g.46717742
g.46717854
g.46717860
g.46717999
g.46718095
g.46718361
g.46718436
g.46718895
g.46718964
g.46719042
g.46719906
D
Minimum 5.5kb Haplotype
MinimumRegion
g.46713478
g.46715974
g.46717451
g.46717528
g.46717742
g.46717854
g.46717860
g.46717999
g.46718095
g.46718361
g.46718436
g.46718895
g.46718964
g.46719042
27
Lamina_h3k27ac
0
27
Lamina_h3k4me1
0
27
Lamina_h3k4me3
0
27
Lamina h3k27me3
0
27
Lamina_input
0
Regulatory Elements from hg38
SharedSpanHg38
Part.OREG1514229
eQTLregion

153

154 **Figure 1**. A. Overview of the 19.6 kb 2018 selective sweep region including all annotated genes in the
155 region (13). B. Zoomed in view of the 2018 selective sweep region including the 17 SNPs used in the
156 haplotype analysis. C. Overview of the 5.5 kb minimum shared haplotype within the Coldblooded
157 trotters and Standardbreds, including all SNPs from the haplotype analysis. D. Horse functional data as
158 generated by chip-seq of histone modifications from hoof lamina (from Equine Functional Annotation
159 of Animal Genomes, FAANG) indicating an active enhancer in our region of interest. In addition, ~ 2
160 kb of the associated haplotype region aligned with the human reference, including segments of an
161 ORegAnno regulatory element, pinpointing a curated regulatory annotation of this region,
162 (OREG1514229), GTEx cis-eQTL variants regulating *EDN3* in esophagus mucosa tissue.
163
164 In step two, we performed variant discovery to increase variant density across the 19.6 kb
165 sweep region. Here, Illumina short-read whole genome sequences (WGSs) were generated
166 from two CBTs and two SBs (the most common breed used for harness racing). In each breed,
167 the horses were selected to be homozygous for different alleles at SNP g.46717860 (CC and
168 TT, respectively) and to have either high (TT) or low (CC) earnings per start (Supplementary
169 Table S2). From the 19.6 kb sweep region, 78 SNPs and six indels (one 400 bp deletion, one
170 23 bp deletion, three single bp deletions, and one single bp insertion) were identified. To
171 possibly detect new structural variants, targeted long-read sequencing using Oxford Nanopore
172 technology (ONT) was performed across the sweep region for eight horses: four CBTs with
173 g.46717860-TT genotype and high performance, and four CBTs with the opposite features,
174 including the two CBTs used for WGS (Supplementary Table S2). However, no new structural
175 genetic variants were identified.
176
177 In step three, we prioritized variants for further analysis. While the whole genome sequenced
178 elite CBT (high earnings per start) was homozygous for the 400 bp deletion
179 (ECA22:46,714,602-46,715,003), the elite SB was heterozygous, and the deletion was not

180 detected in the low-performing (low earnings per start) horses. The deletion was, thus, tested
181 for association with racing performance traits in 479 CBTs using linear models, but the results
182 were not significant (Supplementary Table S3). From the remaining variants identified, a total
183 of 24 SNPs and one single base pair deletion, evenly distributed within the region, were
184 selected (see methods) and genotyped in four donkeys and 412 horses sampled from 13
185 different breeds. SNP g.46717860 was not included in this set but was genotyped separately in
186 170 horses and imputed in the remainder. At this stage, 24 SNPs and 407 horses passed quality
187 control measures. Pairwise LD estimation with SNP g.46717860 revealed 15 SNPs in strong
188 LD ($r^2 \geq 0.53$, 14 SNPs with an r^2 score > 0.98) (ECA22: 46,708,983-46,719,042). LD decayed
189 outside the region ($r^2 < 0.07$), leaving 17 SNPs (15 SNPs in LD and two flanking SNPs) for
190 haplotype analysis (Figure 1B). Phasing revealed 15 different haplotypes, five showing a
191 frequency > 2 %. All haplotypes with a frequency > 5 % in the total sample set or within each
192 breed are presented in Table 1 (disregarding breeds represented by less than five horses).

193 Table 1. Haplotype frequencies in different breeds

Haplotype (ECA22: 46,708,110-46,719,906)	All breeds (n=411)	CBT (n=222)	Arabians (n=29)	TB (n=30)	SB (n=45)	Ardennes (n=20)	NSH (n=18)	Exmoor (n=17)	Icelandic (n=7)
1) GCAAGCGTTTCTCAAAA	0.30	0.27	0.74	0.68	0.36	0.00	0.03	0.00	0.00
2) GTGGTTACCCTCTGGGA	0.28	0.25	0.00	0.07	0.00	0.60	0.86	1.00	0.00
3) ACAAGCGTTTCTCAAAG	0.27	0.32	0.26	0.25	0.41	0.10	<0.01	0.00	0.00
4) ACGGTTACCCTCTGGGG	0.11	0.14	0.00	0.00	0.20	0.00	0.01	0.00	0.43
5) ACGGTTACCCTCTGGGA	0.03	0.00	0.00	0.00	0.04	0.18	0.03	0.00	0.36
6) ACAAGCGTTTCTCGAAA	<0.01	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
7) ACGGGCGTTCCTCAAAG	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
8) GCAAGCGTTTCTCAAGG	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
9) GCAAGTGTTTCTCAAGG	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07

194 [1] Order of SNPs. Chr 22: 46708110, 46708983, 46713478, 46715974, 46717451, 46717528,
195 46717742, 46717854, 46717860, 46717999, 46718095, 46718361, 46718436, 46718895, 46718964,
196 46719042, 46719906

197 CBT: Coldblooded trotter, TB: Thoroughbred, SB: Standardbred, NSH: North-Swedish draught horse

198

199 The association between haplotype and harness racing performance traits was performed in
200 CBTs (n=180) and SBs (n=38) using generalized linear model regression (GLM) analysis in
201 the statistical software R (23) (Table 2 and 3). Only haplotypes with a frequency > 2 % in the
202 population were included in the analysis. Of the four haplotypes in CBTs, haplotype 3 (H3),
203 carrying the g.46717860-T high performance associated allele, was the most common (0.32)
204 and was defined as the base haplotype. Two haplotypes, H2 and H4, respectively, demonstrated
205 a significant negative effect on all performance traits tested, and each carried the g.46717860-
206 C low performance associated allele (Table 2). Four haplotypes were also available for testing
207 within SBs with performance data. Three haplotypes were shared with CBTs (H1, H3 and H4)
208 and one (H5) was shared with other breeds (Table 3). As for the CBTs, H3 was the most
209 common haplotype (0.43) and was defined as the base haplotype in this analysis. Two
210 haplotypes demonstrated significant negative effects on performance traits: H4 (earnings and
211 earnings per start) and H5 (all traits tested). The above analysis revealed a minimum 5,564 bp,
212 14 SNP shared haplotype, significantly associated with racing performance within CBTs
213 (ECA22: 46,713,478-46,719,042). The same variants were also significantly associated with
214 racing performance in SBs. This allowed for the definition of an elite-performing haplotype
215 (EPH: AAGCGTTTCTCAAA), and a sub-elite-performing haplotype (SPH:
216 GGTTACCCTCTGGG) (Supplementary Table S4) (Figure 1C). Of note, the reference
217 EquCab3.0 genome represents the SPH variant.

218

219

220

221

Table 2. Haplotype frequencies, haplotype coefficient and *P*-values for the performance analysis in Coldblooded trotters (n=180)

	Haplotype[2] (H3)		**Haplotype (H1)**		**Haplotype (H2)**		**Haplotype (H4)**	
Haplotype[1]	ACAAGCGTTTCTCAAAG		GCAAGCGTTTCTCAAAA		GTGGTTACCCTCTGGGA		ACGGTTACCCTCTGGGG	
Frequency	0.32		0.27		0.27		0.14	
	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]
No. of starts[5]	-	-	-0.30	0.43	-1.00	**0.004**	-1.05	**0.02**
No. of wins[5]	-	-	-0.04	0.31	-0.14	**<0.001**	-0.17	**<0.001**
No. of placings[5]	-	-	-0.06	0.13	-0.12	**<0.001**	-0.14	**0.002**
Earnings (SEK)[4,5]	-	-	-0.31	0.05	-0.53	**<0.001**	-0.56	**0.002**
Earnings/start[5]	-	-	-0.19	0.07	-0.45	**<0.001**	-0.42	**<0.001**
Best time (sec/km)[5]	-	-	0.03	0.16	0.07	**<0.001**	0.08	**0.002**

[1]Order of SNPs. Chr 22: 46708110, 46708983, 46713478, 46715974, 46717451, 46717528, 46717742, 46717854, 46717860, 46717999, 46718095, 46718361, 46718436, 46718895, 46718964, 46719042, 46719906

[2]The most common haplotype was used as the base haplotype

[3]A GLM regression analysis was performed in R. Sex, age, birth country, number of starts (when applicable) and *DMRT3* genotype were included as fixed effects. Significant results (P≤ 0.05) in bold

[4] SEK, Swedish kronor

[5] Log transformed values (log10 +1) were used for wins and placings. Number of starts were square root transformed. Earnings and earnings per start were transformed applying ln(Earnings + 1000).

CBT: Coldblooded trotter, TB: Thoroughbred, SB: Standardbred, NSH: North-Swedish draught horse

240 Table 3. Haplotype frequencies, haplotype coefficient and *P*-values for the performance analysis in

241 Standardbreds (n=38)

	Haplotype[2] (H3)		**Haplotype (H1)**		**Haplotype (H4)**		**Haplotype (H5)**	
Haplotype[1]	ACAAGCGTTTCTCAAAG		GCAAGCGTTTCTCAAAA		ACGGTTACCCTCTGGGG		ACGGTTACCCTCTGGGA	
Frequency	0.43		0.34		0.18		0.04	
	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]	Haplotype coefficient	P-value[3]
No. of starts[5]	-	-	0.12	0.82	0.48	0.41	-0.61	**0.003**
No. of wins[5]	-	-	-0.02	0.74	-0.08	0.35	-0.47	**<0.001**
No. of placings[5]	-	-	-0.03	0.63	-0.11	0.06	-0.25	**<0.001**
Earnings (SEK)[4, 5]	-	-	-0.17	0.53	-1.42	**<0.001**	-1.32	**<0.001**
Earnings/start[5]	-	-	-0.10	0.66	-1.14	**<0.001**	-1.12	**<0.001**
Best time (sec/km)[5]	-	-	-0.10	0.37	0.15	0.24	0.65	**0.01**

242 [1]Order of SNPs. Chr 22: 46708110, 46708983, 46713478, 46715974, 46717451, 46717528,

243 46717742, 46717854, 46717860, 46717999, 46718095, 46718361, 46718436, 46718895, 46718964,

244 46719042, 46719906

245 [2]The most common haplotype was used as the base haplotype

246 [3]A GLM regression analysis was performed in R. Sex, age, birth country and number of starts (when

247 applicable) were included as fixed effects. Significant results (P≤ 0.05) in bold

248 [4] SEK, Swedish kronor

249 [5] Log transformed values (log10 +1) were used for wins and placings. Number of starts were square

250 root transformed. Earnings and earnings per start were transformed applying ln(Earnings + 1000).

251 CBT: Coldblooded trotter, TB: Thoroughbred, SB: Standardbred, NSH: North-Swedish draught horse

252

253 In step four, we assessed the potential functional impact of the minimal 5.5 kb region and each

254 of the 14 SNP haplotype variants. Using horse functional data from two Thoroughbreds (from

255 the Equine Functional Annotation of Animal Genomes, FAANG), we found three SNPs

256 (g.46717860, g.46718095 and g.46718113) that overlapped potential active enhancers, i.e.,

257 H3K4me1 and H3K27ac marks (24,25) (Figure 1D). Both Thoroughbreds were homozygous

258 for EPH, with the histone marks indicating the presence of an active enhancer in the hoof
259 lamina, present but less active in lung and ovarian tissue, and repressed in the skin(24,25). In
260 comparative analyses, a genome liftover to hg38 indicated that, although not well conserved,
261 ~ 2 kb of the associated haplotype region aligned with the human reference, including segments
262 of an ORegAnno regulatory element (OREG1514229), GTEx cis-eQTL variants regulating
263 *EDN3* in esophagus mucosa tissue (Figure 1D). These combined regulatory signals prompted
264 us to test the potential for transcription factor binding affinity changes in relation to the EPH
265 and SPH alleles. Most SNP allele changes (12 out of 14 SNPs) altered transcription factor
266 binding (Supplementary Table S4). For EPH, this meant the perturbation of matrices for stress
267 responders, e.g., IRF2 and NFKB1, while creating sites for developmental transcription factors
268 such as NR4A2 and HOXA5.
269
270 **The elite performing haplotype has an additive effect on performance traits.**
271 Given that all 14 associated variants within the minimum shared 5.5 kb haplotype were in
272 perfect LD, we used the genotypes at g.46718095 as a proxy for EPH (C allele) and SPH (T
273 allele) haplotypic pairs and re-assessed the association between haplotypes and racing
274 performance. For CBTs (n=539), EPH homo- or heterozygotes outperformed the SPH
275 homozygotes (Table 4). In SBs, there were no statistically significant differences between the
276 three genotypes (Table 5). However, when analyzing horses carrying at least one T allele versus
277 CC horses, the results were significant for several performance traits (Table 5), including the
278 number of wins, placings and earnings. However, the number of wins and placings was higher
279 in horses with at least one T allele, while the CC horses earned more money than the TT/TC
280 horses (Table 5).
281
282

283 Table 4. Performance results in 539 Coldblooded trotters for SNP g.46718095 T>C

Genotype	TT (n=65)		TC (n=256)		CC (n=218)		P[2]		
Performance trait[1]	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**TT/TC**	**TT/CC**	**TC/CC**
No. of starts	31.9	22.0	40.2	28.0	41.3	31.0	**0.01**	**0.02**	1.00
No. of wins	2.8	1.0	5.7	3.0	5.9	3.0	**0.003**	**0.02**	0.65
No. of placings (1-3)	9.3	4.0	14.2	9.0	14.5	9.0	**0.005**	0.07	0.38
Earnings (SEK)	207,295	51,772	438,700	140,161	453,911	115,312	**0.006**	**0.05**	0.58
Earnings per start (SEK)	4,752	2,976	9,666	4,362	8,469	3,771	**0.003**	0.08	0.20
Best km time (sec)[3]	90.9	91.4	88.5	88.3	88.9	88.9	**0.001**	**0.02**	0.44

284 [1]Log transformed values (log10 +1) were used for wins and placings. Earnings and earnings per start
285 were transformed applying ln(earnings + 1000). Square root transformations were used for number of
286 starts
287 [2]A linear model was performed in R. Sex, age, birth country, number of starts and *DMRT3* genotype
288 were included as fixed effects (when significant). Least square means and Tukey were used as post
289 hoc tests. Significant results ($P \leq 0.05$) in bold
290 [3]Only races where the horse did not gallop were included
291
292
293
294
295
296
297
298
299
300

301 Table 5. Performance results in 271 Standardbreds for SNP g.46718095 T>C

Genotype	**TT (n=7)**		**TC (n=62)**		**CC (n=202)**		**P[2]**			
Perf. trait[1]	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**TT/TC**	**TT/CC**	**TC/CC**	**CC/ TT+TC**
No. of starts	43.0	48.0	42.9	38.0	35.6	31.0	0.85	0.60	0.52	0.20
No. of wins	6.7	5.0	6.3	4.0	5.7	5.0	0.82	0.99	0.10	**0.04**
No. of placings (1-3)	15.6	12.0	16.1	16.0	13.7	11.0	0.86	0.96	0.08	**0.03**
Earnings (SEK)	700,008	322,300	730,124	332,300	1,075,529	369,675	0.92	0.43	0.06	**0.01**
Earnings per start (SEK)	11,981	5,463	16,673	9,765	35,346	9,897	0.86	0.44	0.16	**0.04**
Best km time (sec)[3]	73.1	73.9	73.3	72.8	73.1	72.9	0.85	1.00	0.19	0.10

302 [1]Log transformed values (log10 +1) were used for wins and placings. Earnings and earnings per start
303 were transformed applying ln(earnings + 1000). Square root transformations were used for number of
304 starts
305 [2]A linear model was performed in R. Sex, age, birth country and number of starts were included as
306 fixed effects (when significant). Least square means and Tukey were used as post hoc tests.
307 Significant results ($P \leq 0.05$) in bold
308 [3]Only races where the horse did not gallop were included
309
310 Our racing performance analyses revealed that the g.46718095 genotypes did not distribute
311 according to HWE. This prompted us to ask if there had been selection on this position or

312 perhaps the minimal haplotype in other breeds. While each individual's performance status is
313 unknown, a trend for increased frequency of the C allele in traditional performance breeds was
314 observed, e.g., Arabian horses, Thoroughbreds and Warmbloods. In contrast, this allele was
315 low or absent in draft horses and ponies (Table 6).
316
317 Table 6. Allele frequency of SNP g.46718095 T>C

Breed	**n**	**T**	**C**
Arabian horses	30	0.00	1.00
Thoroughbred	30	0.07	0.93
Standardbred	306	0.14	0.86
Warmblood	3	0.17	0.83
Coldblooded trotters	539	0.36	0.64
Finnhorses	4	0.50	0.50
Gotland pony	2	0.50	0.50
Shetland pony	18	0.64	0.36
Ardennes	20	0.78	0.22
Icelandic horses	11	0.86	0.14
North-Swedish draught	25	0.88	0.12
Exmoor pony	23	1.00	0.00
Fjord horse	1	1.00	0.00
Przewalski	4	1.00	0.00

318
319

320 **Increased frequency of the favorable haplotype coincides with the time of horse**
321 **domestication**

322 For all 14 SNPs within the minimum shared haplotype in the trotters, we investigated the allele
323 frequency over time using the mapDATAge package (26). For all SNPs but two (SNP 6 and 7
324 in Figure 2) there was an increase of the alternate allele from 7,500-5,500 years ago (Figure 2).
325 For all SNPs considered, the two alleles segregated in specimens pre-dating the rise and spread
326 of the DOM2 genetic lineage of modern domestic horses (data not shown). The dataset
327 considered included 431 ancient genomes previously characterized (2,27–31).
328

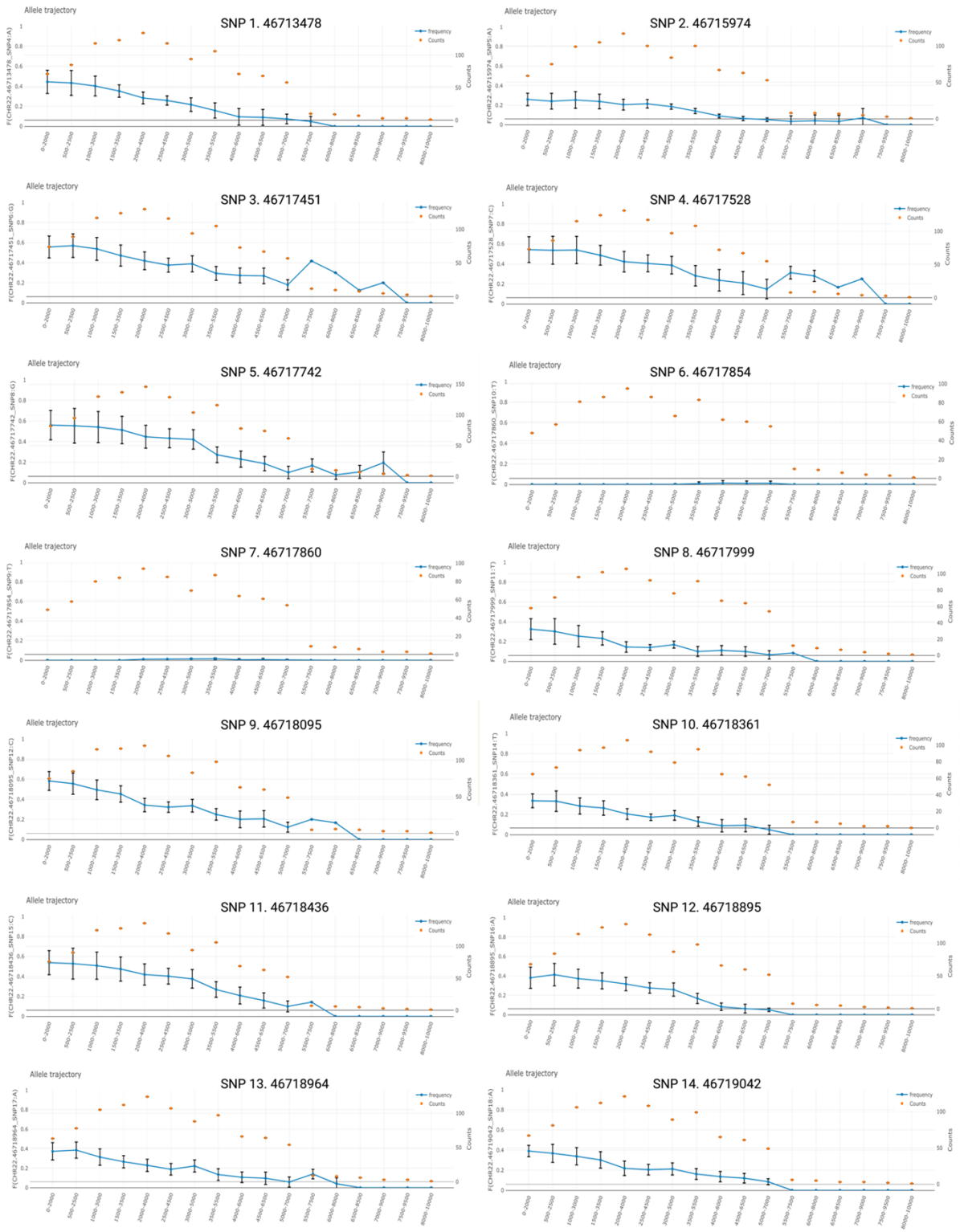


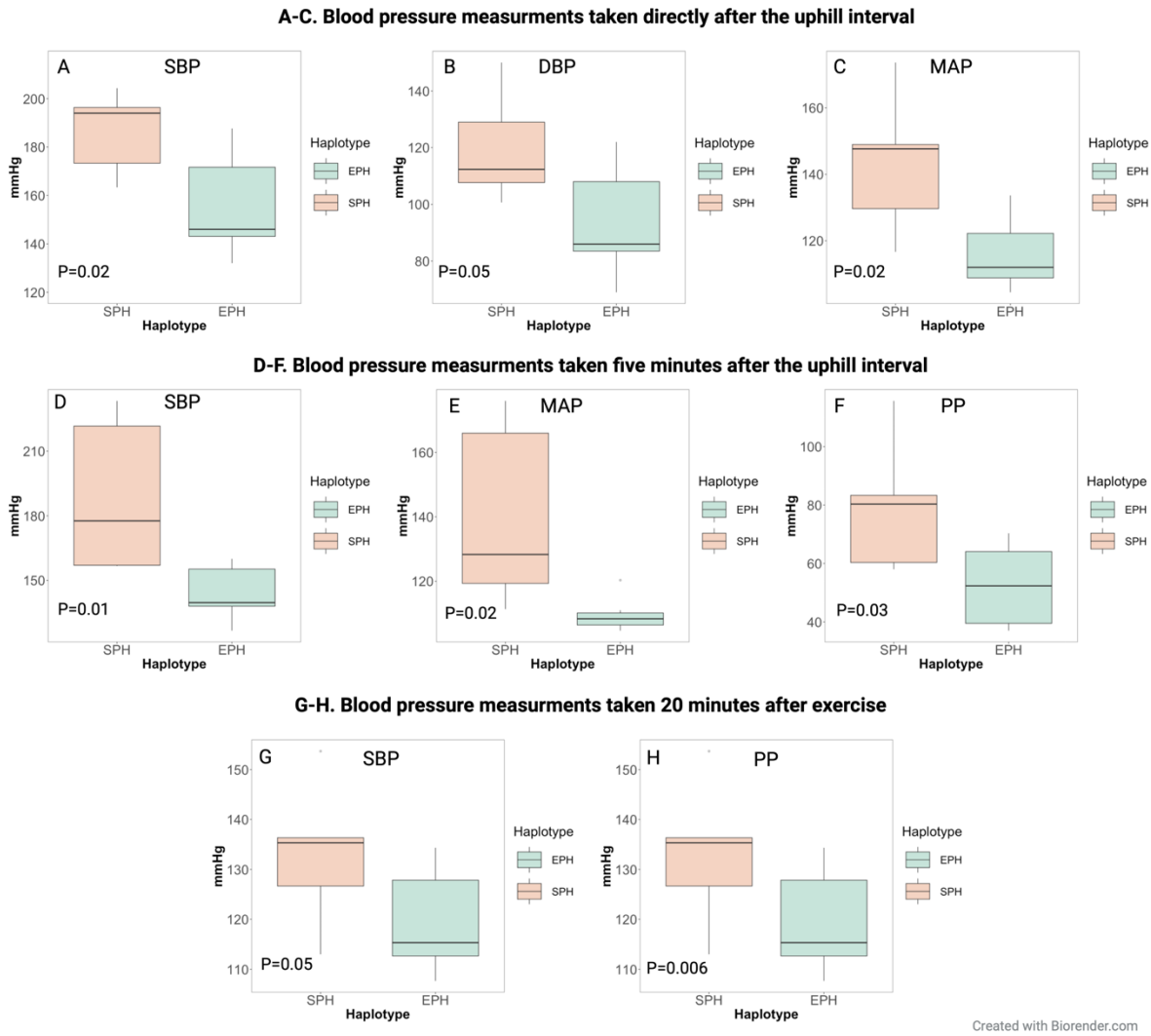
Figure 2. Frequency of the alternative allele over time from the mapDATAge package (26) for all 14 SNPs within the minimum shared haplotype in the trotters. The orange dots represent the number of horse samples considered in a given time bin of 2,000 years, and the blue line represents the frequency for the alternative allele estimated from allele counts. Figure created with Biorender.com.

334 **Elite performance haplotype linked to lower blood pressure**

335 To link genotype to phenotype, we assessed the physiological difference between CBTs
336 homozygous for SPH (n=5) and EPH (n=13) before, during and post-exercise (see methods).
337 Blood pressure measurements were also taken from eight heterozygous horses (HET) at rest.
338 Horses homozygous for SPH had, on average, significantly higher systolic blood pressure
339 (SBP), diastolic blood pressure (DBP) and mean arterial pressure (MAP) during exercise,
340 measured directly after the last uphill interval (Figure 3A-C). In addition, five mins after the
341 uphill interval, the SPH group showed significantly higher SBP, MAP, and pulse pressure (PP)
342 relative to the EPH group (Figure 3D-F). Horses homozygous for SPH also had significantly
343 higher SBP and PP 20 mins after completing the exercise, compared to EPH horses (Figure
344 3G-H). Twenty mins after the exercise EPH horses were back to resting values, while it took
345 ten additional mins for SPH horses to recover to resting values. There were no statistically
346 significant differences in blood pressure measurements at rest, before the exercise.
347 Furthermore, SPH horses had a higher heart rate than EPH horses during exercise (97.5 vs.
348 86.4) and five mins after exercise (85.7 vs. 77.0), although the difference was not significant.
349 All blood pressure values are presented in Supplementary Table S5.

350

351



352

353 **Figure 3**. Blood pressure measurements during and after exercise in horses homozygous for the sub-
354 elite performing haplotype (SPH) and horses homozygous for the elite-performing haplotype (EPH).
355 A-C: Blood pressure during exercise measured directly after the uphill interval (SPH n = 5, EPH n=7).
356 D-F: Blood pressure during exercise measured five mins after the uphill interval (SPH n = 5, EPH n=7).
357 G-H. Blood pressure measured 20 mins after completing the exercise (SPH n = 5, EPH n=11). SBP:
358 Systolic blood pressure, DBP: Diastolic blood pressure, MAP: Mean arterial blood pressure, PP: Pulse
359 pressure. PP is defined as the systolic blood pressure minus the diastolic blood pressure.

360 **Plasma concentrations of EDN1 and EDN3 were significantly different between the**

361 **haplotype groups**

362 ELISA tests were used to measure the plasma concentration of EDN1 and EDN3 at rest and

363 during exercise for each haplotype group (Figure 4). Horses homozygous for SPH (n=8) had a

364 significantly higher plasma concentration of EDN1 both at rest and during exercise compared

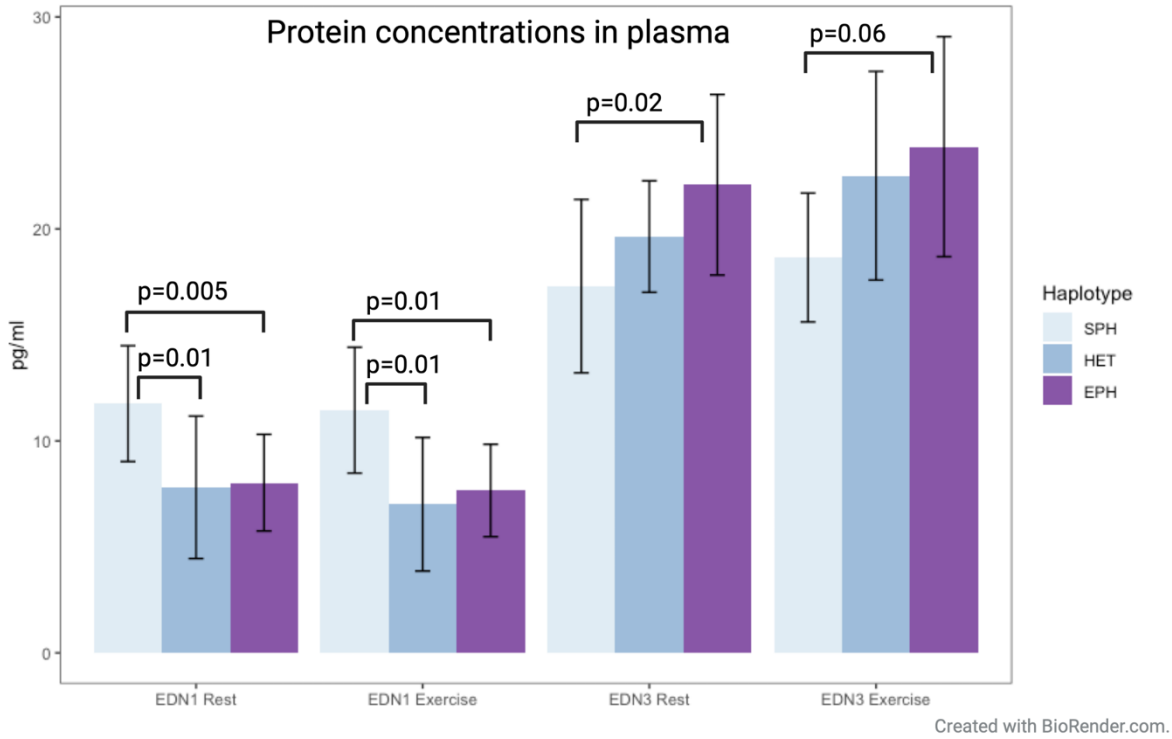
365 to the other groups (Tukey´s HSD test). In addition, SPH horses had a significantly lower

366 plasma concentration of EDN3 at rest and during exercise (borderline significant) compared to

367 horses homozygous for the EPH (Figure 4). Individual plasma values are presented in

368 Supplementary Tables S6 and S7.

369



370

371 **Figure 4**. Plasma concentrations (pg/ml) of EDN1 and EDN3 at rest and during exercise in horses

372 homozygous for the sub-elite performing haplotype (SPH) (n=8), horses heterozygous for the EPH

373 and SPH (HET) (n=11) and horses homozygous for the elite-performing haplotype (EPH) (n=21).

374 ANOVA and Tukey´s HSD test were performed. Standard deviations are presented as error bars.

375 **Significant differences in protein expression between haplotype groups**

376 To identify potential biological pathways including EDN1 and EDN3, and to compare the
377 expression of different proteins between EPH (n=6) and SPH (n=6) homozygous carriers, we
378 performed a global relative quantitative proteomic analysis using Tandem Mass Tag (TMT)
379 and mass spectrometry (MS), using plasma samples taken at rest or during exercise. In total,
380 582 distinct proteins were identified, but neither EDN1 nor EDN3 (Supplementary Table S8).
381 Proteins quantified from at least three horses from each haplotype group were included in the
382 statistical calculations. This resulted in 383 proteins that were analyzed in the five following
383 comparisons; 1) exercise vs. rest in SPH, 2) exercise vs. rest in EPH, 3) EPH vs. SPH at rest,
384 4) EPH vs. SPH during exercise, 5) paired ratio analysis, where the individual difference at rest
385 and exercise was compared between EPH and SPH groups.

386

387 Fifty proteins were found to show significantly different expression levels across the five
388 comparisons ($P < 0.05$; Supplementary Table S8). Table 7 presents the top proteins with the
389 highest fold change in each comparison, including all proteins with different levels that were
390 significant in more than one comparison.

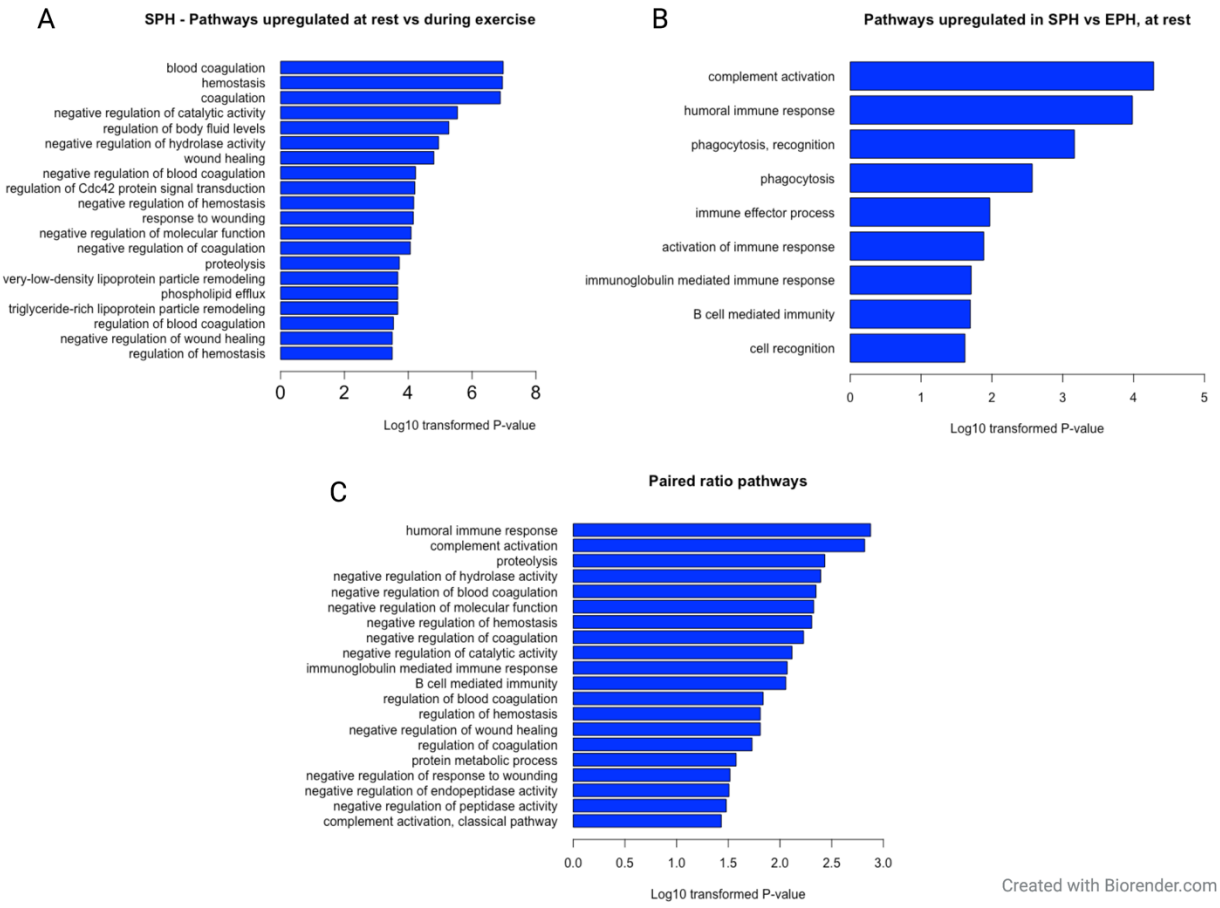
391

392 **Place for Table 7, see end of document**

393

394 When comparing protein levels during exercise vs. at rest (comparisons 1 and 2), there were
395 21 proteins in the SPH group and three proteins in the EPH group with significantly different
396 levels (Supplementary Table S8). A gene ontology (GO) analysis for the 21 proteins showing
397 changing expression profiles at rest vs. during exercise in the SPH group revealed pathways
398 related to blood coagulation, hemostasis, coagulation and regulation of body fluid levels
399 (Figure 5A) (32). When comparing EPH and SPH at rest (comparison 3), there were 23 proteins

400 upregulated in the SPH group (Supplementary Table S8), implicated in pathways related to
401 immune response and complement activation (Figure 5B). There were no statistically
402 significant differences in protein levels when comparing EPH and SPH during exercise
403 (comparison 4). The paired ratio comparison (comparison 5) revealed 19 proteins showing
404 significant difference in the individual difference at rest and exercise in SPH and EPH groups
405 (Supplementary Table S8). These proteins are involved in pathways related to immune
406 response, complement activation and catalytic and hydrolase activity (Figure 5C).



407

408 **Figure 5**. A. The 20 most significant pathways from the GO analysis based on the significant proteins
409 in the comparison of exercise vs. rest in the SPH (sub-elite performing haplotype) group. B. All
410 significant pathways from the GO analysis based on the significant proteins in the comparison of SPH
411 (sub-elite performing haplotype) and EPH (elite performing haplotype) at rest. C. The 20 most
412 significant pathways from the GO analysis based on the significant proteins in the paired ratio analysis
413 of SPH vs. EPH.

414 **Discussion**

415 The EDN3 protein has an established functional role in blood pressure regulation in humans
416 and other species. However, to the best of our knowledge, there are no published reports
417 establishing associations between the *EDN3* gene and equine athletic performance, except for
418 our previous selective sweep study (13). Given the role of EDN3 in blood pressure regulation,
419 we hypothesized that 1) an underlying phenotype for elite athletic performance is blood
420 pressure regulation, and 2) the identified non-coding region harbors a regulatory element that
421 acts on the *EDN3* gene.

422

423 All the SNPs defining the haplotypes were found to segregate in the genome of horses that
424 lived prior to the rise and spread of the modern domestic lineage around ~4,200 years ago
425 (DOM2) (31). In fact, temporal allelic trajectories portray an increase in the frequency of the
426 alternative allele, from 7,500-5,500 years ago, prior to the earliest archaeological evidence of
427 horse husbandry (30,33). The concerted rise in frequency of all alternate alleles following the
428 spread of the DOM2 lineage is compatible with a haplotype undergoing positive selection,
429 possibly following artificial selection for animals showing improved athletic performance.
430 Interestingly, the allele trajectories at *EDN3* are in striking contrast with those previously
431 described at the *MSTN* locus, driving performance in short-distance racing (6), which indicated
432 selection starting within approximately the last 1,000 years only. This may indicate
433 performance traits representing recurrent selection targets in the history of horse domestication.
434 We next demonstrated significant differences in blood pressure values, as well as protein levels
435 of EDN1 and EDN3, when contrasting homozygous EPH to SPH CBT carriers. The 5.5 kb
436 region contained multiple marks pointing at a potential *cis*-regulatory module. First, the region
437 spanned an equine enhancer (H3K4me1) active in tissues relevant to racing performance.
438 Second, comparative evidence from human gTEX v8 implicated this region in the regulation

439 of *EDN3* expression across multiple tissues. Third, of the 14 SNPs forming the haplotypes,
440 alleles from 12 showed differential transcription factor binding affinities, suggesting
441 differential regulatory activity between the EPH and SPH carriers. Intriguingly, the SPH
442 haplotype was predicted to have binding affinities for both immune system actors, NFATC2
443 and IRF2, where IRF2 has been shown to be a negative regulator of NFAT target genes (34)
444 (Supplementary Table S4). On the other hand, the EPH haplotype harbored motifs for factors
445 involved during embryogenesis and cell differentiation, e.g., SRF, HOXA5 and CREB1 (35–
446 37) (Supplementary Table S4). These differing regulatory roles may point to the multifaceted
447 role of EDN3, from early neural crest development and proper function of enteric neurons and
448 melanoblasts (38). The EDN3 ligand has a broad expression range across tissues and has been
449 implicated in diseases such as Hirschsprung's disease, Multiple Sclerosis and Waardenburg
450 syndrome in humans, as well as melanocyte development in mice (39–43). However, a key role
451 for this ligand is in blood flow homeostasis.
452
453 Endothelins are involved in one of the most potent vasoregulatory systems, where EDN1 and
454 EDN3 have opposite effects and act synergistically to regulate blood pressure (18). EDN1
455 increases blood pressure by vasoconstriction, while EDN3 is involved in nitric oxide release,
456 which results in vasodilation and decreased blood pressure (18). Additionally, EDN3
457 stimulates the secretion of vasopressin, which increases the blood volume by retaining water
458 in the kidneys, while EDN1 has the opposite effect in the kidneys (44–47). In the present study,
459 horses homozygous for EPH showed significantly higher plasma concentrations of EDN3 and
460 lower blood pressure during exercise. In contrast, horses homozygous for SPH had a
461 significantly higher plasma concentration of EDN1 and higher blood pressure during exercise.
462 Blood pressure depends on cardiac output and resistance in peripheral blood vessels, while
463 total peripheral resistance is affected by the diameter of the arteries and the viscosity of the

464 blood. An increase of EDN3 likely leads to a decrease in blood pressure and an increase in
465 vasodilation of the blood vessels, primarily in the working muscles and the lungs. This means
466 that blood can flow more easily, allowing improved arterial and venous oxygenation. This
467 would, in turn, be beneficial for athletic performance. In humans, infusion of EDN3 caused a
468 decrease in mean arterial blood pressure and vasoconstriction to the visceral organs (48). On
469 the other hand, infusion of EDN1 leads to an increased mean arterial pressure and decreased
470 cardiac output (49,50). This is comparable to a study on horses during anesthesia, which
471 revealed that horses receiving nitric oxide, a vasodilator, had better arterial and venous
472 oxygenation and lower EDN1 plasma concentration than the control group without nitric oxide
473 administration (51). In our study, we did not see any effect of exercise on the endothelin levels,
474 as there were no statistically significant differences in EDN1 or EDN3 levels when contrasting
475 resting and exercising. This is in concordance with previous studies in which no differences in
476 EDN1 levels before and during exercise were observed in healthy horses (22,52). In terms of
477 exercise performance, the present study showed that SPH homozygous horses had a slower
478 recovery than the horses carrying the EPH. This could be because the SPH group had higher
479 blood pressure during the exercise, so it took longer for these individuals to return to resting
480 blood pressure values.
481
482 Although the current study demonstrated statistically significant differences in EDN1 and
483 EDN3 levels between the different haplotype groups, there were no differences in blood
484 pressure at rest. In adult horses, normal blood pressure at rest is approximately 130/95 mmHg
485 (systolic/diastolic), in line with the values observed in the current study (53). In humans, it has
486 been demonstrated that elevated blood pressure at rest is negatively correlated with athletic
487 performance (21). Individuals with elevated blood pressure had significantly lower maximal
488 oxygen consumption, ventilatory anaerobic thresholds and heart rate reserves (difference
489 between an individual's resting heart rate and maximum heart rate) (21). However, unlike

490 humans, high blood pressure in horses is uncommon and most blood pressure measurements
491 on horses are performed to evaluate and monitor hypotension. An increase in blood pressure at
492 rest is most commonly seen as a result of diseases such as laminitis, chronic renal failure or
493 equine metabolic syndrome (54–56).
494
495 The spleen represents one important organ for the distribution of blood, acting as a reservoir
496 for red blood cells. The equine spleen reservoir is larger than in any other domestic animal
497 (57). When there is an increased demand, red blood cells stored in the spleen can be released
498 into the system (57). During exercise, the spleen is contracted, which causes an increase in the
499 volume of circulating blood cells, while the volume of plasma is unchanged or even reduced.
500 This leads to an increase in the packed cell volume/hematocrit (i.e., the proportion of blood
501 made up of red blood cells), hemoglobin concentration, and red blood cell count (57). Previous
502 studies have demonstrated significantly larger spleen size in horses used for racing compared
503 to breeds that are not traditionally used for racing (58). In addition, one study demonstrated a
504 significant correlation between blood pressure and spleen volume (59), with the splenic volume
505 decreasing when hypertension was induced. However, when hypotension was induced, there
506 were no significant changes in the splenic volume (59). The results by (59) demonstrated the
507 importance of the spleen for the cardiovascular system in horses. Further studies examining the
508 relationship between blood pressure, spleen volume and expression of EDN3 represent
509 promising avenues for enhancing understanding of the cardiovascular system in horses in
510 relation to exercising.
511
512 Our global proteomic analysis resulted in the identification of a total of 582 proteins. Only one
513 such protein, Cathepsin X, was encoded as part of the 20 protein-coding genes located within
514 2 Mb surrounding the genomic region investigated (ECA22: 45,718,095 – 47,718,095). In

515 contrast to the ELISA analysis, none of the endothelins were detected. A major obstacle in MS-
516 based plasma proteomics is the large dynamic range in protein concentrations and that the
517 abundant high molecular weight proteins mask the identification and quantification of lower
518 molecular weight proteins, such as the endothelins. The highly abundant proteins contribute to
519 approximately 90 % of the total protein content in plasma. In this study, plasma depletion of
520 the most abundant proteins was performed to facilitate the quantification of less abundant
521 proteins and to improve the sequence coverage (60). However, a limitation of the present study
522 is the quality of the available proteome database for *Equus caballus*, which only includes a
523 limited number of proteins. Additionally, many proteins were excluded due to a low number
524 of biological replicates, and in total, 383 proteins could be included in the comparisons between
525 groups.
526
527 When comparing protein levels at rest and during exercise for the different haplotype groups,
528 we observed a clear difference in the number of regulated proteins (Supplementary Table S8).
529 In the SPH group, 21 proteins were detected at higher levels at rest, than during exercise. The
530 GO analysis of these proteins identified several pathways related to injuries, including blood
531 coagulation, hemostasis, and wound healing (Figure 5A). Several of these proteins also have
532 functions related to the immune system. When comparing the protein levels between the
533 haplotype groups at rest, 23 proteins were found regulated. The GO analysis revealed
534 significant pathways mainly related to the immune system. This is consistent with previous
535 studies in which an increase of plasma proteins involved in pathways related to inflammation,
536 coagulation and immune modulation was observed (61,62). A previous genome-wide
537 association analysis of racing performance in CBTs also found correlations for genes related
538 to immune response (63). Racing horses perform intense exercise, which induces a physical
539 stress response to obtain energy for the working muscles. The released stress hormones (e.g.,

540 cortisol) stimulate gluconeogenesis and lipolysis (64). Additionally, the stress hormones affect
541 inflammatory and coagulation processes (61,62,65,66). In a non-inflammatory state, intact
542 endothelial cells maintain blood flow and control coagulation. However, chronic inflammation,
543 oxidative stress and mechanical stress from turbulent blood flow negatively affect the function
544 of the endothelial cells (67,68). Furthermore, in a human study, multi-omics changes in
545 response to exercise were observed, involving pathways related to energy metabolism,
546 oxidative stress and immune response (69). Potentially, SPH horses have more difficulties
547 recovering from exercise due to elevated blood pressure, resulting in higher levels of immunity-
548 related proteins and proteins related to injury healing in the plasma at rest.
549
550 It is interesting to note that the protein showing the highest expression difference between SPH
551 and EPH groups at rest was the proteasome subunit beta (PSMB10). Interestingly, the *psmb10*
552 gene has previously been linked to blood pressure regulation in mice. Knockout of the *psmb10*
553 gene was protective for high blood pressure, cardiac fibrosis and inflammation in DOCA salt-
554 treated mice (70). Although the *PSMB10* gene may be involved in blood pressure regulation in
555 the horse, it is notable that there were no differences in blood pressure between the haplotype
556 groups at rest. In the EPH group comparison of protein ratios at rest versus exercise, there were
557 only three proteins with significant differences. Notably, one of the proteins detected was the
558 secretoglobin family 1D member (LOC102150542). Secretoglobins are known to have anti-
559 inflammatory effects in airway disorders such as asthma in both horses and humans (71,72).
560 The increased levels of this protein may play an anti-inflammatory role during exercise by
561 preserving the function of the vascular endothelial cells, contributing to a higher tolerance for
562 intense exercise.
563

564 Several studies have demonstrated the effect of the *EDN1* gene and the endothelin receptor
565 gene, *EDNRB*, in horses and humans (73–81). Studies have suggested EDN1 as a marker for
566 different diseases in horses, for example, hyperinsulinemia, neonatal diseases, and cardiac
567 disease (76–78). Moreover, the EDNRB receptor is known to be associated with lethal white
568 foal syndrome (73–75). However, to the best of our knowledge, there is only one study
569 performed on horses involving the *EDN3* gene (13). The current study demonstrates significant
570 differences in blood pressure and plasma levels of both EDN3 and EDN1 in CBTs with
571 different haplotypes. Although the identified gene variant may also act on other genes in the
572 region or other parts of the genome, we believe that we have identified a regulatory variant
573 influencing the transcription of *EDN3*. Based on the ELISA results and racing performance
574 data, the identified candidate SNP appears to have an additive effect, with homozygous carriers
575 and heterozygous horses performing significantly better than homozygous wild-type horses.
576 The obtained results in horses emphasize the importance of the endothelin system as critical
577 for performance and support the functional role previously shown for EDN3 in human blood
578 pressure regulation.
579

580 **Conclusion**

581 Here we demonstrate, for the first time, a significant association between *EDN3*, blood pressure
582 regulation and athletic performance in horses. We identified a 5.5 kb candidate genomic
583 interval and showed that horses homozygous for the elite performing haplotype had a higher
584 plasma concentration of EDN3, lower plasma concentration of EDN1, and lower blood
585 pressure during exercise compared to horses with the sub-elite performing haplotype. Horses
586 carrying the elite performing haplotype also recovered from exercise faster than those with the
587 sub-elite haplotype. The results reported here are important for understanding the biological
588 mechanisms behind blood pressure regulation, as well as inflammation and coagulation

589 systems in relation to racing performance in both horses and other mammals, including
590 humans.
591

592 Materials and Methods

593 Horse material and DNA extraction

594 A summary of the horses included in the study, and specifying which experiments they were
595 used for, is presented in Supplementary Table S2. Genomic DNA from all horses was extracted
596 from hair roots or blood samples. For hair preparation, 186 µL of 5 % Chelex® 100 Resin (Bio-
597 Rad Laboratories, Hercules, CA, USA) and 14 µL of proteinase K (20 mg/mL; Merck KgaA,
598 Darmstadt, Germany) were added to each sample. The mixture was incubated at 56°C for two
599 hours at 600 rpm and inactivated with proteinase K for 10 mins at 95°C. Blood samples were
600 extracted on the Qiasymphony instrument with the Qiasymphony DSP DNA mini or midi kit
601 (Qiagen, Hilden, Germany).
602

603 Selective sweep racing performance association analysis with additional horse
604 material

605 We used the previously identified 19.6 kb selective sweep region, associated with CBT racing
606 performance, as the starting point for this analysis (ECA22:46,702,297-46,721,892;
607 EquCab3.0) (13). In that study, g.46717860-T was the variant most significantly associated
608 with elite performance, and LD to this variant was used to define a five SNP haplotype (r^2 >
609 0.9). We extracted the 19.6 kb region from an existing 670K Axiom Equine Array genotyping
610 experiment of 661 CBTs (including the 400 horses used for the haplotype analysis in the 2018
611 study) (13,63). Using g.46717860 as the lead variant, we calculated pairwise r^2 across the
612 region, and performed racing performance association analyses between this variant, and any
613 additional variant with $r^2 > 0.6$. Given the high LD, a significance threshold was set at $P \leq 0.05$.

614

615 The available phenotypic records included a) pedigree information b) *DMRT3* genotypes and
616 c) performance data as individual race records for each horse (Swedish Trotting Association
617 (Svensk Travsport) and the Norwegian Trotting Association (Norsk Rikstoto). The *DMRT3*
618 variant was genotyped using StepOnePlus Real-Time PCR System (Thermo Fisher Scientific,
619 Waltham, MA, USA) with the custom TaqMan SNP genotyping assay (82), and included as a
620 factor in the model, since previous studies have demonstrated major effects of the variant on
621 harness racing performance (10,12). Only competitive races were included (i.e., premie and
622 qualification races were excluded). The performance data included: number of starts, number
623 of wins, number of placings (the number of times a horse finished a race in first, second or
624 third place), fastest kilometer (km) time in a race where the horse did not gallop (in seconds),
625 earnings and earnings per start. The earnings for most horses were recorded in Swedish
626 currency (SEK). Earned prize money in Norwegian currency was converted to SEK based on
627 the average exchange rate for the year the race took place.

628

629 All statistical analyses were performed in the R statistical environment. The performance data
630 were tested for normality by computing the skewness coefficient using the package moments
631 v0.14. Non-normally distributed values were transformed, i.e., log transformed values (log10
632 +1) were used for wins and placings, number of starts were square root transformed and
633 earnings and earnings per start were transformed according to the previously reported formula
634 (ln(earnings + 1000)) (83). Genetic variants were tested in a linear model using ANOVA as a
635 *post hoc* test. Number of starts, age at first start, sex, birth year, country of birth, and the
636 *DMRT3* genotype were included as fixed effects, when significant. Significant values ($P \leq$
637 0.05) were further tested using the function lsmeans (Least-Squares Means) with the package
638 emmeans followed by the multiple comparison test Tukey's HSD-test.

639

640 **Additional variant discovery**

641 To find and assess variants missing from the Axiom Equine Array, but with the potential to
642 influence racing performance, we performed both Illumina short-read WGS and MinION
643 Oxford Nanopore targeted sequencing of PCR products.

644

645 First, two CBTs and two SBs were selected for WGS based on their g.46717860 genotype and
646 their performance values. From each breed, horses homozygous for opposite alleles of SNP
647 g.46717860 (CC respectively TT) with either high (TT) or low (CC) earnings per start were
648 selected for sequencing (Supplementary Table S4). Paired-end 150 bp Illumina HiSeqX data
649 was generated to a depth of 15X. Read data were processed according to the GATK 3.8-0 best
650 practices (https://software.broadinstitute.org). Briefly, raw data were trimmed with TrimGalore
651 0.4.4 and quality checked with QualiMap 2.2. The reads were aligned to the EquCab3.0
652 reference genome (84) (GCF_002863925.1) using BWA-MEM 0.7.17. SAM files were
653 converted to BAM files using SAMtools, and Picard 2.10.3 was used to sort the BAM files and
654 remove potential PCR duplications. The BAM files were recalibrated in two steps using
655 BaseRecalibrator and PrintReads. Variant calling was performed using HaplotypeCaller with
656 a list of known variants in horses downloaded from the NCBI web page annotation release 103
657 (https://www.ncbi.nlm.nih.gov). Structural variants were called using default parameters in
658 FindSV comprising the software programs Manta, TIDDIT and CNVNator
659 (https://github.com/J35P312/FindSV).

660

661 Next, as input for Oxford Nanopore sequencing (ONT, Oxford, UK), a 25 kb PCR product was
662 designed to span the 19.6 kb sweep region and adjacent repeats. Primers were designed in
663 Primer3Plus and included unique 10 bp barcodes (Supplementary Table S9). DNA input was
664 taken from four CBTs with g.46717860-TT genotype and high performance and four CBTs

665 with the opposite features, including the two CBTs from the WGS analysis (horse selection as
666 per WGS above and Supplementary Table S2). The PCR region was amplified on a ProFlex
667 PCR cycler (Thermo Fisher Scientific, Waltham, MA, USA) and the long-range PCR using
668 AccuPrime™ Taq DNA Polymerase, High Fidelity kit following manufacturers specifications.
669 The sizes of PCR products were confirmed on 0.8 % TBE Agarose gels, and product
670 concentration quantified with Agilent 4150 TapeStation using the Agilent Genomic DNA
671 ScreenTape Assay (Agilent Technologies, Santa Clara, CA, USA). Samples of the respective
672 PCR products were pooled in equimolar concentration and libraries were prepared using the
673 Genomic DNA by Ligation (SQK-LSK109) protocol. The samples were loaded and sequenced
674 on the MinION. Demultiplexed sequencing data were mapped to EquCab3.0 ECA 22
675 (GCF_002863925) using minimap2 v. 2.4 sequence alignment program
676 (https://github.com/lh3/minimap2). The resulting SAM files were converted to BAM format
677 using samtools v. 1.8. Single nucleotide variants were called using the variant calling tool
678 longshot v. 0.4.3 with default parameters (https://github.com/pjedge/longshot). Structural
679 genomic variations were identified with NanoSV v. 1.2.4 software
680 (https://github.com/mroosmalen/nanosv), with default parameters. Only variants with a quality
681 above 20 were retained.
682
683 **Genotyping**
684 The largest INDEL identified from the WGS data was analyzed for association with racing
685 performance traits in 369 CBTs, using linear models. For MassArray SNP genotyping the
686 number of variant positions available was limited to 25. For this reason, 24 SNPs and one
687 INDEL, evenly distributed across the 19.6 kb region, were selected for genotyping. In total,
688 DNA from 412 horses of 13 different breeds and four donkeys were genotyped (Supplementary
689 Table S10). The horses were carefully selected to include both traditional performance horse

690 breeds, as well as non-performing breeds. Genotyping was performed at the Mutation Analysis
691 Facility at Karolinska University Hospital (Huddinge, Sweden), using iPLEX® Gold chemistry
692 and the MassARRAY® mass spectrometry system (85) (Agena Bioscience, San Diego, CA,
693 USA). Multiplexed assays were designed using MassARRAY® Assay Design Suite v2.2
694 software (Agena Bioscience, San Diego, CA, USA), genotyping the 25 markers in one reaction
695 per sample. The protocol for allele-specific base extension was performed according to Agena
696 Bioscience's recommendation. Analytes were spotted onto a 384-element SpectroCHIP II
697 array (Agena Bioscience, San Diego, CA, USA), using Nanodispenser RS1000 and
698 subsequently analyzed by MALDI-TOF on a MassARRAY® Analyzer 4 mass spectrometer
699 (Agena Bioscience, San Diego, CA, USA). Genotype calls were manually checked by two
700 persons individually using MassARRAY® TYPER v4.0 Software (Agena Bioscience, San
701 Diego, CA, USA).
702
703 The variant g.46717860 was not included as part of the MassArray, but was genotyped for 170
704 horses (9 Ardennes, 96 CBTs, 4 donkeys, 4 Finnhorses, 2 Fjord horses, 2 Gotlandsruss, 10
705 North-Swedish draught horses, 3 Pzrewalski and 40 SBs) using the StepOnePlus Real-Time
706 PCR System (Thermo Fisher Scientific, Waltham, MA, USA) and the custom TaqMan SNP
707 genotyping assay (82). The reaction volume of 15 µl comprised 1.5 µl DNA, 0.38 µl
708 Genotyping Assay 40X, 7.5 µl Genotyping Master Mix 2X, and 5.62 µl deionized water.
709
710 **Haplotype analysis**
711 Pairwise r^2 calculations with all 24 SNPs and g.46717860 were performed using the LD-
712 function in R. Based on the high LD between all genotyped SNPs, it was possible to impute
713 g.46717860 alleles for ungenotyped samples. All SNPs with a pairwise r^2 value >0.6, plus two
714 flanking SNPs were selected for haplotype analysis. Haplotype racing performance association

715 tests used a GLM regression analysis, which was performed in 180 CBTs and 38 SBs using the
716 haplo.glm function from the haplo.stats package in R (86). The model included the effects of
717 sex, age, country of registration, *DMRT3* genotype (only for the CBTs) and number of starts,
718 when significant. All CBTs had been genotyped for the DMRT3 mutation, using the
719 StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) with
720 the custom TaqMan SNP genotyping assay (82). Haplotypes with a frequency < 2 % were
721 considered rare, and were not analyzed for association with performance. A minimum shared
722 haplotype of 14 SNPs between CBTs and SBs was used to define the elite-performing
723 haplotype (EPH) and the sub-elite performing haplotype (SPH).

724 **Association analysis with athletic performance**

725 Allele calls at SNP g.46718095 were a proxy for elite-performing haplotype (EPH) and the
726 sub-elite performing haplotype (SPH). Samples not genotyped on the MassArray were
727 genotyped using the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific,
728 Waltham, MA, USA) with the custom TaqMan SNP genotyping assay (82). Association with
729 racing performance was performed as per "Selective sweep racing performance association
730 analysis with additional horse material" using 539 CBTs born between 1994 and 2017 and 271
731 SBs born between 2005 and 2014. Horses of 13 different breeds and three other types of equids
732 (wild ass, onager and donkey) were also genotyped to investigate the distribution of allele
733 frequencies.

734 **Functional bioinformatic analyses**

735 ChIP-seq data were provided by the Equine Functional Annotation of Animal Genomes
736 (FAANG) team (www.faang.org) (24,25) including four histone marks (H3K4me1, H3K4me3,
737 H3K27ac and H3K27me) across nine tissues (adipose, brain, heart, lamina, liver, lung, muscle,
738 ovary and skin) from two Thoroughbreds. FAANG bam and bed files from the different

739 available tissues, the four histone marks and the input DNA control (project: PRJEB35307)
740 were downloaded to a local server. First, the integrity of the data was checked using checksums.
741 Bam files were sorted with samtools and bedgraphs were subsequently created using bedtools
742 (87,88). The Integrative Genomics Viewer (IGV) browser was used to visualize the output data
743 along with the EquCab3.0 fasta file and its gff3 annotation file (84,89). The presence of
744 H3K27ac and H3K4me1 was considered to mark active enhancers, whereas H3K4me1 and
745 H3K27me3 marked poised enhancers. H3K27ac and H3K4me3 marked active promoters and
746 H3K27me3 was considered to mark repressed regions. The region defined by the minimum
747 shared haplotype (see above) was investigated using IGV for the different histone marks in all
748 available tissues.
749
750 A comparative analysis to hg38 was conducted for the minimum shared haplotype region. Co-
751 ordinates were lifted in UCSC (liftOver) and potential overlaps between ENCODE Candidate
752 Cis-Regulatory Elements (cCREs), ENCODE Integrated Regulation tracks and ORegAnno
753 Regulatory elements were annotated (90,91). For each of the SNPs in the minimum shared
754 haplotype region, the potential of alleles to alter transcription factor binding potential was
755 tested with sTRAP (92). Here, 31 bp sequences centered on the allele of interest were tested
756 using JASPAR vertebrates, on a background model of human promoters. In the cases where
757 two SNPs were within 10 bp of each other, they were tested as a single phased fragment.
758
759 **Ancient DNA screening**
760 We leveraged the availability of an extensive ancient genome time-series in the horse to assess
761 the temporal trajectory and spatial distribution of the SNPs in the minimum shared haplotype
762 within the CBTs, in the past. More specifically, we re-aligned against the EquCab3.0 reference
763 genome (84) supplemented with the Y-chromosomal contigs from (93) and a sub-selection of

764 the sequencing data from (2,27,29,30,94,95), representing a total of 431 ancient genomes.
765 Alignment files were generated using the Paleomix pipeline (96), and further rescaled and
766 trimmed, following the procedures described in (31). This procedure ensured minimal
767 sequencing error rates, especially at sites potentially affected by post-mortem DNA damage.
768 The number of occurrences of each individual allele was counted for each individual using
769 ANGSD (v0.933), with the -baq 0, -remove_bads -minMapQ 25 -minQ 30 -rmTriallelic 1e-4
770 -SNP_pval 1 -C 50 parameters (97). The resulting occurrences were tabulated together with
771 metadata available for each individual, consisting of their corresponding archaeological site
772 (name and GPS coordinates) and their average calibrated radiocarbon date (or age as assessed
773 from the archaeological context otherwise). The tabulated file was visualized using the
774 mapDATAge package (26), which provides spatial distributions within user-defined time
775 periods and/or spatial ranges and estimates allelic trajectories (i.e., allelic frequencies through
776 time) based on the resampling of individual allelic counts. The final allelic trajectories were
777 assessed considering time bins of 2,000 years (step-size = 500). To explore the genetic variation
778 for the SNPs within the haplotype, we considered the 14 polymorphic sites i.e.,
779 ECA22:46703380, 46708110, 46708983, 46713478, 46715974, 46717451, 46717528,
780 46717742, 46717854, 46717999, 46718095, 46718113, 46718361, 46718436, 46718895,
781 46718964, 46719042, 46719906, 46720030, 46720999, and 46721822.

782 **Exercise test, sample collection and blood pressure measurements**

783 **Horse material**

784 Thirty horses were enrolled in the exercise test (Supplementary Table S2). All horses were
785 genotyped for SNP g.46717860 and g.46718095 using the StepOnePlus Real-Time PCR
786 System (Thermo Fisher Scientific, Waltham, MA, USA) with the custom TaqMan SNP
787 genotyping assay as above (82). The two SNPs were used as a proxy for the haplotype variants

homozygote EPH (TT respectively CC) or homozygote SPH (CC respectively TT). Horses were then divided into three groups based on haplotype variant, i.e., homozygote EPH or SPH and HET (heterozygote). The horses were carefully selected to include an even sex and age distribution across the different groups. The age of the horses ranged from two to 13 years (mean age = 5.5 years). All horses, except for the two-year old, were in competing condition. However, due to various reasons including practical problems, blood pressure measurements were not taken from all horses at all time points and some horses only had measurements at rest (Supplementary Table S2). For the ELISA and proteomics analysis, plasma samples were collected from 40 horses, before and/or during exercise. As for blood pressure measurement, some horses only had samples collected at rest.

Exercise

All horses, except for the two two-year old horses, were challenged on the same standardized exercise session (n=28). The test was performed on several days during autumn and winter of the years 2020 and 2021. Each horse, with a trainer, completed a 1.5-hour exercise route that included both flat and uphill interval training. This route was commonly used in their daily training program. A Polar M460 sensor with GPS (Polar Electro, Kempele, Finland) was used to record heart rate (HR), speed, distance and amplitude during exercise. The horses performed the exercise on a track with hilly terrain in the forest. The uphill interval section consisted of the horses trotting four times up a hill with four degrees inclination. Each uphill interval lasted for five mins with a distance of 1.7 km. The HR for each horse reached above 200 beats/min. After the uphill intervals, the horses walked down to the flat part of the track, where they completed 10 mins of intervals, at a speed of 35 km/h and HR of about 180-200 beats/min. Finally, the horses jogged one km back to the stable at a speed of 15-20 km/h. During the exercise, the average HR reached above 190 beats/min for approximately eight mins. Due to their young age the two-year old horses (n=2) performed a standardized exercise session at the

813 racetrack. Following warm-up, they performed a number of intervals with HR reaching over
814 200 beats/min. The two-year old horses performed the exercise test at the same time.

815 **Blood pressure measurements**

816 The following measurements were collected with a Cardell Veterinary Monitor 9402 cuff
817 (Cardell 10), at the middle coccygeal artery: systolic, diastolic, mean arterial pressure, pulse
818 pressure, and heart rate. The same cuff size was used for all the horses. The monitor has
819 previously been successfully evaluated for use in horses (98). Blood pressure was measured at
820 rest in the stable before the exercise and during exercise (directly and five mins after the last
821 uphill interval) (Figure 6). Additionally, blood pressure was measured every ten mins after
822 exercise, up to 40 mins after the exercise was completed. In order for the horse to acclimate to
823 the equipment, the Polar M460 sensor and the cuff were put on at rest, before the exercise route,
824 and the horse was left in the box for 10 mins before measurements were taken. Similarly, blood
825 pressure was measured five times at rest for the horse to get used to the measurement procedure.
826 The measurements with the highest and lowest systolic blood pressure values were removed.
827 During and after exercise the blood pressure was measured three times and an average
828 calculated. For the association analyses with blood pressure, the mean values of the triplicates
829 were analyzed using ANOVA and Tukey's HSD test.
830

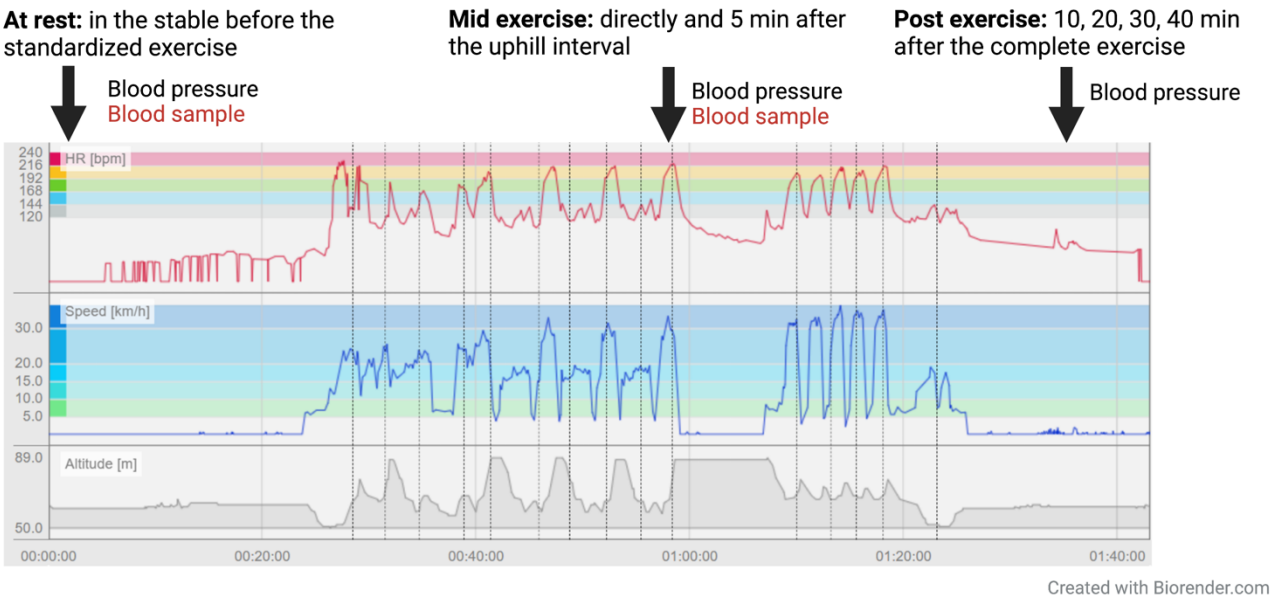


Figure 6. A schematic overview of the parameters recorded during the exercise for each horse, including heart rate, speed and altitude. The arrows indicate when blood samples and blood pressure measurements were taken.

Biological sample collection and analysis

Blood samples were collected at rest before the exercise and during exercise directly after the last uphill interval. Blood was also collected for plasma extraction. Here, a sample was collected in an EDTA tube and mixed gently four times after sampling, to allow the EDTA to mix with the blood, then centrifuged at 1000 x g (3000 rpm) for 20 mins as soon as possible and within two hours after sampling. The tubes were kept cold until centrifugation. The EDTA plasma was aliquoted in Eppendorf tubes and immediately frozen at -20°C and within five days put at -80°C until analysis.

ELISA analysis of EDN1 and EDN3

Plasma EDN1 concentrations were measured using the Horse Endothelin ELISA Kit PicoKine™ EK0945-EQ (Boster Bio, Pleasanton, CA, USA). Plasma EDN3 was quantified

847 with the horse EDN3 ELISA Kit MBS9901200 (MyBioSource, San Diego, CA, USA). The
848 analyses were performed according to the manufacturer's manual. A standard curve was
849 created, and the samples were run in duplicates. Mean values of the concentration (pg/ml) were
850 used in the statistical analysis, and the standard deviation (SD) and coefficient of variation
851 (CV) were calculated. Differences in concentrations between the three groups (EPH, HET and
852 SPH) before and during exercise were analyzed using ANOVA and Tukey´s HSD test.
853 Significance was set at P-values ≤ 0.05.

854 **Relative quantitative proteomic analysis**

855 A global relative quantitative proteomic analysis was performed to identify differentially
856 expressed proteins in plasma before and during exercise. Six horses of each haplotype group
857 (EPH or SPH homozygote) were included in the analyses (Supplementary Table S2). The
858 horses were selected based on earnings per start, age and sex. The two-year-old horses were
859 matched, one from each haplotype group, as they did not perform the same standardized
860 exercise as the other horses.

861 **Sample preparation**

862 The plasma samples and references (10 µl) were depleted from albumin and IgG, using the
863 Pierce™ Albumin/IgG Removal Kit (Thermo Fisher Scientific, Waltham, MA, USA)
864 according to the manufacturer's instructions. The reference was a representative pool that
865 included four groups: EPH at rest, EPH during exercise, SPH at rest and SPH during exercise.
866 The samples were processed using the modified filter-aided sample preparation (FASP) method
867 (99). Half the volume, corresponding to approximately 30 µg, was concentrated on Microcon-
868 30kDa Centrifugal Filter Units (Merck & Co, Kenilworth, NJ, USA), reduced with 100 mM
869 dithiothreitol at room temperature for two hours, washed several times with 8 M urea and once
870 with digestion buffer (DB) (25 mM TEAB, 0.5 % sodium deoxycholate) before alkylation with

871 10 mM methyl methanethiosulfonate in digestion buffer for 30 mins. Digestions were
872 performed by adding 1 µg Pierce MS grade Trypsin (Thermo Fisher Scientific, Waltham, MA,
873 USA) in DB and incubating overnight at 37°C. An additional portion of 1 µg trypsin was added,
874 and the samples were incubated for another three hours. Peptides were collected by
875 centrifugation and labeled using Tandem Mass Tag (TMTpro 16plex) reagents (Thermo Fisher
876 Scientific, Waltham, MA, USA), according to the manufacturer's instructions. The labeled
877 samples were combined into two sets, where both sets included all four groups and a reference.
878 The paired samples were analyzed in the same set. Sodium deoxycholate was removed by
879 acidification with 10 % TFA and the sets were desalted using Pierce peptide desalting spin
880 columns (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's
881 instructions. The two sets were pre-fractionated into 40 fractions with basic reversed-phase
882 chromatography, using a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific,
883 Waltham, MA, USA), an XBridge BEH C18 column (3.5 µm, 3.0x150 mm, Waters Corp,
884 Milford, MA, USA) and a gradient from 3 % to 90 % acetonitrile in 10 mM ammonium formate
885 with pH 10.0 for 25 mins. The fractions were concatenated into 20 fractions, dried, and
886 reconstituted in 3 % acetonitrile and 0.2 % formic acid.

887 **Liquid chromatography–Mass Spectrometry/ Mass Spectrometry analysis**

888 The fractions were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer
889 interfaced with an Easy-nLC 1200 liquid chromatography system (Thermo Fisher Scientific,
890 Waltham, MA, USA). Peptides were trapped on an Acclaim Pepmap 100 C18 trap column (100
891 µm x 2 cm, particle size 5 µm, Thermo Fisher Scientific, Waltham, MA, USA) and separated
892 on an analytical column (75 µm x 35 cm) packed in-house with Reprosil-Pur C18 material
893 (particle size 3 µm, Dr. Maisch, Ammerbuch, Germany) at a flow of 300 nL/min using a
894 gradient from 4 % to 80 % acetonitrile in 0.2 % Formic Acid over 90 mins. The Orbitrap was
895 equipped with the FAIMS Pro ion mobility system alternating between the compensation

896 voltages of -50 and -75. MS scans were performed at 120,000 resolutions, and the 12 most
897 abundant doubly or multiply charged precursors were isolated using 0.7 m/z isolation window
898 and dynamic exclusion within 10 ppm for 60 seconds. The precursors were fragmented by
899 collision induced dissociation (35 %) and detected in the ion trap, followed by multinotch
900 (simultaneous) isolation of the top 10 MS2 fragment ions within the m/z range 400-1400, MS3
901 fragmentation by higher-energy collision dissociation (55 %) and detection of the reporter ions
902 in the Orbitrap at 50 000 resolution, m/z range 100-500.

903 **Proteomic data analysis**

904 The data files in each set were merged for identification and relative quantification using
905 Proteome Discoverer version 2.4 (Thermo Fisher Scientific, Waltham, MA, USA). The Uniprot
906 *Equus caballus* proteome database (July 2021, combined SwissProt and TrEMBL 44,488
907 entries) and Mascot search engine v. 2.5.1 (Matrix Science, London, UK) were used in the
908 database matching with peptide and fragment ion tolerance of 5 ppm and 0.6 Da. Tryptic
909 peptides with zero missed cleavage, variable modifications of methionine oxidation and fixed
910 cysteine alkylation were accepted. ProTMT modifications of N-terminal and lysine were
911 selected. Percolator was used for the peptide-spectrum match validation with a strict false
912 discovery rate (FDR) threshold of 1 %. The TMT reporter ions were identified with 3 mmu
913 mass tolerance and no normalization was applied to the samples. For the protein quantification,
914 only unique peptide sequences at a strict FDR threshold of 1 % with a minimum synchronous
915 precursor selection match of 65 % and an average S/N above 10, were taken into account. The
916 reference in each set was used to calculate protein ratios and quantified proteins were filtered
917 at 5 % FDR.

918 **Processing of quantitative proteome data**

919 The protein ratios of quantified and normalized proteins were processed in Microsoft Excel to
920 calculate the group means and fold changes. All plasma samples were grouped into two sets;
921 before and after exercise. Only proteins quantified in at least three horses from each haplotype
922 group (EPH and SPH), i.e., proteins with at least three biological replicates in each haplotype
923 group, were included in the analyses. Perseus version 1.6.15.20 was used to analyze differences
924 between the groups, by applying t-test on log2 protein ratios. Five different comparisons were
925 made; SPH and EPH at rest and during exercise as well as rest versus exercise within each
926 haplotype. Additionally, paired ratio analysis was performed, to compare differences in protein
927 levels during exercise vs. at rest for each haplotype i.e., (SPH-rest/SPH-exercise)/(EPH-
928 rest/EPH-exercise). For each comparison, fold change values were calculated by dividing the
929 average values for the compared groups and P-values ≤ 0.05 were considered to be significant.
930 GO-annotations downloaded from the Uniprot database into Perseus were used to support the
931 identification of changing biological processes. Furthermore, the GO enrichment analysis and
932 visualization tool g:GOst from g:Profiler was used to identify the biological processes affected
933 in each of the different contrasts (32).

934

935 **Declarations**

936 **Ethics approval and consent to participate**

937 Blood sample collection was approved by the ethics committee for animal experiments in
938 Uppsala, Sweden (number: 5.8.18-15453/2017 and 5.8.18-01654/2020).

939 **Consent for publication**

940 Not applicable.

941 **Availability of data and materials**

942 The datasets used and analyzed during the current study will be deposited to the SRA.

943 **Competing interests**

944 The authors declare that they have no competing interests.

945 **Funding**

946 The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning
947 (FORMAS) and The Swedish Research Council (VR). This project has also received funding
948 from the CNRS, University Paul Sabatier (AnimalFarm IRP), and the European Research
949 Council (ERC) under the European Union's Horizon 2020 research and innovation program
950 (grant agreement 681605-PEGASUS).

951 **Authors´ contributions**

952 G.L. designed the project and led the genetic mapping together with J.R.S.M. M.K.R. was
953 supervised with main input from G.L. and R.N. to perform the sampling, genetic mapping and
954 blood pressure measurements. M.Å. designed genotyping assays. A.J. performed genotyping.
955 M.K.R. and K.F. performed haplotype analysis. L.O. performed the mapDATAge analysis.
956 R.N. designed the long-read sequencing experiment, performed ChIP-sequencing and
957 proteomics data analysis. A.T. performed proteomic experiments and analysis. B.D.V.
958 contributed with supervision of M.K.R. and A.R. contributed with mentoring of K.F. B.E
959 .advised the blood pressure measurements. C.M.M. assisted with human comparisons and in
960 the interpretation of the blood pressure results. G.A. performed transcription factor binding site
961 analyses. J.R.S.M. assisted with the comparative genomic analysis and performed the
962 transcription factor binding site analysis. K.F. lead the writing of the manuscript and created
963 figures, with input from all co-authors.

964 **Acknowledgements**

965 Thanks to all horse owners and trainers who participated in the study, The Swedish Trotting
966 Association and the Norwegian Trotting Association for providing pedigree and performance
967 data, Anna Darlene van der Heiden for the ONT sequencing, Tytti Vanhala for help with
968 genotyping, Anna Johansson at the National Bioinformatics Infrastructure Sweden at
969 SciLifeLab for bioinformatics advice, and the Department of Clinical Sciences Clinical training
970 center (KTC) at SLU for providing the Cardell Veterinary monitor for the blood pressure
971 measurements. Anna Svensson´s lab at the Department of Clinical Sciences SLU for running
972 ELISA. Adnan Niazi at SLU Bioinformatics Infrastructure (SLUBI), Uppsala, Sweden SLUBI
973 for supporting Nanopore data analysis. WGS was performed by the SNP&SEQ Technology
974 Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden
975 and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish
976 Research Council and the Knut and Alice Wallenberg Foundation. The computations and data
977 handling were enabled by resources in project snic2018-8-226 provided by the Swedish
978 National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish
979 Research Council through grant agreement no. 2018-05973. The global relative quantitative
980 proteomic analysis was performed at the Proteomics Core Facility, Sahlgrenska Academy at
981 the University of Gothenburg, Gothenburg, Sweden.

982

983 **References**

984 1. Andersson L. Domestic animals as models for biomedical research. Ups J Med Sci.
985 2016;121(1):1–11.

986 2. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking
987 Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell.
988 2019 May;177(6):1419–35.

989 3. McGivney BA, Han H, Corduff LR, Katz LM, Tozaki T, MacHugh DE, et al. Genomic
990 inbreeding trends, influential sire lines and selection in the global Thoroughbred horse
991 population. Sci Rep. 2020 Jan 16;10(1):466.

4. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet. 2013;9(1):1–17.

5. Meadows JRS, Lindblad-Toh K. Dissecting evolution and disease using comparative vertebrate genomics. Nat Rev Genet. 2017 Oct;18(10):624–36.

6. Hill EW, Gu J, Eivers SS, Fonseca RG, McGivney BA, Govindarajan P, et al. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. PloS One. 2010 Jan 20;5(1):1–6.

7. Tozaki T, Sato F, Hill EW, Miyake T, Endo Y, Kakoi H, et al. Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. J Vet Med Sci. 2011 Dec;73(12):1617–24.

8. McGivney BA, Browne JA, Fonseca RG, Katz LM, Machugh DE, Whiston R, et al. MSTN genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. Anim Genet. 2012 Dec;43(6):810–2.

9. François L, Jäderkvist Fegraeus K, Eriksson S, Andersson LS, Tesfayonas YG, Viluma A, et al. Conformation Traits and Gaits in the Icelandic Horse are Associated with Genetic Variants in Myostatin (MSTN). J Hered. 2016 Sep;107(5):431–7.

10. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin CJ, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012 Aug;488(7413):642–6.

11. Jäderkvist Fegraeus K, Johansson L, Mäenpää M, Mykkänen A, Andersson LS, Velie BD, et al. Different DMRT3 Genotypes Are Best Adapted for Harness Racing and Riding in Finnhorses. J Hered. 2015 Dec;106(6):734–40.

12. Jäderkvist K, Andersson LS, Johansson AM, Árnason T, Mikko S, Eriksson S, et al. The DMRT3 "Gait keeper" mutation affects performance of Nordic and Standardbred trotters. J Anim Sci. 2014 Oct;92(10):4279–86.

13. Jäderkvist Fegraeus K, Velie BD, Axelsson J, Ang R, Hamilton NA, Andersson L, et al. A potential regulatory region near the EDN3 gene may control both harness racing performance and coat color variation in horses. Physiol Rep. 2018 May;6(10):1–12.

14. Horinouchi T, Terada K, Higashi T, Miwa S. Endothelin Receptor Signaling: New Insight Into Its Regulatory Mechanisms. J Pharmacol Sci. 2013;123(2):85–101.

15. Rapoport RM, Merkus D. Endothelin-1 Regulation of Exercise-Induced Changes in Flow: Dynamic Regulation of Vascular Tone. Front Pharmacol. 2017;8:1–9.

16. Arai H, Hori S, Aramori I, Ohkubo H, Nakanishi S. Cloning and expression of a cDNA encoding an endothelin receptor. Nature. 1990 Dec 20;348(6303):730–2.

17. Sakurai T, Yanagisawa M, Takuwa Y, Miyazaki H, Kimura S, Goto K, et al. Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. Nature. 1990 Dec 20;348(6303):732–5.

1030 18. Rossi GP, Sacchetto A, Cesari M, Pessina AC. Interactions between endothelin-1 and the
1031 renin-angiotensin-aldosterone system. Cardiovasc Res. 1999 Aug 1;43(2):300–7.

1032 19. Ostrander EA, Huson HJ, Ostrander GK. Genetics of Athletic Performance. Annu Rev
1033 Genomics Hum Genet. 2009;10(1):407–29.

1034 20. Davis PG, Ferguson MA, Alderson NL, Pate RR, Bodary PF, Durstine JL. Effect of
1035 exercise duration on plasma endothelin-1 concentration. J Sports Med Phys Fitness. 2005
1036 Sep;45(3):419–23.

1037 21. Mazic S, Suzic Lazic J, Dekleva M, Antic M, Soldatovic I, Djelic M, et al. The impact of
1038 elevated blood pressure on exercise capacity in elite athletes. Int J Cardiol. 2015
1039 Feb;180:171–7.

1040 22. McKeever KH, Antas LA, Kearns CF. Endothelin response during and after exercise in
1041 horses. Vet J Lond Engl 1997. 2002 Jul;164(1):38–46.

1042 23. R. Foundation for Statistical Computing, Vienna, Austria. R: The R Project for Statistical
1043 Computing [Internet]. [cited 2022 Jan 17]. Available from: https://www.r-project.org/

1044 24. Kingsley NB, Kern C, Creppe C, Hales EN, Zhou H, Kalbfleisch TS, et al. Functionally
1045 Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq.
1046 Genes. 2019 Dec 18;11(1):3.

1047 25. Kingsley NB, Hamilton NA, Lindgren G, Orlando L, Bailey E, Brooks S, et al. "Adopt-a-
1048 Tissue" Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare.
1049 Front Genet. 2021 Mar 26;12:1–9.

1050 26. Liu X, Orlando L. mapDATAge: a ShinyR package to chart ancient DNA data through
1051 space and time. Wren J, editor. Bioinformatics. 2022 Aug 10;38(16):3992–4.

1052 27. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al.
1053 Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc
1054 Natl Acad Sci [Internet]. 2014 Dec 30 [cited 2022 Aug 17];111(52). Available from:
1055 https://pnas.org/doi/full/10.1073/pnas.1416991111

1056 28. Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, Albrechtsen A, et al.
1057 Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to
1058 subarctic environments. Proc Natl Acad Sci. 2015 Dec 15;112(50):6889–97.

1059 29. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al.
1060 Ancient genomic changes associated with domestication of the horse. Science. 2017 Apr
1061 28;356(6336):442–5.

1062 30. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, Schubert M, et al. Ancient
1063 genomes revisit the ancestry of domestic and Przewalski's horses. Science. 2018 Apr
1064 6;360(6384):111–4.

1065 31. Librado P, Khan N, Fages A, Kusliy MA, Suchan T, Tonasso-Calvière L, et al. The origins
1066 and spread of domestic horses from the Western Eurasian steppes. Nature. 2021
1067 Oct;598(7882):634–40.

1068 32. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web
1069 server for functional enrichment analysis and conversions of gene lists (2019 update).
1070 Nucleic Acids Res. 2019 Jul 2;47(1):191–8.

1071 33. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, Zaibert V, et al. The Earliest
1072 Horse Harnessing and Milking. Science. 2009 Mar 6;323(5919):1332–5.

1073 34. Carneiro FRG, Ramalho-Oliveira R, Mognol GP, Viola JPB. Interferon Regulatory Factor
1074 2 Binding Protein 2 Is a New NFAT1 Partner and Represses Its Transcriptional Activity.
1075 Mol Cell Biol. 2011 Jul 15;31(14):2889–901.

1076 35. Miano JM. Role of serum response factor in the pathogenesis of disease. Lab Invest. 2010
1077 Sep;90(9):1274–84.

1078 36. Jeannotte L, Gotti F, Landry-Truchon K. Hoxa5: A Key Player in Development and
1079 Disease. J Dev Biol. 2016 Mar 25;4(2):13.

1080 37. Zheng T, Huang J, Xiang X, Li S, Yu J, Qu K, et al. Systematical analysis reveals a strong
1081 cancer relevance of CREB1-regulated genes. Cancer Cell Int. 2021 Dec;21(1):530.

1082 38. Bondurand N, Dufour S, Pingault V. News from the endothelin-3/EDNRB signaling
1083 pathway: Role during enteric nervous system development and involvement in neural crest-
1084 associated disorders. Dev Biol. 2018 Dec;444:S156–69.

1085 39. Baynash AG, Hosoda K, Giaid A, Richardson JA, Emoto N, Hammer RE, et al. Interaction
1086 of endothelin-3 with endothelin-B receptor is essential for development of epidermal
1087 melanocytes and enteric neurons. Cell. 1994 Dec;79(7):1277–85.

1088 40. Kusafuka T, Wang Y, Puri P. Mutation analysis of the RET, the endothelin-B receptor, and
1089 the endothelin-3 genes in sporadic cases of Hirschsprung's disease. J Pediatr Surg. 1997
1090 Mar;32(3):501–4.

1091 41. Kusafuka T, Puri P. Mutations of the endothelin-B receptor and endothelin-3 genes in
1092 Hirschsprung's disease. Pediatr Surg Int. 1997 Jan;12(1):19–23.

1093 42. Monti L, Arrigucci U, Rossi A. Insights into Endothelin-3 and Multiple Sclerosis. Biomol
1094 Concepts. 2020 Jun 25;11(1):137–41.

1095 43. Read AP, Newton VE. Waardenburg syndrome. J Med Genet. 1997 Aug 1;34(8):656–65.

1096 44. Emori T, Hirata Y, Ohta K, Kanno K, Eguchi S, Imai T, et al. Cellular mechanism of
1097 endothelin-1 release by angiotensin and vasopressin. Hypertension. 1991 Aug;18(2):165–
1098 70.

1099 45. Imai T, Hirata Y, Emori T, Yanagisawa M, Masaki T, Marumo F. Induction of endothelin-1
1100 gene by angiotensin and vasopressin in endothelial cells. Hypertension. 1992
1101 Jun;19(6_pt_2):753–7.

1102 46. Rossi NF. Effect of endothelin-3 on vasopressin release in vitro and water excretion in vivo
1103 in Long-Evans rats. J Physiol. 1993 Feb;461:501–11.

1104 47. Rossi NF. Cation channel mechanisms in ET-3-induced vasopressin secretion by rat
1105 hypothalamo-neurohypophysial explants. Am J Physiol-Endocrinol Metab. 1995
1106 Mar;268(3):E467–75.

1107 48. Weitzberg E, Hemsen A, Lundberg JM, Ahlborg G. ET-3 is extracted by and induces potent
1108 vasoconstriction in human splanchnic and renal vasculatures. J Appl Physiol. 1995
1109 Oct;79(4):1255–9.

1110 49. Tycho Vuurmans JL, Boer P, Koomans HA. Effects of endothelin-1 and endothelin-1-
1111 receptor blockade on renal function in humans. Nephrol Dial Transplant. 2004 Nov
1112 1;19(11):2742–6.

1113 50. Weitzberg E, Ahlborg G, Lundberg JM. Long-lasting vasoconstriction and efficient
1114 regional extraction of endothelin-1 in human splanchnic and renal tissues. Biochem
1115 Biophys Res Commun. 1991 Nov 14;180(3):1298–303.

1116 51. Grubb T, Edner A, Frendin JH, Funkquist P, Rydén A, Nyman G. Oxygenation and plasma
1117 endothelin-1 concentrations in healthy horses recovering from isoflurane anaesthesia
1118 administered with or without pulse-delivered inhaled nitric oxide. Vet Anaesth Analg. 2013
1119 Nov;40(6):e9–18.

1120 52. Benamou AE, Art T, Marlin DJ, Roberts CA, Lekeux P. Effect of exercise on
1121 concentrations of immunoreactive endothelin in bronchoalveolar lavage fluid of normal
1122 horses and horses with chronic obstructive pulmonary disease. Equine Vet J Suppl. 1999
1123 Jul;(30):92–5.

1124 53. Heesch CM, Kline DD, Hasser EM. Control Mechanisms of the Circulatory System. In:
1125 Duke´s Physiology of Domestic Animals. 13th ed. John Wiley & Sons, Inc.; 2015. p. 353.

1126 54. Navas de Solis C, Slack J, Boston RC, Reef VB. Hypertensive cardiomyopathy in horses:
1127 5 cases (1995-2011). J Am Vet Med Assoc. 2013 Jul 1;243(1):126–30.

1128 55. Moore JN, Garner HE, Coffman JR. Haematological changes during development of acute
1129 laminitis hypertension. Equine Vet J. 1981 Oct;13(4):240–2.

1130 56. Bailey SR, Habershon-Butcher JL, Ransom KJ, Elliott J, Menzies-Gow NJ. Hypertension
1131 and insulin resistance in a mixed-breed population of ponies predisposed to laminitis. Am
1132 J Vet Res. 2008 Jan;69(1):122–9.

1133 57. Poole DC, Erickson HH. Exercise Physiology of Terrestrial Animals. In: Duke´s
1134 Physiology of Domestic Animals. 13th ed. John Wiley & Sons, Inc.; 2015. p. 444.

1135 58. Kline H, Foreman J. Heart and spleen weights as a function of breed and somatotype.
1136 Equine Exerc Physiol. 1991;(3):17–21.

1137 59. Lorello O, Heliczer N, Casoni D, Schüpbach G, Navas de Solis C. Correlation of Blood
1138 Pressure With Splenic Volume in Horses, Daily Variation in Blood Pressure, and "White
1139 Coat Hypertension." J Equine Vet Sci. 2018 Apr 1;63:41–5.

1140 60. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, et al. Depletion of
1141 Abundant Plasma Proteins and Limitations of Plasma Proteomics. J Proteome Res. 2010
1142 Oct 1;9(10):4982–91.

1143 61. Scoppetta F, Tartaglia M, Renzone G, Avellini L, Gaiti A, Scaloni A, et al. Plasma protein
1144 changes in horse after prolonged physical exercise: a proteomic study. J Proteomics. 2012
1145 Jul 19;75(14):4494–504.

1146 62. Donovan DC, Jackson CA, Colahan PT, Norton N, Hurley DJ. Exercise-induced alterations
1147 in pro-inflammatory cytokines and prostaglandin F2alpha in horses. Vet Immunol
1148 Immunopathol. 2007 Aug 15;118(3–4):263–9.

1149 63. Velie BD, Fegraeus KJ, Solé M, Rosengren MK, Røed KH, Ihler CF, et al. A genome-wide
1150 association study for harness racing success in the Norwegian-Swedish coldblooded trotter
1151 reveals genes for learning and energy metabolism. BMC Genet. 2018 Dec;19(1):80.

1152 64. Goff JP. The endocrine system. In: Duke´s Physiology of Domestic Animals. 13th ed. John
1153 Wiley & Sons, Inc.; 2015. p. 635–6.

1154 65. Kostrzewa-Nowak D, Kubaszewska J, Nowakowska A, Nowak R. Effect of Aerobic and
1155 Anaerobic Exercise on the Complement System of Proteins in Healthy Young Males. J
1156 Clin Med. 2020 Jul 23;9(8):E2357.

1157 66. Barton MH, Williamson L, Jacks S, Norton N. Effects on plasma endotoxin and eicosanoid
1158 concentrations and serum cytokine activities in horses competing in a 48-, 83-, or 159-km
1159 endurance ride under similar terrain and weather conditions. Am J Vet Res. 2003
1160 Jun;64(6):754–61.

1161 67. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. Nat Rev
1162 Immunol. 2007 Oct;7(10):803–15.

1163 68. Dinh QN, Drummond GR, Sobey CG, Chrissobolis S. Roles of inflammation, oxidative
1164 stress, and vascular dysfunction in hypertension. BioMed Res Int. 2014;2014:1–11.

1165 69. Contrepois K, Wu S, Moneghetti KJ, Hornburg D, Ahadi S, Tsai MS, et al. Molecular
1166 Choreography of Acute Exercise. Cell. 2020 May;181(5):1112-1130.e16.

1167 70. Yan W, Bi HL, Liu LX, Li NN, Liu Y, Du J, et al. Knockout of immunoproteasome subunit
1168 β2i ameliorates cardiac fibrosis and inflammation in DOCA/Salt hypertensive mice.
1169 Biochem Biophys Res Commun. 2017 Aug;490(2):84–90.

1170 71. Mootz M, Jakwerth CA, Schmidt-Weber CB, Zissler UM. Secretoglobins in the big picture
1171 of immunoregulation in airway diseases. Allergy. 2022;77(3):767–77.

1172 72. Côté O, Clark ME, Viel L, Labbé G, Seah SYK, Khan MA, et al. Secretoglobin 1A1 and
1173 1A1A differentially regulate neutrophil reactive oxygen species production, phagocytosis
1174 and extracellular trap formation. PloS One. 2014;9(4):e96217.

1175 73. Metallinos, A.T. Bowling DL, Rine J. A missense mutation in the endothelin-B receptor
1176 gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung
1177 Disease. Mamm Genome. 1998 Jun 1;9(6):426–31.

1178 74. Santschi EM, Purdy AK, Valberg SJ, Vrotsos PD, Kaese H, Mickelson JR. Endothelin
1179 receptor B polymorphism associated with lethal white foal syndrome in horses. Mamm
1180 Genome. 1998 Apr;9(4):306–9.

1181 75. Yang G. A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal
1182 white foal syndrome (LWFS); a horse variant of Hirschsprung disease. Hum Mol Genet.
1183 1998 Jun 1;7(6):1047–52.

1184 76. Gauff F, Patan-Zugaj B, Licka TF. Hyperinsulinaemia increases vascular resistance and
1185 endothelin-1 expression in the equine digit: Endothelin-1 and vasoconstriction after
1186 hyperinsulinaemia. Equine Vet J. 2013 Sep;45(5):613–8.

1187 77. Gehlen H, Shety T, El-Zahar H, Hofheinz I. Measurement of plasma endothelin-1
1188 concentration in healthy horses and horses with cardiac disease during rest and after
1189 exercise. J Vet Med Sci. 2019;81(2):263–8.

1190 78. Giordano A, Castagnetti C, Panzani S, Paltrinieri S, Freccero F, Veronesi MC. Endothelin
1191 1 in healthy foals and in foals affected by neonatal diseases. Theriogenology. 2015
1192 Sep;84(5):667–73.

1193 79. Bye A, Klevjer M, Ryeng E, Silva GJJ da, Moreira JBN, Stensvold D, et al. Identification
1194 of novel genetic variants associated with cardiorespiratory fitness. Prog Cardiovasc Dis.
1195 2020 May;63(3):341–9.

1196 80. Kim DS, Wheeler MT, Ashley EA. The genetics of human performance. Nat Rev Genet.
1197 2022 Jan;23(1):40–54.

1198 81. Rautureau Y, Schiffrin EL. Endothelin in hypertension: an update. Curr Opin Nephrol
1199 Hypertens. 2012 Mar;21(2):128–36.

1200 82. Hui L, DelMonte T, Ranade K. Genotyping Using the TaqMan Assay. Curr Protoc Hum
1201 Genet [Internet]. 2008 Jan [cited 2022 Sep 12];56(1). Available from:
1202 https://onlinelibrary.wiley.com/doi/10.1002/0471142905.hg0210s56

1203 83. Árnason T. The Importance of Different Traits in Genetic Improvement of Trotters. In
1204 Proceedings of World Congress on Genetics Applied to Livestock Production. University
1205 of Guelph, Guelph, Canada; 1994.

1206 84. Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al.
1207 Improved reference genome for the domestic horse increases assembly contiguity and
1208 composition. Commun Biol. 2018 Dec;1(1):197.

1209 85. Jurinke C, van den Boom D, Cantor CR, Köster H. Automated genotyping using the DNA
1210 MassArray technology. Methods Mol Biol Clifton NJ. 2001;170:103–16.

1211 86. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, et al. Estimation and
1212 Tests of Haplotype-Environment Interaction when Linkage Phase Is Ambiguous. Hum
1213 Hered. 2003;55(1):56–65.

1214 87. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1215 Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

1216 88. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
1217 features. Bioinforma Oxf Engl. 2010 Mar 15;26(6):841–2.

1218 89. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
1219 performance genomics data visualization and exploration. Brief Bioinform. 2013
1220 Mar;14(2):178–92.

1221 90. The ENCODE Project Consortium, Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal
1222 V, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes.
1223 Nature. 2020 Jul 30;583(7818):699–710.

1224 91. Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJM, et al. ORegAnno 3.0: a
1225 community-driven resource for curated regulatory annotation. Nucleic Acids Res. 2016 Jan
1226 4;44(D1):D126-132.

1227 92. Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roider HG, et al.
1228 Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and
1229 regulatory SNPs. Nat Protoc. 2011 Dec;6(12):1860–9.

1230 93. Felkel S, Vogl C, Rigler D, Dobretsberger V, Chowdhary BP, Distl O, et al. The horse Y
1231 chromosome as an informative marker for tracing sire lines. Sci Rep. 2019 Dec;9(1):6095.

1232 94. Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al.
1233 Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Curr
1234 Biol. 2015 Oct;25(19):2577–83.

1235 95. Orlando L, Librado P. Origin and Evolution of Deleterious Mutations in Horses. Genes.
1236 2019 Aug 28;10(9):649.

1237 96. Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, et al.
1238 Characterization of ancient and modern genomes by SNP detection and phylogenomic and
1239 metagenomic analysis using PALEOMIX. Nat Protoc. 2014 May;9(5):1056–82.

1240 97. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation
1241 Sequencing Data. BMC Bioinformatics. 2014 Dec;15(1):356.

1242 98. Söder J, Bröjer JT, Nostell KE. Interday variation and effect of transportation on indirect
1243 blood pressure measurements, plasma endothelin-1 and serum cortisol in Standardbred and
1244 Icelandic horses. Acta Vet Scand. 2012 Jun 10;54:37.

1245 99. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method
1246 for proteome analysis. Nat Methods. 2009 May;6(5):359–62.

1248 Table 7. Fold-change and P-value for the top most upregulated proteins in each comparison in the quantitative proteomic analysis, including all

1249 proteins significant for more than one comparison

					Comparison							
					1. SPH exercise vs. rest		2. EPH exercise vs. rest		3. EPH vs. SPH rest		5. EPH vs. SPH Paired ratio	
Protein function		**Protein**	**Gene**	**Accession**	**FC**	**P-value**	**FC**	**P-value**	**FC**	**P-value**	**FC**	**P-value**
Immune System	*Adaptive immune system	Proteasome subunit beta	*PSMB10*	F6SZ35					6.03	0.03	8.2	0.03
	*Adaptive immune system, angiogenesis	CD34 molecule	*CD34*	F6S1S0	4.33	0.03			3.42	0.03	17.97	0.04
	*Antigen binding	Ig-like domain-containing protein	*Unknown*	A0A3Q2HY39					5.58	0.02		
	*Antigen binding	Ig-like domain-containing protein	*Unknown*	A0A3Q2GXU9					5.44	0.03		

*Complement activation, complement and coagulation cascades	Complement C1r	*C1R*	F6Z5L1	4.69	0.02			4.09	0.03
*Complement activation, inflammatory response	C4a, anaphylatoxin	*LOC100059239*	F6XSF7			2.08	0.04	5.6	0.04
*Complement activation, inflammatory response	C3-beta-c	*LOC100060505*	A0A3Q2HWQ6	2.54	0.05			5.83	0.05
*Complement and coagulation cascades, metabolism of proteins	Carboxypeptidase B2	*CPB2*	F6PSH3	4.31	0.01			7.94	0.03

	*Innate immune system	Cystatin C	*CST3*	A0A3Q2I5T2	2.32	0.02			4.23	0.007	5.91	0.02
	*Innate immune system	Canopy FGF signaling regulator 3	*CNPY3*	F7CNK7	5.05	0.04						
	*Innate immune system, neutrophil degranulation, angiogenesis	Serine peptidase inhibitor, Kunitz type 1	*SPINK5*	A0A5F5Q0T5	3.22	0.04			4.65	0.03	18.49	0.04
	*Small molecule binding, allergenic capacity	Equ c 2 allergen	*LOC106781327*	A0A3G1Z1K7			0.28	0.001				
	*Small molecule binding, allergenic capacity	Major allergen Equ c 1	*LOC100056556*	A0A5F5PQP7			0.24	0.04				
Metabolism	*Metabolism gluconeogenesis	Malate dehydrogenase 1	*MDH1*	A0A5F5PWG8	4.11	0.05						

	*Metabolism of proteins	Tripeptidyl aminopeptidase	*TPP1*	F6TJX5					2.88	0.04	10.13	0.05
Neuronal	*Neuronal system	Neurexin 2	*NRXN2*	F6YED9					5.85	0.005		
Others	*Serine-type endopeptidase inhibitor activity	Serine peptidase inhibitor Kazal type 5	*SPINT1*	A0A3Q2IKF1							6.35	0.04
	*Amino acid metabolism	T1-TrpRS	*WARS1*	A0A3Q2KTL8	3.51	0.02			3.48	0.01	13.11	0.005
	*Binding of androgens and other steroids	Secretoglobin family 1D member	*LOC102150542*	A0A3Q2HLK8			0.17	0.0009				
	*Hemostasis and regulation of insulin-like growth factor (IGF) transport and uptake	Fibrinogen gamma chain	*FGG*	A0A5F5PPB8	2.8	0.05					4.28	0.05