

RESEARCH ARTICLE

# An endothelial regulatory module links blood pressure regulation with elite athletic performance

Kim Fegraeus<sup>1</sup> <sup>\*</sup>, Maria K. Rosengren<sup>2</sup> <sup>2</sup> <sup>2</sup> <sup>2,3</sup>, Ludovic Orlando<sup>4</sup>, Magnus Åbrink<sup>5</sup> <sup>5</sup>, Ahmad Jouni<sup>2</sup>, Brandon D. Velie<sup>6</sup> <sup>6</sup>, Amanda Raine<sup>1</sup>, Beate Egner<sup>7</sup>, C Mikael Mattsson<sup>8</sup> <sup>8</sup>, Karin Lång<sup>9</sup> <sup>9</sup>, Artemy Zhigulev<sup>10</sup> <sup>10</sup>, Hanna M. Björck<sup>9</sup> <sup>9</sup>, Anders Franco-Cereceda<sup>11</sup> <sup>11</sup>, Per Eriksson<sup>9</sup> <sup>9</sup>, Göran Andersson<sup>2</sup> <sup>2</sup>, Pelin Sahlén<sup>10</sup> <sup>10</sup>, Jennifer R. S. Meadows<sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup> <sup>12</sup>

Association (Stockholm, Sweden), but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. However, data are available from The Swedish Trotting Association (Svensk Travsport, Box 201 51, 161 02 Bromma, Sweden). The contact person is Christina Olsson, head of the breeding department at The Swedish Trotting Association, [kundtjanst@travsport.se](mailto:kundtjanst@travsport.se). The WGS data has been deposited in the BioProject database (NCBI repository SRA) with ID PRJNA1045044 according to their guidelines. According to the ethical permit related to the study (application number 2006/784-31/1 and 2012/1633-31/4) approved by the Human Research Ethics Committee at Karolinska Stockholm, Sweden, the sequencing data from living patients cannot be shared due to patient privacy regulations. We therefore only share the summary or processed form of the sequencing data in [S5–S7 Tables](#). More information can be found at Etikprövningsmyndigheten, Box 2110, 750 02, Uppsala, or [registrator@etikprovning.se](mailto:registrator@etikprovning.se).

**Funding:** The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) (to GL) and The Swedish Research Council (VR) (to GL). This project has also received funding from the CNRS, University Paul Sabatier (AnimalFarm IRP) (to LO), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 681605-PEGASUS) (to LO). K.F. and A.R. were supported by grants from FORMAS (2020-01135) (to AR). K.F. received salary from FORMAS (2020-01135). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

A previous study discovered that a genomic region close to the Endothelin3 gene was associated with harness racing performance. Here, careful phenotypic documentation of athletic performance and blood pressure measurements in horses, followed by state-of-the-art genomics, allowed us to identify a 5.5 kb regulatory region located approximately 50 kb 3' of the *EDN3* gene. A comparative analysis of the region using human HiCap data supported a regulatory role as, in endothelial cells, interaction was observed between the region and multiple genes relevant to blood pressure regulation and athletic performance. Long range cis-regulatory modules are critical for cooperatively controlling multiple genes located within transcriptionally active domains. We measured blood pressure in Coldblooded trotters during exercise and demonstrated that horses with two copies of the elite-performing haplotype had lower blood pressure during exercise and better racing performance results, compared to horses with two copies of the sub-elite performing haplotype. In addition, horses with the elite-performing haplotype also had higher levels of Endothelin3 in plasma. The results reported here are important for understanding the biological mechanisms behind blood pressure regulation in relation to racing performance in both horses and humans.

## Background

Despite the apparent potential of domestic animal models to provide valuable insights into the natural biological mechanisms driving human traits and disease, their use to date has been limited. In the past, the use of domestic animals as models for genomic research has provided basic knowledge concerning gene function and biological mechanisms and a complementary view on genotype–phenotype relationships [1]. Recent advances in the availability and quality of human and mammalian reference genomes, plus the technological advances required for their alignment, are revealing both coding and non-coding conserved bases, key to unraveling shared gene function and regulation [2,3]. The horse is one of the most popular species for studying athletic performance. They have been intensively selected for centuries, to encompass the optimal physical capacity for strength, speed, and endurance [4,5]. Additionally, their recent population history, involving closed populations selected for 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 [1,4,6,7]. Previous research in horses has begun to unravel the genetics of complex traits, such as muscle mass and locomotion patterns, on athletic performance, with many potential candidate genetic variants identified [8–14]. However, it is much more challenging to understand the mechanisms by which the identified variants exert their functions.

We have previously used a unique, three breed admixture Nordic horse model, to study the genetics of athletic performance traits [15,16]. The basis for this model is that racing Coldblooded trotters (CBTs) predominantly originate from the North Swedish Draught horses (NSDs), a sturdy breed used in farming and forestry. It is well established that some crossbreeding occurred between Standardbreds (SBs) (the most commonly used breed for harness racing) and CBTs before obligatory paternity testing was introduced in Sweden in the 1960s. However, a remarkable improvement in the racing performance of the CBT has occurred during the last fifty years. In part, it is likely that this improvement could be explained by crossbreeding, with a marked increase of favorable genetic variants originating from SBs and

introduced to the CBTs. This process should leave “genetic footprints” in the genome of CBT in the form of chromosome segments originating from SBs. We have identified such a footprint using pair-wise allele frequency distribution and selective sweep mapping [15]. Notably, our model, which is solely based on the hybrid origin of the CBT, provides a unique opportunity to study genes influencing body constitution and complex morphological traits of importance for racing success.

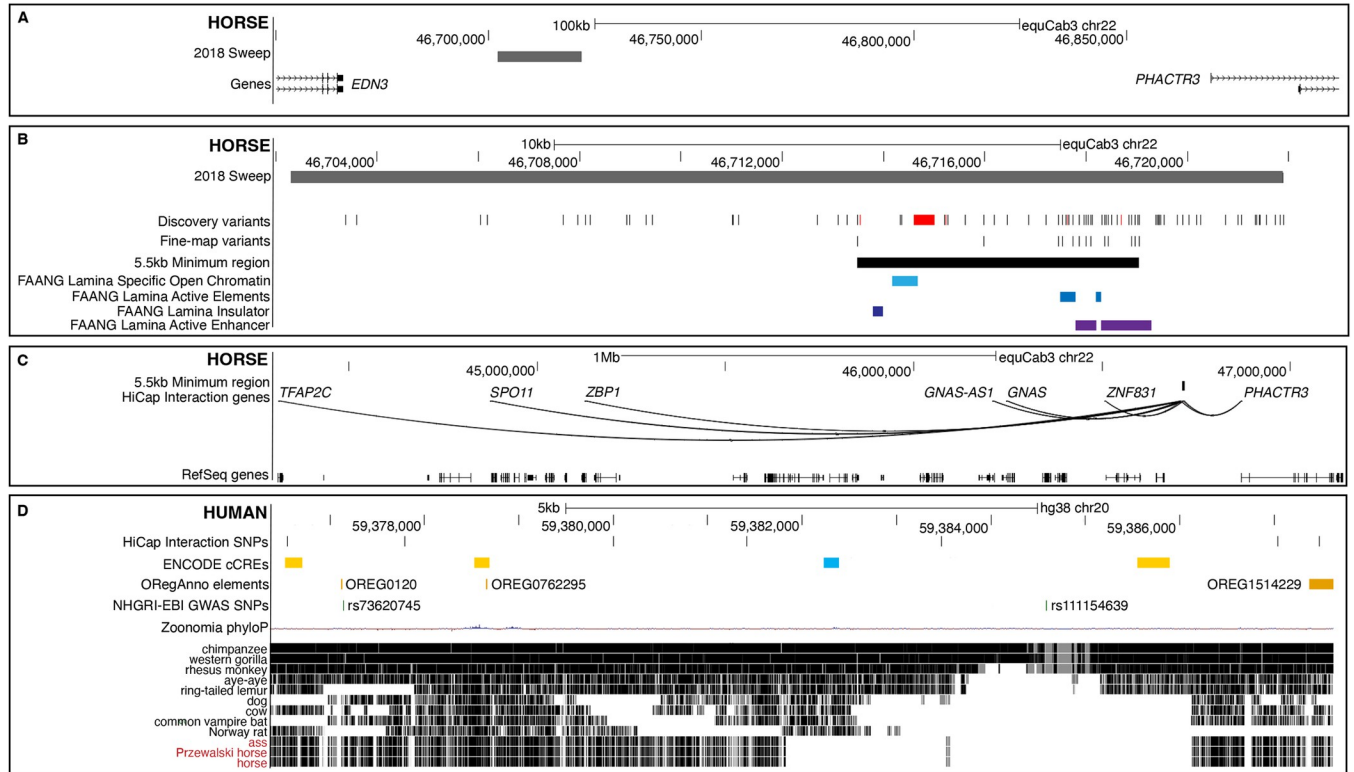
The harness racing selective sweep study revealed a 19.6 kb genomic region on chromosome 22, located approximately 50 kb downstream of the *Endothelin 3* (*EDN3*) gene, under selection in CBT [15]. Five SNVs in high linkage disequilibrium (LD,  $r^2 = 0.92\text{--}0.94$ ) were significantly associated with racing performance, including the number of victories, earnings, and racing times. Variant rs69244086 C>T (EquCab3.0, ECA22:46717860) showed the strongest association within the sweep region, 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. We hypothesized that the identified region might contain a regulatory element influencing either the expression of *EDN3* or other genes nearby [15].

Genomic regulatory regions, such as putative enhancers, are difficult to study simply because they are hard to characterize [17]. They can be cell type and condition specific, acting both upstream and downstream of target genes, or over long genomic distances, where genome looping brings them into proximity to their target genes [18]. Targeted chromosome conformation capture (HiCap) is an experimental method that can detect genomic loops mediating regulatory interactions between such regions and the promoters of their target genes [19,20]. In this study, we fine-mapped the harness racing selective sweep, and used comparative data to interrogate its potential function. In addition to equine epigenetic data, we have used HiCap datasets produced using twelve different human cell types, including vascular endothelial cells, to indicate a role of our potential regulatory region and the target genes involved. Further, since the sweep region flanks the *GNAS-EDN3* region identified in multiple human blood pressure GWAS [21,22], we evaluated the effect of the minimal sweep horse haplotypes on blood pressure, as well as EDN1 and EDN3 plasma levels, before and during exercise. Finally, key genetic variants within the sweep region were characterized for spatial and temporal distribution within past and present horse populations. Taken together, the results presented here indicate the identification of a regulatory unit, likely important for vascular traits and performance in horses and humans.

## Results

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

With a four-step process, we reduced the 19.6 kb sweep region from the previous study [15] to a minimum shared 5.5 kb region (Fig 1A and 1B). In step one, we repeated the racing performance association analysis using 251 more horse samples (total  $n = 629$ ) with SNV genotypes extracted from the 670K Axiom Equine Genotyping Array. However, supplementing the original 378 CBTs with additional samples did not narrow the 19.6 kb region [15]. Seven SNVs within the sweep (five significantly associated with racing performance and two flanking SNVs) [15] were extracted from the array, including the most significant one: rs69244086 [15]. Pairwise LD centered on the previously most associated SNV, rs69244086 [15], resulted in the same core five SNV region (ECA22:46,717,451–46,718,964,  $r^2 > 0.6$ ), and analysis of linear models showed that these SNVs remained significantly associated with harness racing performance traits (i.e., number of victories, earnings, and race times,  $P \leq 0.05$ ) (S1 Table).



**Fig 1. Sweep fine-mapping and functional exploration.** A) Location of the 2018 selective sweep. B) Fine-mapped region including discovery variants, fine-mapped SNVs associated with racing performance, the refined minimum sweep region as well as active elements in lamina from FAANG data. Discovery SNVs are indicated in black, deletions in red. C) HiCap interacting SNV and genes, lifted from humans. D) The span of the minimum sweep region covers an extended 11 kb space in the human reference genome, hg38. HiCap interacting SNVs are indicated in relation to known regulatory elements and SNVs reported from GWAS studies. Zoonomia phyloP scores and cactus alignments show the conservation of this region across mammals. The yellow boxes represent distal enhancer-like signatures and the blue box represent a CTCF-only.

<https://doi.org/10.1371/journal.pgen.1011285.g001>

In step two, we performed whole genome sequencing (WGS) of two CBTs and two SBs to increase variant density across the 19.6 kb sweep region (Fig 1B). In each breed, horses were selected to be homozygous for different alleles at SNV rs69244086 (CC and TT, respectively) and had either high (TT) or low (CC) earnings per start (S2 Table). From the 19.6 kb sweep region, 78 SNVs and six indels (one 400 bp deletion, one 23 bp deletion and four single bp INDELS) were identified (Fig 1B and S3 Table).

In step three, we prioritized variants for further analysis. The 400 bp deletion (ECA22:46,714,602–46,715,003) was absent in horses with low earnings per start and either heterozygous (SB) or homozygous (CBT) in elite performing horses. However, this variant was not significantly associated with racing performance traits following genotyping and analyses in a further 497 CBTs (linear models, S4 Table). Due to technology constraints, 25 of the available 82 single base pair variants were selected for genotyping in a larger horse material. Variant selection was based on available pooled allele frequency data matching admixture expectation in CBTs, SBs and NSD [16], variant spacing across the region, and success in MassArray design (see Methods). The 24 SNVs and one single base pair deletion (Fig 1B) were genotyped in 391 horses sampled from 12 different breeds (including 210 CBTs), as well as three Przewalski horses. SNV rs69244086 was not included in this set, but was either directly genotyped or imputed across the dataset. After quality control, 24 SNVs and 394 horses (including the Przewalskis) were available for further analysis. Pairwise LD calculations with SNV

**Table 1. Haplotype frequencies in different breeds.**

Haplotype (H) <sup>1</sup>	All breeds (n = 394)	CBT (n = 210)	Arabians (n = 29)	TB (n = 30)	SB (n = 45)	Ardennes (n = 20)	NSD (n = 18)	Exmoor (n = 23)	Icelandic (n = 6)
H1) AAGCGTTTCTCAAA	0.58	0.62	1.00	0.93	0.77	0.13	0.06	0.00	0.00
H2) GGTTACCCTCTGGG	0.41	0.38	0.00	0.07	0.23	0.78	0.92	1.00	0.75
H3) AAGCGTTTCTCGAA	<0.01	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00
H4) AAGCGTTTCTCAAG	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
H5) AAGTGTTCCTCAAG	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
H6) GGCGTTCCTCAAA	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
H7) GGCGTTCCTCAAA	<0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05

<sup>1</sup> Order of SNVs, rs395117226, rs397265747, rs396474304, rs69244081, rs69244084, rs69244085, rs69244086, rs69244088, rs69244089, rs396281591, rs394573286, rs69244091, rs69244093, rs69244095.

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

<https://doi.org/10.1371/journal.pgen.1011285.t001>

rs69244086 revealed 15 SNVs in strong LD ( $r^2 \geq 0.53$ , 14 SNVs with  $r^2 > 0.98$ ) (ECA22:46,708,983–46,719,042). LD decayed outside the region ( $r^2 < 0.07$ ), leaving 14 SNVs for haplotype analysis (Fig 1B). Phasing revealed in total 14 different haplotypes, with seven haplotypes found at a frequency  $> 2\%$  in the total sample set, or within each breed represented by five or more individuals (Table 1). Generalized linear model (GLM) regression in the R environment (23) was used to test the association between the two haplotypes harbored by CBTs ( $n = 165$ ) and SBs ( $n = 38$ ), and harness racing performance traits. For both CBTs and SBs, the H1 haplotype, carrying the rs69244086-T high-performance associated allele, was the most common (Table 2). In both breeds, the H2 haplotype demonstrated a significantly negative effect on all performance traits tested, except for the number of starts in SBs ( $P = 0.59$ ) and the number of wins in SBs ( $P = 0.06$ ) (Table 2). The above analysis showed that the minimum 5,564 bp, 14 SNV shared haplotype (ECA22: 46,713,478–46,719,042), significantly associated with racing performance in both CBTs and SBs. This allowed for the definition of H1 as an elite-performing haplotype (EPH: AAGCGTTTCTCAAA) and H2 as a sub-elite-performing haplotype (SPH: GGTTACCCTCTGGG). Of note, the reference EquCab3.0 haploid reference genome, generated from a Thoroughbred, represents the SPH haplotype.

In step four, we assessed the potential functional impact of the minimum shared 5.5 kb region using comparative data from both horse and human resources (Fig 1C and 1D). Epigenetic data drawn from nine tissues was generated by the Equine section of Functional Annotation of Animal Genomes (FAANG) [24] for two Thoroughbreds, both homozygous for the EPH haplotype. No clear signals were observed in the minimal region for four marks (H3K27ac, H3K27me3, H3K4me1, H3K4me3) measured across eight different tissues (adipose, brain, heart, liver, lung, muscle, ovary, skin). But in lamina we observed an insulator and an active enhancer in our 5.5 kb minimum shared region (Fig 1). To access additional functional data points, we lifted the minimum shared 5.5 kb region and contained SNVs to the human reference genome, hg38. In comparative analyses, the liftover indicated that, although not well conserved, the region had regulatory potential, including Open Regulatory Annotation database (OREGAnno) elements [25], ENCODE cCREs [18] and GTEx cis-eQTL variants

**Table 2. Haplotype frequencies, haplotype coefficient and P-values for the performance analysis in Coldblooded trotters (n = 165) and Standardbreds (n = 38).**

	Coldblooded trotters				Standardbreds			
	Haplotype H1		Haplotype H2		Haplotype H1		Haplotype H2	
Haplotype <sup>1</sup>	AAGCGTTTCTCAAA		GGTTACCCCTCTGGG		AAGCGTTTCTCAAA		GGTTACCCCTCTGGG	
Frequency	0.61		0.39		0.78		0.22	
	Haplotype coefficient <sup>2</sup>	P-value <sup>3</sup>	Haplotype coefficient	P-value <sup>3</sup>	Haplotype coefficient	P-value <sup>3</sup>	Haplotype coefficient	P-value <sup>3</sup>
No. of starts <sup>5</sup>	-	-	-1.02	<0.001	-	-	0.25	0.59
No. of wins <sup>5</sup>	-	-	-0.12	<0.001	-	-	-0.12	0.06
No. of placings <sup>5</sup>	-	-	-0.10	<0.001	-	-	-0.12	<b>0.01</b>
Earnings (SEK) <sup>4,5</sup>	-	-	-0.39	<0.001	-	-	-1.30	<0.001
Earnings/start <sup>6</sup>	-	-	-0.37	<0.001	-	-	-1.08	<0.001
Best time (sec/km) <sup>5,6</sup>	-	-	0.05	<0.001	-	-	0.31	<b>0.001</b>

<sup>1</sup> Order of SNVs. rs395117226, rs397265747, rs396474304, rs69244081, rs69244084, rs69244085, rs69244086, rs69244088, rs69244089, rs396281591, rs394573286, rs69244091, rs69244093, rs69244095

<sup>2</sup> Haplotype coefficient, or effect size. This is the effect of each haplotype compared to the base haplotype (H1), i.e., one copy of the H2 haplotype will reduce the square root transformed number of starts with ~1.

<sup>3</sup> A GLM regression analysis was performed in R. Sex, age, birth country, number of starts (when applicable) and DMRT3 genotype (only in CBTs) were included as fixed effects. Significant results ( $P \leq 0.05$ ) in bold

<sup>4</sup> SEK, Swedish kronor

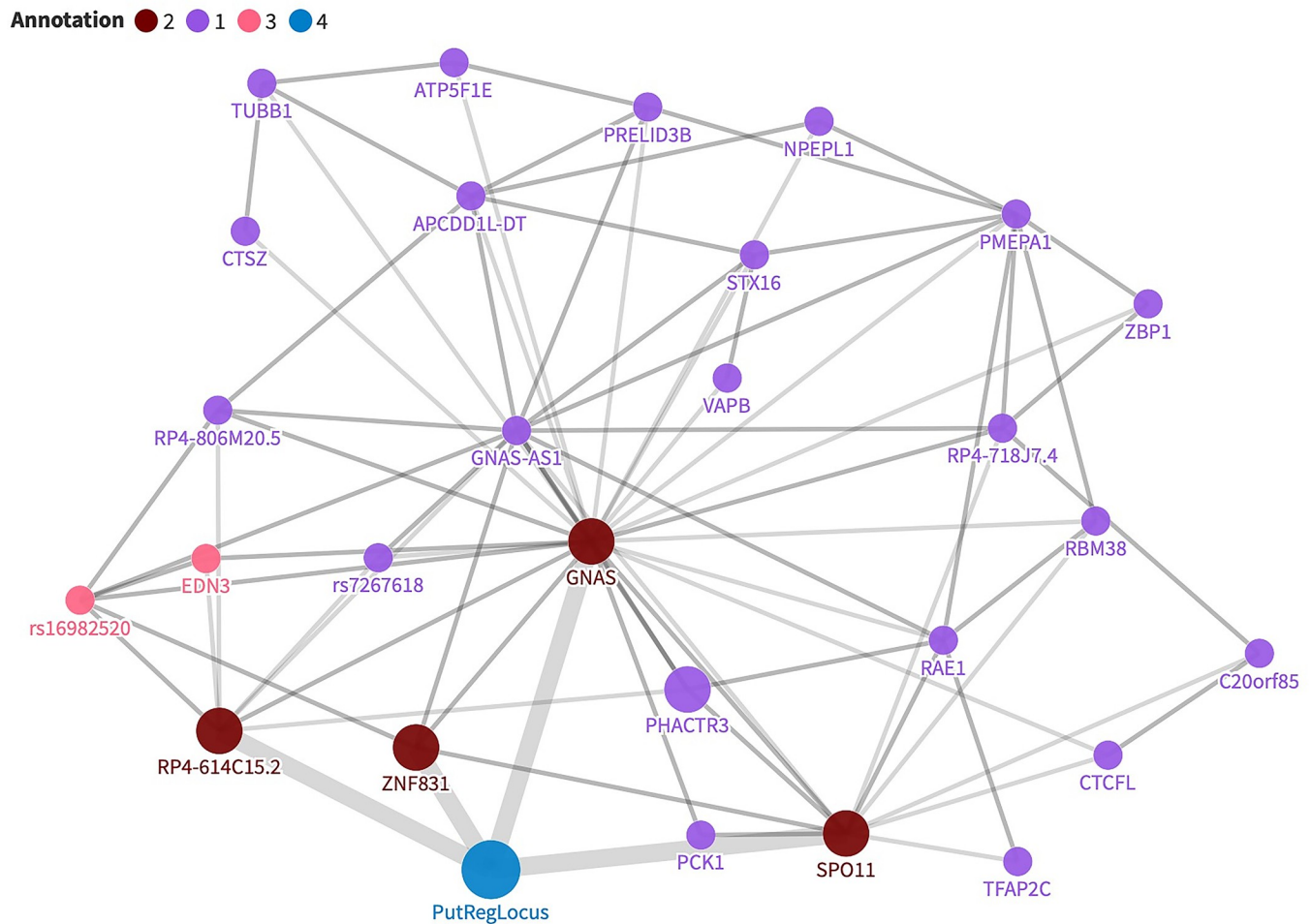
<sup>5</sup> Log transformed values ( $\log_{10} + 1$ ) were used for wins and placings. Number of starts were square root transformed. Earnings and earnings per start were transformed applying  $\ln(\text{earnings} + 1000)$

<sup>6</sup> 173 CBTs and 36 SBs were included in the analysis of best racing time

<https://doi.org/10.1371/journal.pgen.1011285.t002>

regulating *EDN3* in esophagus mucosa tissue [26] (Fig 1D). We further investigated the regulatory potential of the minimum shared 5.5 kb region using comparative chromatin interaction profiles in 12 different human cell types, including iPS cells and a neural cell line (see Methods and S5 Table). SNVs lifted to the minimum 5.5 kb region showed interactions with the promoters of multiple genes, but only in the vascular endothelial cell datasets (Fig 1C and 1D and S5 Table). With a relaxed threshold, requiring support evidence from at least two samples, direct interaction between human to horse lifted SNVs and seven genes were observed (Fig 1C). These interactions included the proximal cell membrane-cytoskeleton dynamics gene, *PHACTR3* [27], through to the megabase distal cell reprogramming transcription factor gene, *TFAP2C* [28]. While specific SNVs in humans are unlikely to have the exact same role in horses, we never-the-less explored the human direct, and indirect, interaction network to gain a wider perspective of genes which may be regulated by the minimum shared 5.5 kb region (Fig 2). Here, support was required from five samples, and all human loci are considered, even if these are not currently supported with horse annotation evidence. We see that the 5.5 kb region (putative regulatory locus, PutRegLocus) interacts with the *EDN3* promoter indirectly, via *GNAS* and two other promoters (Fig 2). GWAS variant rs16982520, located within *ZNF831*, is associated with hypertension [29,30], systolic blood pressure [31] and mean arterial pressure [32] and interacts directly with *EDN3* and four direct interactors of the putative regulatory region (*GNAS*, *RP4-614C15.2*, *SPO11* and *ZNF831*), suggesting a complex regulatory pattern. In total, there were 15 protein-coding genes that directly or indirectly interacted with the putative regulatory region (S6 Table). We performed a gene enrichment analysis for these 15 genes and found that several human phenotype terms related to thyroid hormone regulation were enriched (S7 Table).

Moving back to the horse genome, we investigated the *in silico* regulatory potential of each allele from the 14 SNVs associated with racing performance and included in the minimum



**Fig 2. The direct and indirect interactions of the minimum haplotype lifted over to hg38 (PutRegLocus, blue) in vascular endothelial cells.** There were four direct interactors of the locus (marked in dark red). Some of the direct interactors of the PutRegLocus interacted with EDN3 promoter and rs16982520 variant (pink). The direct interactors of PutRegLocus are connected with thick edges.

<https://doi.org/10.1371/journal.pgen.1011285.g002>

shared 5.5 kb region, to alter transcription factor binding. Using the EquCab3 genome reference and the Homo sapiens Comprehensive Model Collection (HOCOMOCO) transcription factor binding model database [33] in motifbreakR [34], we found that all alleles had the potential to cause alterations of high effect. SNVs rs69244086 C>T and rs69244089 T>C are illustrated as examples in S1 Fig. In S8 Table, the transcription factor binding scores for all analyzed variable sites are listed.

### The elite performing haplotype is significantly associated with improved racing performance

Given that all 14 associated variants within the minimum shared 5.5 kb locus were in perfect LD in CBTs and SBs, we used the genotypes at rs69244089 as a proxy for EPH (C allele) and SPH (T allele) haplotypic pairs and re-assessed the association between haplotypes and racing performance. For CBTs ( $n = 516$ ), EPH homo- or heterozygotes outperformed the SPH homozygotes (Table 3). In SBs, there were no statistically significant differences between the three genotypes (Table 4). However, when analyzing horses carrying at least one T allele versus CC horses, the results were significant for several performance traits (Table 4), including the

**Table 3. Performance results in 516 Coldblooded trotters for SNV rs69244089 T>C.**

Genotype <sup>2</sup>	TT (n = 63)		TC (n = 244)		CC (n = 209)		P <sup>3</sup>		
	Mean	Median	Mean	Median	Mean	Median	TT/TC	TT/CC	TC/CC
Performance trait <sup>1</sup>									
No. of starts	31.6	20.0	41.5	31.0	41.2	31.0	<b>0.01</b>	<b>0.02</b>	<b>0.99</b>
No. of wins	2.8	1.0	5.7	3.0	5.6	3.0	<b>0.006</b>	<b>0.05</b>	0.56
No. of placings (1–3)	9.1	4.0	14.2	9.0	13.9	8.0	<b>0.005</b>	<b>0.04</b>	0.54
Earnings (SEK)	209,486	51,772	453,084	148,197	425,448	116,500	<b>0.009</b>	<b>0.05</b>	0.65
Earnings per start (SEK)	4,760	2,939	9,666	4,362	8,188	3,748	<b>0.003</b>	0.07	0.26
Best km time (sec) <sup>4</sup>	91.1	91.5	88.5	88.3	89.0	89.0	<b>0.002</b>	<b>0.04</b>	0.37

<sup>1</sup> Log transformed values (log<sub>10</sub> + 1) were used for wins and placings. Earnings and earnings per start were transformed applying ln(earnings + 1000). Square root transformations were used for number of starts

<sup>2</sup> The C-allele is tagging the elite performance haplotype (EPH)

<sup>3</sup> A linear model was performed in R. Sex, age, birth country, number of starts and *DMRT3* genotypes were included as fixed effects (when significant). Least square means and Tukey were used as post hoc tests. Significant results ( $P \leq 0.05$ ) in bold

<sup>4</sup> Only races where the horse did not gallop were included

<https://doi.org/10.1371/journal.pgen.1011285.t003>

number of wins, placings and earnings. While the number of wins and placings was higher in horses with at least one T allele, the CC horses earned more money than the TT/TC horses (Table 4).

In both SBs and CBTs the rs69244089 genotypes distributed according to HWE ( $P = 0.16$  and  $P = 0.35$ , respectively). For breeds other than CBT and SB, each individual's performance status is unknown. However, a trend for an increased frequency of the C allele in traditional performance breeds was observed, e.g., Arabian horses, Thoroughbreds, and Warmbloods. In contrast, this allele was low or absent in draft horses and ponies (Table 5).

### Increased frequency of the favorable haplotype coincides with the time of horse domestication

For all 14 SNVs within the minimum shared haplotype in the trotters, we investigated the allele frequency over time using the mapDATAge package (35) and 431 previously characterized ancient genomes [4,36–40]. For all SNVs except two, rs69244085 and rs69244086, there has

**Table 4. Performance results in 271 Standardbreds for SNV rs69244089 T>C.**

Genotype <sup>2</sup>	TT (n = 7)		TC (n = 62)		CC (n = 202)		P <sup>3</sup>			
	Mean	Median	Mean	Median	Mean	Median	TT/TC	TT/CC	TC/CC	CC/ TT+TC
Perf. trait <sup>1</sup>										
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	<b>0.04</b>
No. of placings (1–3)	15.6	12.0	16.1	16.0	13.7	11.0	0.86	0.96	0.08	<b>0.03</b>
Earnings (SEK)	700,008	322,300	730,124	332,300	1,075,529	369,675	0.92	0.43	0.06	<b>0.01</b>
Earnings per start (SEK)	11,981	5,463	16,673	9,765	35,346	9,897	0.86	0.44	0.16	<b>0.04</b>
Best km time (sec) <sup>4</sup>	73.1	73.9	73.3	72.8	73.1	72.9	0.85	1.00	0.19	0.10

<sup>1</sup> Log transformed values (log<sub>10</sub> + 1) were used for wins and placings. Earnings and earnings per start were transformed applying ln(earnings + 1000). Square root transformations were used for number of starts

<sup>2</sup> The C-allele is tagging the elite performance haplotype (EPH)

<sup>3</sup> A linear model was performed in R. Sex, age, birth country and number of starts were included as fixed effects (when significant). Least square means and Tukey were used as post hoc tests. Significant results ( $P \leq 0.05$ ) in bold

<sup>4</sup> Only races where the horse did not gallop were included

<https://doi.org/10.1371/journal.pgen.1011285.t004>



Table 5. Allele and genotype frequencies for SNV rs69244089 T&gt;C.

Breed	n	T	C	TT	TC	CC
Arabian horses	30	0.00	1.00	0.00	0.00	1.00
Thoroughbred	30	0.07	0.93	0.00	0.13	0.87
Standardbred	306	0.14	0.86	0.03	0.22	0.75
Warmblood	3	0.17	0.83	0.00	0.33	0.66
Coldblooded trotters	539	0.36	0.64	0.12	0.48	0.40
Finnhorses	4	0.50	0.50	0.50	0.00	0.50
Gotland pony	2	0.50	0.50	0.50	0.00	0.50
Shetland pony	18	0.64	0.36	0.44	0.39	0.17
Ardennes	20	0.78	0.22	0.65	0.25	0.10
Icelandic horses	11	0.86	0.14	0.73	0.27	0.00
North-Swedish draught	18	0.92	0.08	0.83	0.17	0.00
Exmoor pony	23	1.00	0.00	1.00	0.00	0.00
Fjord horse	1	1.00	0.00	1.00	0.00	0.00
Pzrewalski	3	1.00	0.00	1.00	0.00	0.00

<https://doi.org/10.1371/journal.pgen.1011285.t005>

been an increase of the alternate allele observed from at least 7,500 to 5,500 years ago (Fig 3). For all 14 SNVs, the two alleles were shown to segregate in specimens pre-dating the rise and spread of the DOM2 genetic lineage of modern domestic horses [35].

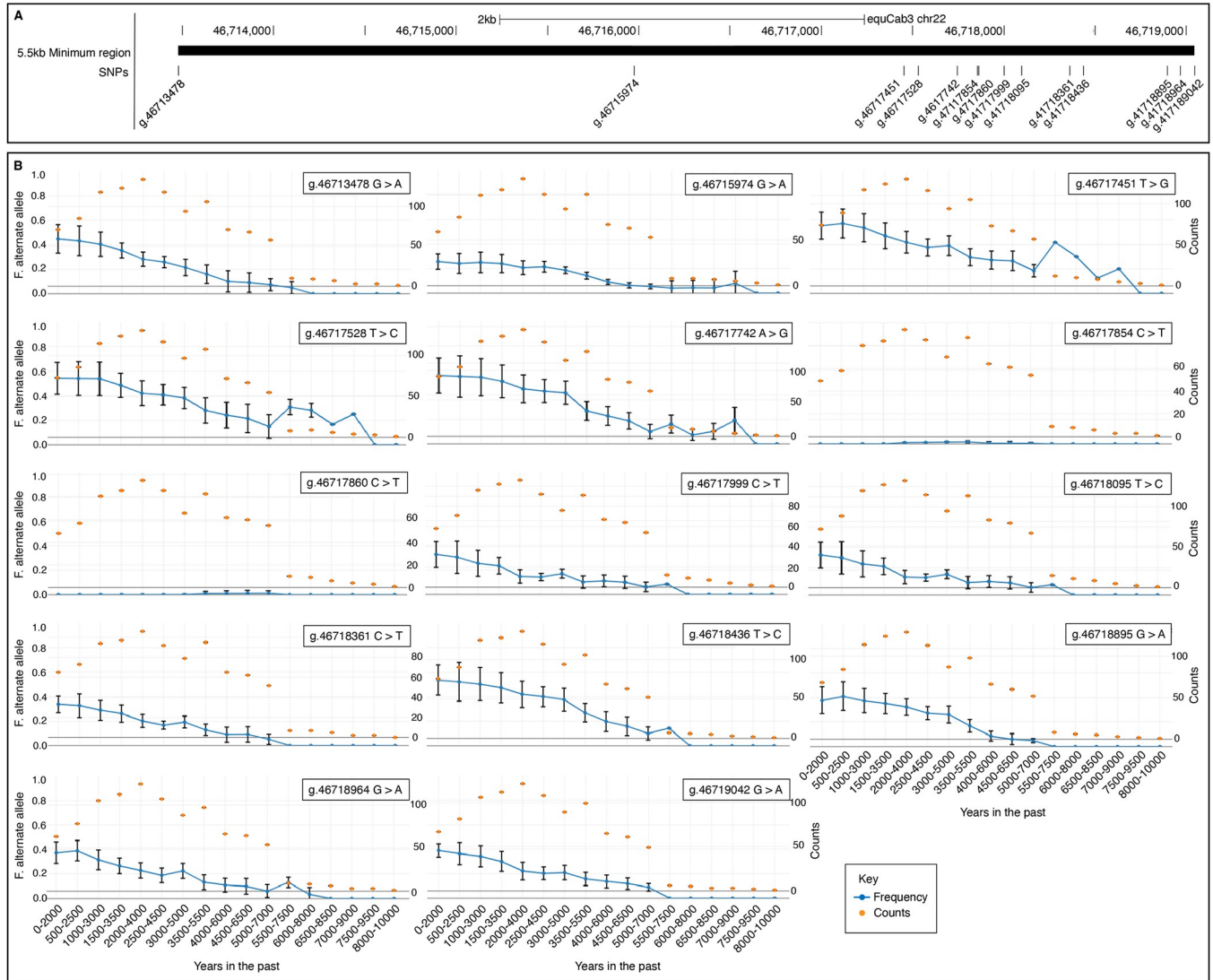
### Elite performance haplotype linked to lower blood pressure

To link genotype to phenotype, we assessed the physiological difference between CBTs homozygous for SPH (n = 5–7) and EPH (n = 17) before, during and post-exercise (see methods).

At rest, blood pressure measurements were also taken from eight heterozygous horses (HET). For all traits, a significant interaction was calculated between haplotype and time point. On average, horses homozygous (HOM) for SPH had significantly higher systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean arterial pressure (MAP) during exercise, measured directly after the last uphill interval (Fig 4A–4C). In addition, five minutes after the uphill interval, the SPH group showed significantly higher SBP, MAP, and pulse pressure (PP) than the EPH group (Fig 4D and 4E). There were no statistically significant differences in blood pressure measurements at rest before the exercise. Although there was a higher proportion of males in the SPH group, there was no significant impact of sex on any of the blood pressure measurements. All blood pressure values are presented in S9 Table.

### Plasma concentrations of EDN1 and EDN3 differ between the haplotype groups

ELISA tests were used to measure the plasma concentration of EDN1 and EDN3 at rest and during exercise for each haplotype group. Samples with a coefficient of variation (CV) value above 20% were excluded from the analysis (S10 and S11 Tables). Horses homozygous for SPH (n = 8) had a significantly higher plasma concentration of EDN1 at rest and during exercise, compared to the other groups (Tukey's HSD test) (Fig 5). In addition, SPH horses had lower plasma concentrations of EDN3 at rest (P = 0.06) and during training (P = 0.003) compared to horses homozygous for the EPH (Fig 5). Individual plasma values are presented in S10 and S11 Tables. There was a significant correlation between the plasma EDN3 and DPB (R = -0.66, P = 0.03) and a borderline significant correlation between EDN3 and MAP (R = -0.57, P = 0.07) during exercise. Also, there was a significant correlation between EDN1 and DBP



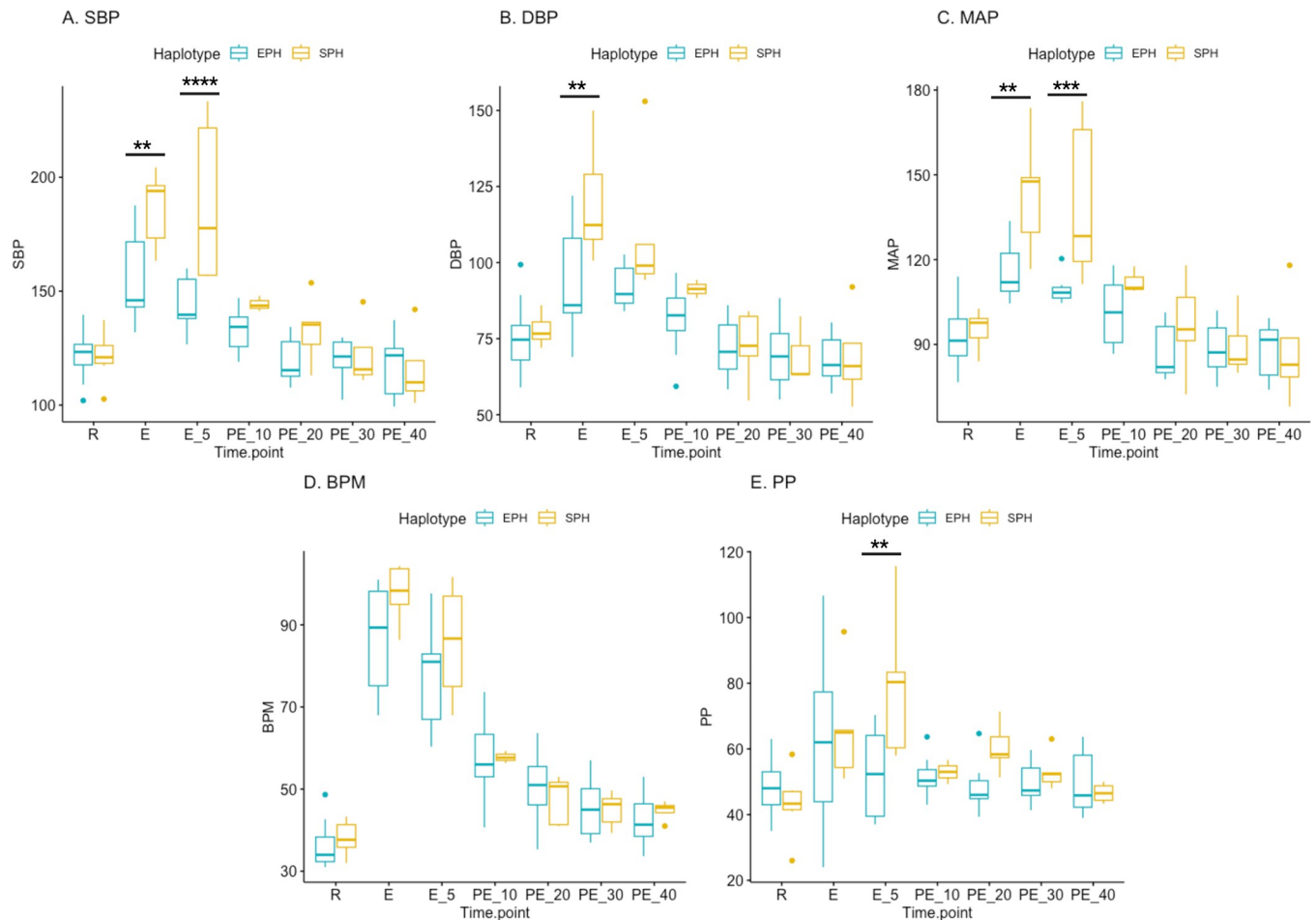
**Fig 3. Distribution of alleles over time.** A) The location of the 14 SNVs included in the minimum shared region. B) From estimated allele counts, the frequency of the alternative allele for each SNP (blue) is plotted over time in bins of 2000 years. The number of horse samples considered is also indicated (orange).

<https://doi.org/10.1371/journal.pgen.1011285.g003>

( $R = 0.73$ ,  $P = 0.01$ ), SBP ( $R = 0.7$ ,  $P = 0.02$ ) as well as MAP ( $R = 0.81$ ,  $P = 0.003$ ) during exercise.

### Discussion

Naturally occurring animal models, such as livestock and companion animal species, can provide complementing views to *de novo* generated animal models, where genome editing is required and the genomic context of the mutation may be lost. Natural model populations can carry specific genetic variants that have been under artificial selection to obtain desired phenotypes. In this study, we have used such a model to study the regulatory genomics of blood pressure modulation. More fundamental knowledge on exercise related blood pressure modulation and blood pressure tuning in general is warranted, since dysregulation of these systems are linked to both cardiovascular disease and metabolic syndrome [41,42]. By

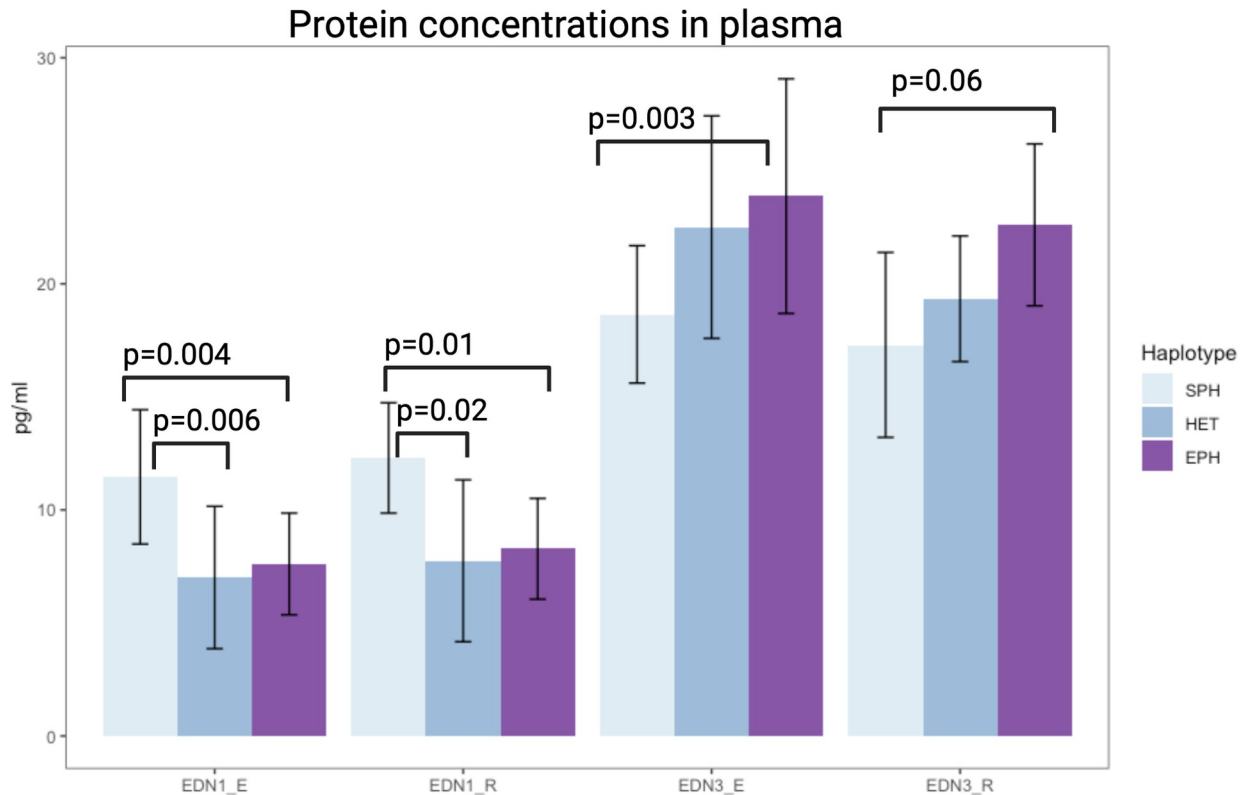


**Fig 4. Blood pressure measurements during and after exercise in horses homozygous for the sub-elite performing haplotype (SPH, n = 5–7) and horses homozygous for the elite-performing haplotype (EPH, n = 7–17).** A. SBP: Systolic blood pressure, B. DBP: Diastolic blood pressure, C. MAP: Mean arterial blood pressure, D. BPM: Beats per minute, E. PP: Pulse pressure. PP is defined as the systolic blood pressure minus the diastolic blood pressure. R = Rest, E = During exercise, directly after the uphill interval, E\_05 = During exercise, 5 minutes after the uphill interval, PE = Post exercise.

<https://doi.org/10.1371/journal.pgen.1011285.g004>

performing multiple experiments, to first refine a selective sweep region associated with racing performance, followed by comparative transcription factor binding analyses and 3D genome interaction mapping, as well as horse plasma ELISA and blood pressure measurements, we identified a regulatory element comparatively active in human endothelial cells, potentially affecting horse blood pressure regulation and athletic performance. Here, we demonstrate a regulatory role of a potentially admixed genomic region, advantageous for harness racing performance in horses, and show that this region interacts with several well-known blood pressure linked genes, as identified in human GWAS studies [21,30].

The intergenic selective sweep identified in 2018 [15] is located 50 kb downstream of the *EDN3* gene. Given both the known role of *EDN3* in blood pressure regulation and its proximal location, we hypothesized that 1) the identified non-coding region harbored a regulatory element that acted on the *EDN3* gene and 2) a key phenotype for elite athletic performance is blood pressure regulation. Our results demonstrated significant associations between the two minimal haplotypes identified in CBTs and SBs and multiple racing performance traits. In addition, in comparison with SPH homozygotes, horses homozygous for the EPH had significantly higher levels of *EDN3* in their plasma and lower blood pressure readings during exercise.



Created with BioRender.com.

**Fig 5. Plasma concentrations (pg/ml) of EDN1 and EDN3 at rest (R) and during exercise (E) in horses homozygous for the sub-elite performing haplotype (SPH) (n = 8), horses heterozygous for the EPH and SPH (HET) (n = 11) and horses homozygous for the elite-performing haplotype (EPH) (n = 21).** ANOVA and Tukey's HSD tests were performed. Standard deviations are presented as error bars.

<https://doi.org/10.1371/journal.pgen.1011285.g005>

The role of endothelins in blood pressure regulation is well studied. Endothelins are involved in one of the most potent vasoregulatory systems, where EDN1 and EDN3 have opposite effects and act synergistically to regulate blood pressure [43]. EDN1 increases blood pressure by vasoconstriction, while EDN3 is involved in nitric oxide release, in turn resulting in vasodilation and decreased blood pressure [43]. Further, EDN3 also stimulates the secretion of vasopressin, which increases blood volume by retaining water in the kidneys, while EDN1 has the opposite effect in the same organ [44–47]. Although EDN3 clearly contributes to blood pressure regulation, and was a strong candidate for our regulatory region, in the absence of appropriate horse samples, we used already produced HiCap datasets from 12 different human cell types, to investigate the comparative chromatin interaction profile of our identified region. Spanning 12 cell types covering multiple organs and life stages, including iPS cells and a neural cell line (S5 Table), the results showed that interactions were only observed between our putative regulatory region and different genes in human primary vascular endothelial cells. Endothelial cells make up the inner layer of blood vessels, and regulate the exchange between the blood and the surrounding tissues, by adjusting the vessel diameters according to the tissue demand [48]. Endothelial cells are critical regulators of the vascular tone and blood pressure regulation, as they influence the contractile ability of vascular smooth muscle cells within small arteries [49].

From the human HiCap data analysis, genes with direct interactions with our putative regulatory region included *GNAS*, *SPO11* and *ZNF831*, all with orthologues and conserved gene order in horses (Fig 1C). Also included was the human specific lincRNA *RP4-614C15.2* (Fig 2). Each of these genes encode proteins with known functional roles related to endothelial dysfunction but are certainly not limited to this function. *GNAS* is a complex locus which is imprinted in humans and mice [50,51]. It gives rise to multiple gene products through the use of both alternative promoters and splicing mechanisms [52]. This inclusion of varied first exons together provides instructions for the alpha subunit of the protein complex, called a guanine nucleotide-binding protein (G protein) [53]. This transmembrane protein complex is involved in calcium and potassium ion concentration changes within cells [54,55]. These changes are crucial for regulating heritable traits such as cardiac output and peripheral vascular resistance. In fact, blood pressure regulation in humans is a classical complex genetic trait with heritability estimates of 30–50% [56]. In line with this, GWAS of blood pressure measures has identified hundreds of genetic variants associated with systolic- and diastolic blood pressure as well as hypertension in humans [30]. A genomic region frequently identified in human GWAS for blood pressure measures is the wider *GNAS-EDN3* region [21,22]. The *SPO11* (Initiator Of Meiotic Double Stranded Breaks) gene is involved in DNA damage repair [57], and the stressors leading to DNA damage can be found during exercise.

In both humans and horses, intense exercise induces a physical stress response, in order to supply optimal energy and oxygen for the working muscles. Endothelial cells in the vascular wall have a significant role in maintaining blood flow by regulating the diameter of the blood vessels and by preventing the blood from clotting [58,59]. The blood flows more efficiently, and with less turbulence if the blood vessels are wide and if the blood is less viscous [60]. As such, blood vessels must be dilated during intense exercise, in order to supply working muscles with sufficient energy and oxygen. An increasing body of evidence suggests that oxidative stress, which results in an excessive generation of reactive oxygen species (ROS), is vital in the pathogenesis of hypertension [61]. Oxidative stress and inflammation significantly contribute to vascular remodeling by promoting exaggerated contractility and proliferation of vascular cells [62]. These factors also favor DNA damage, thereby linking *SPO11* to endothelial dysfunction. It should be noted that several genes with known function in DNA damage and response pathways previously have been associated with pulmonary arterial hypertension [63]. *ZNF831* encodes a transcription factor that has been associated, via GWAS, with a wide range of human phenotypes, including body height [64], platelet component distribution width [65], hair color [66] and diastolic blood pressure [21,67]. The *ZNF831* mRNA is highly expressed in whole blood and spleen, and the protein is identified in the heart and T-lymphocytes [68], making it a likely candidate transcription factor for endothelial cells. While not orthologous in horse, the human lincRNA *RP4-614C15.2* is a gene with a known role in angiogenesis [69]. LincRNAs are found explicitly in the nucleus, functioning in cell differentiation and identity. Imminent evidence suggests that long non-coding RNAs (lncRNAs) play critical roles in pulmonary vascular remodeling and pulmonary arterial hypertension (PAH) [70]. LncRNAs are implicated in regulating chromatin structure, thereby leading to pulmonary arterial endothelial dysfunction by modulating endothelial cell proliferation, angiogenesis, endothelial mesenchymal transition, and metabolism [71]. Targeting epigenetic regulators may lead to new, potential therapeutic possibilities in treating PAH [71]. It is clear that the search for potential lncRNAs regulating this wider genomic region in horses are required, but even so, these four directly interactive genes comprise promising candidates that warrant further investigation. The results observed from the human HiCap data were supported by the analysis of ChIP-seq data provided by the equine section of FAANG [72,73]. We demonstrated that there is an active enhancer in lamina, located in our 5.5 kb region (Fig 1). Laminae can be described as

finger-like protrusions of tissue in the hoof and is affected in equine laminitis. There is plentiful evidence that a compromised blood flow through the lamina develops in horses with laminitis e.g., [74], but little is known about the cause. One piece to this puzzle has shown that the concentration of EDN1 in laminar connective tissues obtained from laminitic horses was higher than in non-laminitic horses [75,76]. Horses with laminitis are hypertensive [77], consistent with digital hypoperfusion. All in all, these findings may be linked to the active enhancer in our 5.5 kb region in lamina.

Each SNV within the minimum shared 5.5 kb span was found to segregate in the genome of horses that lived before the rise and spread of the modern domestic lineage, around ~4,200 years ago (DOM2) [40]. Temporal allelic trajectories portray an increase in the alternative allele frequency, from 7,500–5,500 years ago, prior to the earliest archaeological evidence of horse husbandry [39,78]. This concerted rise in alternate allele frequency following the spread of the DOM2 lineage is compatible with a haplotype undergoing positive selection, possibly following artificial selection for horses showing improved athletic performance. Interestingly, the allele trajectories are in striking contrast with those previously described at the *myostatin* locus, which is driving performance in short-distance racing [8]. In that case selection was recent, indicated to start only within the last 1,000 years. It may be that performance traits represent recurrent selection targets over the history of horse domestication.

Although the current study demonstrated statistically significant differences in horse EDN1 and EDN3 levels between the different haplotype groups, there were no observed differences in blood pressure at rest. In adult horses, normal resting blood pressure is approximately 130/95 mmHg (systolic/diastolic) [79], in line with the values observed in the current study. In humans, it has been demonstrated that elevated blood pressure at rest is negatively correlated with athletic performance [80]. Individuals with elevated blood pressure had significantly lower maximal oxygen consumption, ventilatory anaerobic thresholds, and heart rate (HR) reserves (difference between an individual's resting HR and maximum HR) [80]. However, unlike humans, high blood pressure in horses is very uncommon, and most blood pressure measurements on horses are performed to evaluate and monitor hypotension. An increase in blood pressure at rest is most commonly seen as a result of diseases such as laminitis, chronic renal failure, or equine metabolic syndrome [81–83].

Perhaps the differences in observed blood pressure measurements between humans and horses is driven by species differences in the red blood cell reservoir, the spleen [60]. A previous study on horses demonstrated a significant correlation between blood pressure and spleen volume, with the splenic volume decreasing when hypertension was induced [84]. The equine spleen reservoir is larger than in any other domestic animal, and when there is increased demand, erythrocytes stored in the spleen can be released into the system [60]. Given the relationship between spleen size and blood pressure, further studies examining the relationship between the identified regulatory region, spleen size, and blood pressure represent promising avenues for enhancing the understanding of the cardiovascular system in horses in relation to exercise. While it may be viewed as a limitation, in the current study, blood pressure was measured using an external cuff. In general, direct blood pressure measurements via a fluid-filled cannula inserted into an artery are more accurate. However, the indirect blood pressure measurements are less invasive and horse cuff protocols have previously been successfully evaluated [85].

The regulatory potential of the region identified in this study is likely not limited to blood pressure regulation. For example, the EDN3 ligand has a broad expression across a range of tissues and has been implicated in diseases such as Hirschsprung's disease, Multiple Sclerosis, and Waardenburg syndrome in humans, as well as melanocyte development in mice [86–90]. Also, the *GNAS* gene encodes a G protein with multifaceted functionality. Much of vertebrate

physiology is based on the versatile functionality of G protein complexes associated with their G protein coupled receptors (GPCR) [91]. Notably, the G protein encoded by the *GNAS* gene helps stimulate the activity of an enzyme called adenylyl cyclase [92]. Adenylyl cyclase is the key enzyme that synthesizes cAMP and it is involved in cellular processes that help regulate the activity of endocrine glands such as the thyroid, pituitary gland, gonads, and adrenal glands [93,94]. In addition, adenylyl cyclase is thought to influence bone development signaling pathways, limiting the spatial production of bone with tissue [95]. Further, we see a trend associated with the haplotype and conformational type differences across horse breeds, warranting further studies. Another intriguing avenue is via thyroid hormone regulation, a network with profound effects on the cardiovascular system [96,97]. Therefore, it is unsurprising that the 15 protein-coding genes that directly, or indirectly, interacted with the putative regulatory region, showed gene ontology terms related to thyroid hormone regulation in an enrichment analysis (S7 Table). Thyroid hormone likely mediates the new demands and adaptations on the cardiovascular system needed for elite performance.

Taken together, this study underscores the pivotal role of a specific genomic region in regulating endothelial tissue integrity across species boundaries. By shedding light on the intricate mechanisms governing blood pressure modulation, these findings hold promise for advancing our understanding of cardiovascular health and disease. Moreover, the implications extend beyond blood pressure regulation, potentially encompassing a myriad of physiological processes influenced by hormone signaling. This elucidation of molecular interactions underscores the interconnectedness of mammalian physiology and highlights avenues for future research into therapeutic interventions and preventive measures against cardiovascular disorders.

## Conclusion

By combining state of the art genomics with the careful phenotyping of equine athletic performance and blood pressure measurements, we have identified a 5.5 kb potential regulatory region distal to *EDN3*, but with influence on multiple genes. The application of comparative interactome and bioinformatics datasets, allowed us to suggest that this is an enhancer region regulating the transcription of multiple blood pressure relevant genes in endothelial cells. These results are key to understanding the biological mechanisms behind blood pressure regulation in both human and horse elite athletic performance.

## Materials and methods

### Ethics statement

Horse blood sample collection was approved by the ethics committee for animal experiments in Uppsala, Sweden (number: 5.8.18-15453/2017 and 5.8.18-01654/2020). Human samples were collected following approvals for the Human Research Ethics Committee at Karolinska Institute (application number 2006/784-31/1 and 2012/1633-31/4), Solna, Sweden. Written informed consent was obtained from all the individuals according to the Declaration of Helsinki and methods were carried out following relevant guidelines.

### Horse material and DNA extraction

A summary of the various horses used across the experiments detailed below, is presented in S2 Table. Metadata and genotype information for all tested horses is included in S12 Table. Genomic DNA was extracted from hair roots or blood samples. For hair preparation, 186  $\mu$ L of 5% Chelex 100 Resin (Bio-Rad Laboratories, Hercules, CA, USA) and 14  $\mu$ L of proteinase K (20 mg/mL; Merck KgaA, Darmstadt, Germany) were added to each sample. The mixture was

incubated at 56°C for two hours at 600 rpm and inactivated with proteinase K for 10 mins at 95°C. Blood samples were extracted on the Qiasymphony instrument with the Qiasymphony DSP DNA mini or midi kit (Qiagen, Hilden, Germany).

### Racing performance association analysis with known selective sweep

All horse genomic coordinates in the study refer to the EquCab3.0 assembly [98]. In 2018, a 19.6 kb selective sweep (ECA22:46,702,297–46,721,892; EquCab3.0) was identified in CBTs [15]. Subsequently, the racing results and Illumina SNP670K Genotyping BeadChip genotypes from 378 CBTs were used to identify five SNVs ( $r^2 > 0.9$  sweep region) as significantly associated with elite racing performance in that breed [15]. In this study, we repeat this sweep-performance analysis using horse metadata and 670K Axiom Equine Genotyping Array genotypes available from 629 CBTs [99]. The 629 CBTs include the 378 horses used in the 2018 study [15]. To define the region for analysis from within the sweep, we centered our analysis on the most significant variant from the 2018 study: rs69244086-T, and used available 670 K Axiom Equine Genotyping Array genotypes to calculate pairwise  $r^2$  across the original 19.6 kb region. Variants with  $r^2 > 0.6$  were included in the association analysis. Given the high LD, tests were not considered independent, and so the significance threshold was set at  $P \leq 0.05$ . The available metadata included a) pedigree information, b) *DMRT3* gait associated genotypes [12] and c) performance data as individual race records for each horse (Swedish Trotting Association (Svensk Travsport) and the Norwegian Trotting Association (Norsk Rikstoto)). The *DMRT3* variant rs1150690013 (ECA23:22391254) was genotyped using StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) with the custom TaqMan SNP genotyping assay [100]. Reactions (15  $\mu$ l) contained 1.5  $\mu$ l DNA, 0.38  $\mu$ l Genotyping Assay 40X, 7.5  $\mu$ l Genotyping Master Mix 2X, and 5.62  $\mu$ l deionized water. The *DMRT3* variant was included as a factor in the association model, as previous studies have demonstrated major effects of the variant on harness racing performance [12,14]. Only performance data from competitive races were included in the analysis (i.e., premie and qualification races were excluded). The performance data included: number of starts, number of wins, number of placings (the number of times a horse finished a race in first, second or third place), fastest kilometer (km) time in a race where the horse did not gallop (in seconds), earnings and earnings per start. The earnings for most horses were recorded in Swedish currency (SEK). Earned prize money in Norwegian currency was converted to SEK based on the average exchange rate for the specific race year.

All statistical analyses were performed in the R statistical environment versions 4.1.1 and 4.2.2 [23]. The performance data were tested for normality by computing the skewness coefficient using the package moments v0.14. Non-normally distributed values were transformed, i.e., log transformed values ( $\log_{10} + 1$ ) were used for wins and placings, number of starts were square root transformed and earnings and earnings per start were transformed according to the previously reported formula ( $\ln(\text{earnings} + 1000)$ ) [101]. Genetic variants were tested in a linear model using ANOVA as a *post hoc* test. Number of starts, age at first start, sex, birth year, country of birth, and the *DMRT3* genotypes were included as fixed effects, when significant. Significant values ( $P \leq 0.05$ ) were further tested using the function lsmeans (Least-Squares Means) with the package emmeans followed by the multiple comparison test Tukey's HSD-test.

### SNV, indel and SV variant discovery

To find variants not included on the 670K Axiom Equine Genotyping Array, but with the potential to influence racing performance we performed Illumina short-read WGS. Two CBTs and two SBs were selected based on their rs69244086 genotype and their performance



earnings. For each breed, one horse with high earnings per start and homozygous T at SNV rs69244086, and one horse with low earnings per start homozygous C at SNV rs69244086, were selected for sequencing (S2 Table). Paired-end 150 bp Illumina HiSeqX data was generated to a depth of 15X. Read data were processed according to the GATK 3.8–0 best practices (<https://software.broadinstitute.org>). Briefly, raw data were trimmed with TrimGalore 0.4.4 and quality checked with QualiMap 2.2. The reads were aligned to the EquCab3.0 reference genome (98) using BWA-MEM 0.7.17. SAM files were converted to BAM files using SAMtools 1.8, and Picard 2.10.3 was used to sort the BAM files and remove potential PCR duplications. The BAM files were recalibrated (BaseRecalibrator and PrintReads) and small variants were called with HaplotypeCaller using a list of known variants for calibrations (NCBI release 103, <https://www.ncbi.nlm.nih.gov>). Structural variants were called using default parameters in FindSV v0.4.0 combining the programs Manta, TIDDIT and CNVnator (<https://github.com/J35P312/FindSV>).

### Variant filtration and genotyping

In total, 84 variants, including 78 SNVs and 6 indels of various sizes, were discovered from the WGS data. From the SVs, only a 400 bp deletion (ECA22:46,714,602–46,715,003) matched the required segregation pattern of present in CBTs and SBs, reflective of a selection on crossbreed admixture [15]. The 400 bp deletion was subsequently genotyped in a further 497 CBTs with performance data and available blood sample. The genotyping was performed with digital droplet PCR (ddPCR) using the ddPCR Supermix for Probes (No dUTP) (Bio-rad) protocol. An amplicon of 70 bp (primer sequences; fwd: CTGAGCACAGGGCAGTGT, rev: GAGC GGACAAGAGCGATTG and probe FAM–CCGGGAAAACAGCCCCCTTCC) was used to detect the deletion. A 97 bp amplicon of the Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) gene was used as a control (fwd CGATGCTGGTGCTGAATATGTT, rev: GGTCAACTCCCCTCATCTTTAGC and probe sequence TCTTCACTACCTTGAGAGAAG with HEX fluorophore). Genomic DNA restriction enzyme digestion of DNA was performed in the ddPCR reaction using 1 ul of FastDigest TRUII EA restriction enzyme (Thermo Fisher Scientific). Droplets were generated with Automated Droplet Generator (Bio-Rad Laboratories Inc). The PCR reaction volume was 20 ul and the PCR program included enzyme activation at 95°C for 10 min, 40 cycles of denaturation at 94°C for 30 seconds, and annealing at 60°C for 1 minute, followed by enzyme deactivation at 98°C for 10 minutes. The samples were run in the QX100 or QX200 Droplet reader, and the data were analyzed in the QuantaSoft Software (Bio-Rad Laboratories Inc). Thresholds were set manually in the 2D amplitude tab.

For the smaller variants, technical limitations restricted us to 25 multiplexed SNVs (including indels) (MassARRAY Assay Design Suite v2.2 software, Agena Bioscience, San Diego, CA, USA). As such, discovery variants were filtered for inclusion based on i) genotype segregation patterns from a previous pooled genotype experiment [16] and then ii) for even spacing across the 19.6 kb sweep region. The pooled experiment included the three Nordic horse breeds represented in the original 2018 selective sweep analysis, CBTs, SBs and the NSD [15]. Following the pattern of suspected introgression of SB into CBT, WGS discovery variants were retained if their genotyping pattern in the pooled data was similar between the CBT and SB pools, and discordant in the NSD pool. After filtering, 24 SNVs and one INDEL were selected for MassArray genotyping.

To assess breed or species specificity, the 25 multiplexed variants were genotyped across 394 samples (S13 Table). This included 210 CBTs, plus 181 samples from 11 other breeds, as well as 3 Przewalski horses. The representative samples were selected to include both traditional performance horse breeds (e.g., Thoroughbreds), and non-performing breeds (e.g.,

Shetland pony) (S13 Table). The genotyping was performed using iPLEX Gold chemistry and the MassARRAY mass spectrometry system [102] (Agena Bioscience) at the Mutation Analysis Facility at Karolinska University Hospital (Huddinge, Sweden), according to Agena Bioscience's recommendations. In brief, analytes were spotted onto a 384-element SpectroCHIP II array (Agena Bioscience), using Nanodispenser RS1000 and subsequently analyzed by MALDI-TOF on a MassARRAY Analyzer 4 mass spectrometer (Agena Bioscience). Genotype calls were manually curated based on concordance calls from two operators using MassARRAY TYPER v4.0 Software (Agena Bioscience). Due to design limitations, the known performance associated SNV rs69244086 [15] was not included as part of the MassArray. Rather a custom TaqMan SNV assay (100) was designed, and assayed using the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). Here, due to DNA availability, a reduced set of 242 samples was genotyped (183 CBTs, 38 SBs, 5 NSD, 2 Ardennes, 4 Finnhorses, 1 Fjord horse, 2 Gotlandsruss, 2 Icelandic horses, 2 shetland ponies and 3 Pzrewalski).

### Association analyses with discovery variants

For the SV association analysis, a linear model was developed as per above section, "Racing performance association analysis with known selective sweep", including horse metadata and the genotypes from the 400 bp deletion for 497 CBTs and 41 SBs. A haplotype block for association analysis was then defined as all SNVs from the MassArray that were in pairwise LD  $r^2 > 0.95$  with rs69244086 (LD-function in R). Prior to association analysis, maximum likelihood estimates were used to determine the 14 SNVs haplotypes with the haplo.em function in the haplo.stat package [103]. In total, 166 CBTs and 38 SBs were available for analysis. Haplotype racing performance association tests used a GLM regression analysis using the haplo.glm function from the haplo.stats package in R [103]. The model included the effects of sex, age, country of registration, *DMRT3* genotype (only for the CBTs) and number of starts, when significant. All CBTs were genotyped for the *DMRT3* mutation, using a custom TaqMan SNP genotyping assay [100] as described in the section "Racing performance association analysis with known selective sweep". CBTs not typed for the *DMRT3* genotype were excluded from the analysis.

Haplotypes with a frequency  $< 2\%$  were considered rare, and therefore not analyzed for association with performance. A minimum haplotype span was based on haplotype sharing between CBTs and SBs, and was used to define the elite-performing haplotype (EPH) and the sub-elite performing haplotype (SPH). As the core of the EPH and SPH haplotypes are defined by extremely high pairwise LD ( $r^2 > 0.95$ ) alleles, any of the variants within this span can be used as proxy for the haplotypes themselves. We therefore extended the number of horses we could include in the association analysis by genotyping 519 CBTs and 271 SBs for rs69244089 using a custom TaqMan SNP genotyping assay [100], and methods described in the section "Racing performance association analysis with known selective sweep". Three CBTs lacked *DMRT3* genotypes, and were excluded from the analysis. In total, 516 CBTs born between 1994 and 2017, and 271 SBs born between 2005 and 2014 were then available for association with racing performance. A linear model was performed in R. Sex, age, birth country, number of starts, and *DMRT3* genotype were used as fixed effects. Least square means and Tukey were used as post hoc tests.

### "Performance" SNV frequency in other populations

As per the rs69244086 SNV in section, "Variant filtration and genotyping", horses from 13 different breeds as well as Pzrewalski horses, were also genotyped using the TaqMan assay for rs69244089, in order to investigate the distribution of this variant. The Hardy-Weinberg

equilibrium test (hwe) from the gap package in R, was used to test if the genotypes observed deviated from HWE in SB and CBT.

### Comparative genomics to infer haplotype function

In the absence of empirical data from genotyped horses, we used comparative data from horses and human sources. For horses, ChIP-seq data were provided by the equine section of FAANG [72,73]. These samples from project PRJEB35307 were generated from two Thoroughbreds and included four histone marks (H3K4me1, H3K4me3, H3K27ac and H3K27me) across nine tissues (adipose, brain, heart, lamina, liver, lung, muscle, ovary and skin). After local download, data integrity was confirmed with checksums. Bam files were sorted with samtools and bedgraphs were created with bedtools [104,105]. Bigwig files were viewed as custom tracks in UCSC, EquCab3.0.

To assess the per SNV potential intersection from the minimum haplotype and surrounding genes, we had access to twelve different human HiCap [106] datasets produced at the Sahlén Laboratory (S5 Table). HiCap combines Hi-C and targeted sequencing, to obtain high-resolution genome interaction maps [19]. In short, live cells are crosslinked using formaldehyde, to immobilize chromatin. The cells are lysed and chromatin is digested using a frequent cutter restriction enzyme (such as 4-cutter *DpnII*). The digested chromatin is then end-repaired using biotinylated nucleotides, and ligated back to itself, to capture the proximity information (ends that are close to each other in 3D nuclear space are more likely to be ligated than those that are far away). Once the ligation is complete, the crosslinking is reversed using heat, proteins are removed and DNA is purified, completing the Hi-C step [107]. A sequencing library is then prepared using the Hi-C material. The Hi-C library is hybridized to a set of probes targeting around 21,000 promoters, to obtain only the proximities of promoters, providing up-to 110-fold increase in promoter interaction resolution [106]. Custom designed probes were purchased from Agilent Technologies Inc (Santa Barbara, CA, USA) including the reagents for sequencing library preparations. The Hi-C material was captured and sequencing libraries were prepared using HS2 XT Library Prep Kit (Agilent Technologies Inc, article number: 5190–9685, 5500–0146 and 5191–6686) with small modifications [106]. In several steps, the final libraries were sequenced, on average, ~640 million paired-end for each experiment, at the National Genome Infrastructure Sweden (paired end 2x150 bases). The reads were mapped to hg38 assembly using Bowtie2 (version 2.2.9) using the -very-sensitive option [108]. HiCaptools was used to call interactions in all samples using default settings. GENCODE (version 37) comprehensive gene annotation using only reference chromosomes [109] was used to annotate promoters. We required at least four supporting pairs and true P-value less than 0.05 for each interaction. We overlapped all promoter-interacting regions with the variants around the EDN3 gene using bedtools intersect function via the bedr package v1.0.7 [110] implemented in R (version 4.2.1) [23] with no padding. We searched the twelve HiCap datasets for the interactions of the minimal haplotype block lifted to either hg19 (chr20:57950706–57964856) or hg38 (chr20:59375651–59389801) depending on the dataset. We used interaction sets called at FDR 0.1 thresholds except the vascular endothelial cells [111]. No interactions were present in any of the cell types apart from the vascular endothelial cells. The primary vascular endothelial cells were collected from 16 patients undergoing elective open-heart surgery at the Cardiothoracic Surgery Unit, Karolinska University Hospital, Stockholm, Sweden. HiCap was performed on 2.5 million cells derived from each patient according to the latest protocol [111]. All interactions that were statistically significant in each patient ( $P < 0.05$ ) were called. There were, in total, 49 interactions overlapping with variants. For comparative plotting on the horse reference genome, interactions that were present in at

least two individuals were considered. A network generation figure was generated using the Flourish visualization app (<https://flourish.studio>). To reduce complexity of the network figure, only interactions that were present in at least five individuals were included. For all genes with significant interactions with our region of interest, we performed a gene enrichment analysis using the ToppGene Suite [112].

We could not further fine map the minimum haplotype to identify candidate causative SNVs in the region. However, the impact on transcription factor binding was assessed with MotifbreakR [34] in R v4.2.2 and for 14 SNVs in the minimum shared 5.5 kb locus. Using `forBSgenomeDataPkg` (included in `motifbreakR`) and the UCSC `equCab3.2bit` file (<http://hgdownload.soe.ucsc.edu/goldenPath/equCab3/bigZips/>), we created the genome library required for `motifbreakR` as it is not supplied with this package. We assessed the impact of allele difference between the EPH and SPH haplotypes with `hocomoco` motifs [33], default position weight matrix to score each, at threshold of  $1 \times 10^{-4}$ , before retaining the strong category of effects.

### Ancient DNA screening

We leveraged the availability of an extensive ancient genome time-series in the horse to assess the temporal trajectory and spatial distribution of the SNVs in the minimum shared haplotype within the CBTs, in the past. More specifically, we re-aligned against the `EquCab3.0` reference genome (98) supplemented with the Y-chromosomal contigs from [113] and a sub-selection of the sequencing data from [4,36,38,39,114,115], representing a total of 431 ancient genomes. Alignment files were generated using the `Paleomix` pipeline [116], and further rescaled and trimmed, following the procedures described in [40]. This procedure ensured minimal sequencing error rates, especially at sites potentially affected by post-mortem DNA damage. The number of occurrences of each individual allele was counted for each individual using `ANGSD` (v0.933), with the `-baq 0`, `-remove_bads` `-minMapQ 25` `-minQ 30` `-rmTriallelic 1e-4` `-SNP_pval 1` `-C 50` parameters [117]. The resulting occurrences were tabulated together with metadata available for each individual, consisting of their corresponding archaeological site (name and GPS coordinates) and their average calibrated radiocarbon date (or age as assessed from the archaeological context otherwise). The tabulated file was visualized using the `map-DATAge` package [35], which provides spatial distributions within user-defined time periods and/or spatial ranges and estimates allelic trajectories (i.e., allelic frequencies through time) based on the resampling of individual allelic counts. The final allelic trajectories were assessed considering time bins of 2,000 years (step-size = 500). To explore the genetic variation for the SNVs within the haplotype, we considered the 14 polymorphic sites i.e., rs395117226, rs397265747, rs396474304, rs69244081, rs69244084, rs69244085, rs69244086, rs69244088, rs69244089, rs396281591, rs394573286, rs69244091, rs69244093 and rs69244095.

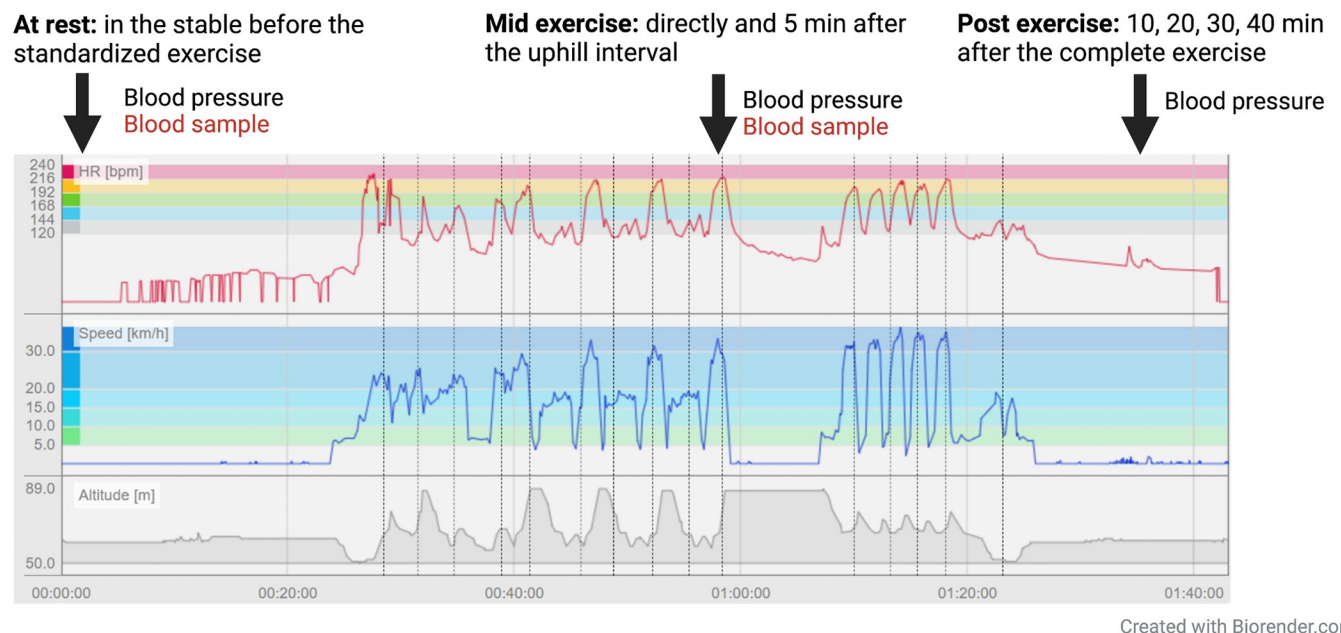
### Exercise test, sample collection and blood pressure measurements

**Horse material.** Thirty horses were enrolled in the exercise test (S2 Table). All horses were genotyped for SNVs rs69244086 and rs69244089 using the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) with the custom TaqMan SNP genotyping assay as above [100]. The two SNVs 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. 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 (S2 Table). For the ELISA, 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 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 minutes with a distance of 1.7 km. The HR for each horse reached above 200 beats/minute. After the uphill intervals, the horses walked down to the flat part of the track, where they completed 10 minutes of intervals, at a speed of 35 km/h and HR of about 180–200 beats/minute. 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/minute for approximately eight mins. Due to their young age the two-year old horses ( $n = 2$ ) performed a standardized exercise session at the racetrack. Following warm-up, they performed a number of intervals with HR reaching over 200 beats/minute. The two-year old horses performed the exercise test at the same time.

**Blood pressure measurements.** The following measurements were collected with a Cardell Veterinary Monitor 9402 cuff (Cardell 10), at the middle coccygeal artery: systolic, diastolic, mean arterial pressure, pulse pressure, and HR. The same cuff size was used for all the horses. The monitor has previously been successfully evaluated for use in horses [85]. Blood pressure was measured at rest in the stable before the exercise and during exercise (directly and five minutes after the last uphill interval) (Fig 6). Additionally, blood pressure was



**Fig 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.

<https://doi.org/10.1371/journal.pgen.1011285.g006>

measured every ten minutes after exercise, up to 40 minutes after the exercise was completed. In order for the horse to acclimate to the equipment, the Polar M460 sensor and the cuff were put on at rest, before the exercise route, and the horse was left in the box for 10 minutes before measurements were taken. Similarly, blood pressure was measured five times at rest for the horse to get used to the measurement procedure. The measurements with the highest and lowest systolic blood pressure values were removed. During and after exercise the blood pressure was measured three times and an average value was calculated. For the association analyses with blood pressure, the mean values of the triplicates were analyzed using Linear models and ANOVA, with pairwise tests using emmeans, corrected for multiple testing with Bonferroni. We also analyzed the correlation between the different blood pressure measurements and plasma EDN1 and EDN3 with ggplot and Pearson Correlation coefficient in R.

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

## Supporting information

**S1 Table. Performance analysis results for the five significant SNPs in Coldblooded trotters.**

(XLSX)

**S2 Table. Information about all horses sampled and used for the various analyses.**

(XLSX)

**S3 Table. Location of discovery and genotyped variants with the minimum 5.5 kb region.**

(XLSX)

**S4 Table. Performance results for the 400 bp deletion in 497 Coldblooded trotters.**

(XLSX)

**S5 Table. The HiCap interactions of the putative regulatory element in different human cell types.**

(XLSX)

**S6 Table. The protein-coding genes that directly or indirectly interact with the regulatory region.**

(DOCX)

**S7 Table. Enrichment of 15 protein-coding genes directly or indirectly interacting with the putative regulatory element.**

(XLSX)

**S8 Table. List of all analyzed variable sites in the 5.5 kb region.**

(XLSX)

**S9 Table. Mean values  $\pm$  SD for all blood pressure measurements.**

(DOCX)

**S10 Table. Plasma concentration of Endothelin 1 (EDN1) (mean + standard deviation) at rest and during exercise.**

(DOCX)

**S11 Table. Plasma concentration of Endothelin 3 (EDN3) (mean + standard deviation) at rest and during exercise.**

(DOCX)

**S12 Table. Metadata and genotypes for horses included in the study.**

(XLSX)

**S13 Table. List of horses included in the MassArray genotyping analysis.**

(DOCX)

**S1 Fig. Illustration of the five most significant high impact transcription factor binding changes that may result from a switch of alleles contained within the reference SPH to the non-reference EPH. SNVs rs69244086 C>T (A) and rs69244089 T>C (G/A) (B) are illustrated.**

(TIF)

## Acknowledgments

Thanks to all horse owners and trainers who participated in the study, The Swedish Trotting Association and the Norwegian Trotting Association for providing pedigree and performance data, Tytti Vanhala at the Department of Animal Breeding and Genetics, SLU, for help with genotyping, Anna Johansson at the National Bioinformatics Infrastructure Sweden at SciLife-Lab for bioinformatics advice, and the Department of Clinical Sciences Clinical training center (KTC) at SLU for providing the Cardell Veterinary monitor for the blood pressure measurements. We also thank Mats Pettersson for statistical advice and Anna Svensson's laboratory at the Department of Clinical Sciences SLU for running ELISA. WGS was performed by the SNP&SEQ Technology Platform, Uppsala University. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The computations and data handling were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) in Uppsala partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973.

## Author Contributions

**Conceptualization:** Göran Andersson, Gabriella Lindgren.

**Data curation:** Maria K. Rosengren, Rakan Naboulsi, Ludovic Orlando, Magnus Åbrink, Pelin Sahlén, Jennifer R. S. Meadows.

**Formal analysis:** Kim Fegraeus, Maria K. Rosengren, Karin Lång, Artemy Zhigulev, Hanna M. Björck, Anders Franco-Cereceda, Per Eriksson, Pelin Sahlén, Jennifer R. S. Meadows.

**Funding acquisition:** Amanda Raine, Gabriella Lindgren.

**Investigation:** Kim Fegraeus, Maria K. Rosengren, Ahmad Jouni, Jennifer R. S. Meadows.

**Project administration:** Gabriella Lindgren.

**Resources:** Maria K. Rosengren, Ludovic Orlando, Karin Lång, Artemy Zhigulev, Hanna M. Björck, Anders Franco-Cereceda, Per Eriksson, Pelin Sahlén.

**Software:** Ludovic Orlando.

**Supervision:** Rakan Naboulsi, Brandon D. Velie, Amanda Raine, Beate Egner, C Mikael Mattsson, Göran Andersson, Jennifer R. S. Meadows, Gabriella Lindgren.

**Validation:** Maria K. Rosengren.

**Visualization:** Kim Fegraeus, Pelin Sahlén, Jennifer R. S. Meadows.

**Writing – original draft:** Kim Fegraeus, Jennifer R. S. Meadows, Gabriella Lindgren.

**Writing – review & editing:** Kim Fegraeus, Maria K. Rosengren, Rakan Naboulsi, Ludovic Orlando, Magnus Åbrink, Ahmad Jouni, Brandon D. Velie, Amanda Raine, Beate Egner, C Mikael Mattsson, Karin Lång, Artemy Zhigulev, Hanna M. Björck, Anders Franco-Cereceda, Per Eriksson, Göran Andersson, Pelin Sahlén, Jennifer R. S. Meadows, Gabriella Lindgren.

## References

1. Andersson L. Domestic animals as models for biomedical research. *Ups J Med Sci.* 2016; 121(1):1–11. <https://doi.org/10.3109/03009734.2015.1091522> PMID: 26479863
2. Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, Li X, et al. Evolutionary constraint and innovation across hundreds of placental mammals. *Science.* 2023 Apr 28; 380(6643):eabn3943. <https://doi.org/10.1126/science.abn3943> PMID: 37104599
3. Sullivan PF, Meadows JRS, Gazal S, Phan BN, Li X, Genereux DP, et al. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science.* 2023 Apr 28; 380(6643):eabn2937. <https://doi.org/10.1126/science.abn2937> PMID: 37104612
4. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell.* 2019 May; 177(6):1419–35. <https://doi.org/10.1016/j.cell.2019.03.049> PMID: 31056281
5. McGivney BA, Han H, Corduff LR, Katz LM, Tozaki T, MacHugh DE, et al. Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci Rep.* 2020 Jan 16; 10(1):466. <https://doi.org/10.1038/s41598-019-57389-5> PMID: 31949252
6. 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. <https://doi.org/10.1371/journal.pgen.1003211> PMID: 23349635
7. Meadows JRS, Lindblad-Toh K. Dissecting evolution and disease using comparative vertebrate genomics. *Nat Rev Genet.* 2017 Oct; 18(10):624–36. <https://doi.org/10.1038/nrg.2017.51> PMID: 28736437
8. 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. <https://doi.org/10.1371/journal.pone.0008645> PMID: 20098749
9. 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. <https://doi.org/10.1292/jvms.11-0295> PMID: 21836385
10. 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. <https://doi.org/10.1111/j.1365-2052.2012.02329.x> PMID: 22497477



11. 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. <https://doi.org/10.1093/jhered/esw031> PMID: 27208149
12. 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. <https://doi.org/10.1038/nature11399> PMID: 22932389
13. 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. <https://doi.org/10.1093/jhered/esv062> PMID: 26285915
14. 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. <https://doi.org/10.2527/jas.2014-7803> PMID: 25085403
15. 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. <https://doi.org/10.14814/phy2.13700> PMID: 29845762
16. Velie BD, Lillie M, Fegraeus KJ, Rosengren MK, Solé M, Wiklund M, et al. Exploring the genetics of trotting racing ability in horses using a unique Nordic horse model. *BMC Genomics*. 2019 Dec; 20(1):104. <https://doi.org/10.1186/s12864-019-5484-9> PMID: 30717660
17. Panigrahi AO'Malley BW. Mechanisms of enhancer action: the known and the unknown. *Genome Biol*. 2021 Dec; 22(1):108.
18. The ENCODE Project Consortium, Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020 Jul 30; 583(7818):699–710. <https://doi.org/10.1038/s41586-020-2493-4> PMID: 32728249
19. Sahlén P, Abdullayev I, Ramsköld D, Matskova L, Rilakovic N, Lötstedt B, et al. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol*. 2015 Dec; 16(1):156. <https://doi.org/10.1186/s13059-015-0727-9> PMID: 26313521
20. Åkerborg Ö, Spalinskas R, Pradhananga S, Anil A, Höjer P, Poujade FA, et al. High-Resolution Regulatory Maps Connect Vascular Risk Variants to Disease-Related Pathways. *Circ Genomic Precis Med* [Internet]. 2019 Mar [cited 2023 Jun 14];12(3). Available from: <https://www.ahajournals.org/doi/10.1161/CIRCGEN.118.002353> PMID: 30786239
21. The International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011 Oct; 478(7367):103–9. <https://doi.org/10.1038/nature10405> PMID: 21909115
22. Turner ST, Boerwinkle E, O'Connell JR, Bailey KR, Gong Y, Chapman AB, et al. Genomic association analysis of common variants influencing antihypertensive response to hydrochlorothiazide. *Hypertens Dallas Tex* 1979. 2013 Aug; 62(2):391–7. <https://doi.org/10.1161/HYPERTENSIONAHA.111.00436> PMID: 23753411
23. R. Foundation for Statistical Computing, Vienna, Austria. R: The R Project for Statistical Computing [Internet]. [cited 2023 May 15]. Available from: <https://www.r-project.org/>
24. Giuffra E, Tuggle CK, FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu Rev Anim Biosci*. 2019 Feb 15; 7(1):65–88. <https://doi.org/10.1146/annurev-animal-020518-114913> PMID: 30427726
25. Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJM, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res*. 2016 Jan 4; 44(D1):D126–132. <https://doi.org/10.1093/nar/gkv1203> PMID: 26578589
26. The GTEx Consortium, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020 Sep 11; 369(6509):1318–30. <https://doi.org/10.1126/science.aaz1776> PMID: 32913098
27. Itoh A, Uchiyama A, Taniguchi S, Sagara J. Phactr3/Scapinin, a Member of Protein Phosphatase 1 and Actin Regulator (Phactr) Family, Interacts with the Plasma Membrane via Basic and Hydrophobic Residues in the N-Terminus. Permyakov EA, editor. *PLoS ONE*. 2014 Nov 18; 9(11):e113289. <https://doi.org/10.1371/journal.pone.0113289> PMID: 25405772
28. Wang Y, Chen S, Jiang Q, Deng J, Cheng F, Lin Y, et al. TFAP2C facilitates somatic cell reprogramming by inhibiting c-Myc-dependent apoptosis and promoting mesenchymal-to-epithelial transition. *Cell Death Dis*. 2020 Jun 25; 11(6):482. <https://doi.org/10.1038/s41419-020-2684-9> PMID: 32587258
29. German CA, Sinsheimer JS, Klimentidis YC, Zhou H, Zhou JJ. Ordered multinomial regression for genetic association analysis of ordinal phenotypes at Biobank scale. *Genet Epidemiol*. 2020 Apr; 44(3):248–60. <https://doi.org/10.1002/gepi.22276> PMID: 31879980

30. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet.* 2009 Jun; 41(6):677–87. <https://doi.org/10.1038/ng.384> PMID: 19430479
31. CHARGE-Heart Failure Consortium EchoGen Consortium, Consortium METASTROKE, Consortium GIANT, Consortium EPIC-InterAct, Lifelines Cohort Study, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet.* 2016 Oct; 48(10):1151–61. <https://doi.org/10.1038/ng.3654> PMID: 27618447
32. CHD Exome+ Consortium, ExomeBP Consortium, GoT2DGenes Consortium, T2D-GENES Consortium, Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia, CKDGen Consortium, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet.* 2016 Oct; 48(10):1162–70. <https://doi.org/10.1038/ng.3660> PMID: 27618448
33. Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, et al. HOCO-MOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 2013 Jan 1; 41(D1):D195–202. <https://doi.org/10.1093/nar/gks1089> PMID: 23175603
34. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics.* 2015 Dec 1; 31(23):3847–9. <https://doi.org/10.1093/bioinformatics/btv470> PMID: 26272984
35. Liu X, Orlando L. mapDATAge: a ShinyR package to chart ancient DNA data through space and time. Wren J, editor. *Bioinformatics.* 2022 Aug 10; 38(16):3992–4.
36. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci.* 2014 Dec 30; 111(52):5661–9. <https://doi.org/10.1073/pnas.1416991111> PMID: 25512547
37. Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, Albrechtsen A, et al. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci.* 2015 Dec 15; 112(50):6889–97. <https://doi.org/10.1073/pnas.1513696112> PMID: 26598656
38. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al. Ancient genomic changes associated with domestication of the horse. *Science.* 2017 Apr 28; 356(6336):442–5. <https://doi.org/10.1126/science.aam5298> PMID: 28450643
39. Gaunitz C, Fages A, Hanghøj K, Albrechtsen A, Khan N, Schubert M, et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science.* 2018 Apr 6; 360(6384):111–4. <https://doi.org/10.1126/science.aao3297> PMID: 29472442
40. Librado P, Khan N, Fages A, Kusliy MA, Suchan T, Tonasso-Calvière L, et al. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature.* 2021 Oct; 598(7882):634–40. <https://doi.org/10.1038/s41586-021-04018-9> PMID: 34671162
41. Stevens SL, Wood S, Koshiaris C, Law K, Glasziou P, Stevens RJ, et al. Blood pressure variability and cardiovascular disease: systematic review and meta-analysis. *BMJ.* 2016 Aug 9; i4098. <https://doi.org/10.1136/bmj.i4098> PMID: 27511067
42. Petrie JR, Guzik TJ, Touyz RM. Diabetes, Hypertension, and Cardiovascular Disease: Clinical Insights and Vascular Mechanisms. *Can J Cardiol.* 2018 May; 34(5):575–84. <https://doi.org/10.1016/j.cjca.2017.12.005> PMID: 29459239
43. Rossi GP, Sacchetto A, Cesari M, Pessina AC. Interactions between endothelin-1 and the renin-angiotensin-aldosterone system. *Cardiovasc Res.* 1999 Aug 1; 43(2):300–7. [https://doi.org/10.1016/s0008-6363\(99\)00110-8](https://doi.org/10.1016/s0008-6363(99)00110-8) PMID: 10536660
44. Emori T, Hirata Y, Ohta K, Kanno K, Eguchi S, Imai T, et al. Cellular mechanism of endothelin-1 release by angiotensin and vasopressin. *Hypertension.* 1991 Aug; 18(2):165–70. <https://doi.org/10.1161/01.hyp.18.2.165> PMID: 1909304
45. Imai T, Hirata Y, Emori T, Yanagisawa M, Masaki T, Marumo F. Induction of endothelin-1 gene by angiotensin and vasopressin in endothelial cells. *Hypertension.* 1992 Jun; 19(6\_pt\_2):753–7. <https://doi.org/10.1161/01.hyp.19.6.753> PMID: 1592477
46. Rossi NF. Effect of endothelin-3 on vasopressin release in vitro and water excretion in vivo in Long-Evans rats. *J Physiol.* 1993 Feb; 461:501–11. <https://doi.org/10.1113/jphysiol.1993.sp019525> PMID: 8350273
47. Rossi NF. Cation channel mechanisms in ET-3-induced vasopressin secretion by rat hypothalamo-neurohypophysial explants. *Am J Physiol-Endocrinol Metab.* 1995 Mar; 268(3):E467–75. <https://doi.org/10.1152/ajpendo.1995.268.3.E467> PMID: 7534990
48. Baeyens N, Bandyopadhyay C, Coon BG, Yun S, Schwartz MA. Endothelial fluid shear stress sensing in vascular health and disease. *J Clin Invest.* 2016 Mar 1; 126(3):821–8. <https://doi.org/10.1172/JCI83083> PMID: 26928035

49. Wilson C, Zhang X, Buckley C, Heathcote HR, Lee MD, McCarron JG. Increased Vascular Contractility in Hypertension Results From Impaired Endothelial Calcium Signaling. *Hypertension*. 2019 Nov; 74(5):1200–14. <https://doi.org/10.1161/HYPERTENSIONAHA.119.13791> PMID: 31542964
50. Plagge A, Kelsey G. Imprinting the Gnas locus. *Cytogenet Genome Res*. 2006; 113(1–4):178–87. <https://doi.org/10.1159/000090830> PMID: 16575178
51. Holmes R, Williamson C, Peters J, Denny P, RIKEN GER Group, GSL Members, et al. A Comprehensive Transcript Map of the Mouse Gnas Imprinted Complex. *Genome Res*. 2003 Jun; 13(6b):1410–5. <https://doi.org/10.1101/gr.955503> PMID: 12819140
52. Turan S, Bastepe M. The GNAS Complex Locus and Human Diseases Associated with Loss-of-Function Mutations or Epimutations within This Imprinted Gene. *Horm Res Paediatr*. 2013; 80(4):229–41. <https://doi.org/10.1159/000355384> PMID: 24107509
53. Weinstein LS, Yu S, Warner DR, Liu J. Endocrine Manifestations of Stimulatory G Protein  $\alpha$ -Subunit Mutations and the Role of Genomic Imprinting. *Endocr Rev*. 2001 Oct 1; 22(5):675–705.
54. Lewis DL, Weight FF, Luini A. A guanine nucleotide-binding protein mediates the inhibition of voltage-dependent calcium current by somatostatin in a pituitary cell line. *Proc Natl Acad Sci*. 1986 Dec; 83(23):9035–9. <https://doi.org/10.1073/pnas.83.23.9035> PMID: 2431411
55. Codina J, Yatani A, Grenet D, Brown AM, Birnbaumer L. The  $\alpha$  Subunit of the GTP Binding Protein Gk Opens Atrial Potassium Channels. *Science*. 1987 Apr 24; 236(4800):442–5.
56. Ehret GB, Caulfield MJ. Genes for blood pressure: an opportunity to understand hypertension. *Eur Heart J*. 2013 Apr 1; 34(13):951–61. <https://doi.org/10.1093/eurheartj/ehs455> PMID: 23303660
57. Keeney S. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. In: Egel R, Lankenau DH, editors. *Recombination and Meiosis* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2023 Jun 19]. p. 81–123. (Genome Dynamics and Stability; vol. 2). Available from: [http://link.springer.com/10.1007/7050\\_2007\\_026](http://link.springer.com/10.1007/7050_2007_026)
58. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. *Nat Rev Immunol*. 2007 Oct; 7(10):803–15. <https://doi.org/10.1038/nri2171> PMID: 17893694
59. Dinh QN, Drummond GR, Sobey CG, Chrissobolis S. Roles of inflammation, oxidative stress, and vascular dysfunction in hypertension. *BioMed Res Int*. 2014; 2014:1–11. <https://doi.org/10.1155/2014/406960> PMID: 25136585
60. Poole DC, Erickson HH. Exercise Physiology of Terrestrial Animals. In: *Duke's Physiology of Domestic Animals*. 13th ed. John Wiley & Sons, Inc.; 2015. p. 443–63.
61. Moris D, Spartalis M, Spartalis E, Karachaliou GS, Karaolani GI, Tsourouflis G, et al. The role of reactive oxygen species in the pathophysiology of cardiovascular diseases and the clinical significance of myocardial redox. *Ann Transl Med*. 2017 Aug; 5(16):326–326. <https://doi.org/10.21037/atm.2017.06.27> PMID: 28861423
62. Dubois-Deruy E, Peugnet V, Turkieh A, Pinet F. Oxidative Stress in Cardiovascular Diseases. *Antioxidants*. 2020 Sep 14; 9(9):864. <https://doi.org/10.3390/antiox9090864> PMID: 32937950
63. Sharma S, Aldred MA. DNA Damage and Repair in Pulmonary Arterial Hypertension. *Genes*. 2020 Oct 19; 11(10):1224. <https://doi.org/10.3390/genes11101224> PMID: 33086628
64. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. *Nature*. 2022 Oct 27; 610(7933):704–12. <https://doi.org/10.1038/s41586-022-05275-y> PMID: 36224396
65. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. 2016 Nov; 167(5):1415–1429.e19. <https://doi.org/10.1016/j.cell.2016.10.042> PMID: 27863252
66. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*. 2019 Jan; 104(1):65–75. <https://doi.org/10.1016/j.ajhg.2018.11.008> PMID: 30595370
67. Wain LV, Vaez A, Jansen R, Joehanes R, Van Der Most PJ, Erzurumluoglu AM, et al. Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* [Internet]. 2017 Sep [cited 2023 Jun 19];70(3). Available from: <https://www.ahajournals.org/doi/10.1161/HYPERTENSIONAHA.117.09438> PMID: 28739976
68. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinforma* [Internet]. 2016 Jun [cited 2023 Jun 19];54(1). Available from: <https://onlinelibrary.wiley.com/doi/10.1002/cpbi.5> PMID: 27322403

69. Müller R, Weirick T, John D, Militello G, Chen W, Dimmeler S, et al. ANGIOGENES: knowledge database for protein-coding and noncoding RNA genes in endothelial cells. *Sci Rep*. 2016 Sep 1; 6(1):32475. <https://doi.org/10.1038/srep32475> PMID: 27582018
70. Han Y, Ali MK, Dua K, Spiekerkoetter E, Mao Y. Role of Long Non-Coding RNAs in Pulmonary Arterial Hypertension. *Cells*. 2021 Jul 26; 10(8):1892. <https://doi.org/10.3390/cells10081892> PMID: 34440661
71. Wołowiec Ł, Mędlewska M, Osiak J, Wołowiec A, Grzešek E, Jaśniak A, et al. MicroRNA and lncRNA as the Future of Pulmonary Arterial Hypertension Treatment. *Int J Mol Sci*. 2023 Jun 4; 24(11):9735. <https://doi.org/10.3390/ijms24119735> PMID: 37298685
72. Kingsley NB, Hamilton NA, Lindgren G, Orlando L, Bailey E, Brooks S, et al. “Adopt-a-Tissue” Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare. *Front Genet*. 2021 Mar 26; 12:1–9. <https://doi.org/10.3389/fgene.2021.649959> PMID: 33841506
73. Kingsley NB, Kern C, Creppe C, Hales EN, Zhou H, Kalbfleisch TS, et al. Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq. *Genes*. 2019 Dec 18; 11(1):3. <https://doi.org/10.3390/genes11010003> PMID: 31861495
74. Hinckley KA, Fearn S, Howard BR, Henderson IW. Nitric oxide donors as treatment for grass induced acute laminitis in ponies. *Equine Vet J*. 1996 Jan; 28(1):17–28. <https://doi.org/10.1111/j.2042-3306.1996.tb01586.x> PMID: 8565949
75. Gauff F, Patan-Zugaj B, Licka TF. Hyperinsulinaemia increases vascular resistance and endothelin-1 expression in the equine digit: Endothelin-1 and vasoconstriction after hyperinsulinaemia. *Equine Vet J*. 2013 Sep; 45(5):613–8.
76. Katwa LC, Johnson PJ, Ganjam VK, Kreeger JM, Messer NT. Expression of endothelin in equine laminitis. *Equine Vet J*. 1999 May; 31(3):243–7. <https://doi.org/10.1111/j.2042-3306.1999.tb03180.x> PMID: 10402139
77. Backus M. Lameness in the horse with special reference to acute laminitis. *J Am Vet Med Assoc*. 1937;(91):64.
78. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, Zaibert V, et al. The Earliest Horse Harnessing and Milking. *Science*. 2009 Mar 6; 323(5919):1332–5. <https://doi.org/10.1126/science.1168594> PMID: 19265018
79. Heesch CM, Kline DD, Hasser EM. Control Mechanisms of the Circulatory System. In: *Duke’s Physiology of Domestic Animals*. 13th ed. John Wiley & Sons, Inc.; 2015. p. 353.
80. Mazic S, Suzic Lazic J, Dekleva M, Antic M, Soldatovic I, Djelic M, et al. The impact of elevated blood pressure on exercise capacity in elite athletes. *Int J Cardiol*. 2015 Feb; 180:171–7. <https://doi.org/10.1016/j.ijcard.2014.11.125> PMID: 25460373
81. Navas de Solis C, Slack J, Boston RC, Reef VB. Hypertensive cardiomyopathy in horses: 5 cases (1995–2011). *J Am Vet Med Assoc*. 2013 Jul 1; 243(1):126–30. <https://doi.org/10.2460/javma.243.1.126> PMID: 23786201
82. Moore JN, Garner HE, Coffman JR. Haematological changes during development of acute laminitis hypertension. *Equine Vet J*. 1981 Oct; 13(4):240–2. <https://doi.org/10.1111/j.2042-3306.1981.tb03506.x> PMID: 7318801
83. Bailey SR, Habershon-Butcher JL, Ransom KJ, Elliott J, Menzies-Gow NJ. Hypertension and insulin resistance in a mixed-breed population of ponies predisposed to laminitis. *Am J Vet Res*. 2008 Jan; 69(1):122–9. <https://doi.org/10.2460/ajvr.69.1.122> PMID: 18167097
84. Lorello O, Heliczner N, Casoni D, Schüpbach G, Navas de Solis C. Correlation of Blood Pressure With Splenic Volume in Horses, Daily Variation in Blood Pressure, and “White Coat Hypertension.” *J Equine Vet Sci*. 2018 Apr 1; 63:41–5.
85. Söder J, Bröjer JT, Nostell KE. Interday variation and effect of transportation on indirect blood pressure measurements, plasma endothelin-1 and serum cortisol in Standardbred and Icelandic horses. *Acta Vet Scand*. 2012 Jun 10; 54:37. <https://doi.org/10.1186/1751-0147-54-37> PMID: 22682151
86. Baynash AG, Hosoda K, Giaid A, Richardson JA, Emoto N, Hammer RE, et al. Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. *Cell*. 1994 Dec; 79(7):1277–85. [https://doi.org/10.1016/0092-8674\(94\)90018-3](https://doi.org/10.1016/0092-8674(94)90018-3) PMID: 8001160
87. Kusafuka T, Wang Y, Puri P. Mutation analysis of the RET, the endothelin-B receptor, and the endothelin-3 genes in sporadic cases of Hirschsprung’s disease. *J Pediatr Surg*. 1997 Mar; 32(3):501–4. [https://doi.org/10.1016/s0022-3468\(97\)90616-3](https://doi.org/10.1016/s0022-3468(97)90616-3) PMID: 9094028
88. Kusafuka T, Puri P. Mutations of the endothelin-B receptor and endothelin-3 genes in Hirschsprung’s disease. *Pediatr Surg Int*. 1997 Jan; 12(1):19–23. <https://doi.org/10.1007/BF01194795> PMID: 9035203

89. Monti L, Arrigucci U, Rossi A. Insights into Endothelin-3 and Multiple Sclerosis. *Biomol Concepts*. 2020 Jun 25; 11(1):137–41. <https://doi.org/10.1515/bmc-2020-0012> PMID: 32589590
90. Read AP, Newton VE. Waardenburg syndrome. *J Med Genet*. 1997 Aug 1; 34(8):656–65. <https://doi.org/10.1136/jmg.34.8.656> PMID: 9279758
91. Jastrzebska B. GPCR: G protein complexes—the fundamental signaling assembly. *Amino Acids*. 2013 Dec; 45(6):1303–14. <https://doi.org/10.1007/s00726-013-1593-y> PMID: 24052187
92. Hepler JR, Gilman AG. G proteins. *Trends Biochem Sci*. 1992 Oct; 17(10):383–7. [https://doi.org/10.1016/0968-0004\(92\)90005-t](https://doi.org/10.1016/0968-0004(92)90005-t) PMID: 1455506
93. Weis WI, Kobilka BK. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu Rev Biochem*. 2018 Jun 20; 87(1):897–919. <https://doi.org/10.1146/annurev-biochem-060614-033910> PMID: 29925258
94. Gilman AG. G PROTEINS: TRANSDUCERS OF RECEPTOR-GENERATED SIGNALS. *Annu Rev Biochem*. 1987 Jun; 56(1):615–49. <https://doi.org/10.1146/annurev.bi.56.070187.003151> PMID: 3113327
95. MedlinePlus [Internet]. Bethesda (MD): National Library of Medicine (US). GNAS gene [Internet]. 2015 [cited 2023 Jun 20]. Available from: <https://medlineplus.gov/genetics/gene/gnas/>
96. Danzi S, Klein I. Thyroid hormone and blood pressure regulation. *Curr Hypertens Rep*. 2003 Dec; 5(6):513–20. <https://doi.org/10.1007/s11906-003-0060-7> PMID: 14594573
97. Gu Y, Zheng L, Zhang Q, Liu L, Meng G, Yao Z, et al. Relationship between thyroid function and elevated blood pressure in euthyroid adults. *J Clin Hypertens*. 2018 Oct; 20(10):1541–9. <https://doi.org/10.1111/jch.13369> PMID: 30260550
98. Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al. EquCab3, an Updated Reference Genome for the Domestic Horse [Internet]. *Genomics*; 2018 Apr [cited 2022 May 16]. Available from: <http://biorxiv.org/lookup/doi/10.1101/306928>
99. Velie BD, Fegraeus KJ, Solé M, Rosengren MK, Røed KH, Ihler CF, et al. A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genet*. 2018 Dec; 19(1):80. <https://doi.org/10.1186/s12863-018-0670-3> PMID: 30157760
100. Hui L, DeMonte T, Ranade K. Genotyping Using the TaqMan Assay. *Curr Protoc Hum Genet* [Internet]. 2008 Jan [cited 2022 Sep 12];56(1). Available from: <https://onlinelibrary.wiley.com/doi/10.1002/0471142905.hg0210s56> PMID: 18428424
101. Árnason T. The Importance of Different Traits in Genetic Improvement of Trotters. In *Proceedings of World Congress on Genetics Applied to Livestock Production*. University of Guelph, Guelph, Canada; 1994.
102. Jurinke C, van den Boom D, Cantor CR, Köster H. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol Clifton NJ*. 2001; 170:103–16. <https://doi.org/10.1385/1-59259-234-1:103> PMID: 11357675
103. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, et al. Estimation and Tests of Haplotype–Environment Interaction when Linkage Phase Is Ambiguous. *Hum Hered*. 2003; 55(1):56–65. <https://doi.org/10.1159/000071811> PMID: 12890927
104. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl*. 2010 Mar 15; 26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
105. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
106. Zhigulev A, Sahlén P. Targeted Chromosome Conformation Capture (HiCap). In: Sexton T, editor. *Spatial Genome Organization* [Internet]. New York, NY: Springer US; 2022 [cited 2023 May 15]. p. 75–94. (Methods in Molecular Biology; vol. 2532). Available from: [https://link.springer.com/10.1007/978-1-0716-2497-5\\_5](https://link.springer.com/10.1007/978-1-0716-2497-5_5)
107. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 2009 Oct 9; 326(5950):289–93. <https://doi.org/10.1126/science.1181369> PMID: 19815776
108. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012 Apr; 9(4):357–9. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
109. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. *Nucleic Acids Res*. 2021 Jan 8; 49(D1):D916–23. <https://doi.org/10.1093/nar/gkaa1087> PMID: 33270111

110. Haider S, Waggott D, Lalonde E, Fung C, Liu FF, Boutros PC. A bedr way of genomic interval processing. *Source Code Biol Med*. 2016 Dec; 11(1):14. <https://doi.org/10.1186/s13029-016-0059-5> PMID: 27999613
111. Anil A, Spalinskas R, Åkerborg Ö, Sahlén P. HiCapTools: a software suite for probe design and proximity detection for targeted chromosome conformation capture applications. Berger B, editor. *Bioinformatics*. 2018 Feb 15; 34(4):675–7. <https://doi.org/10.1093/bioinformatics/btx625> PMID: 29444232
112. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009 Jul 1; 37(Web Server):W305–11. <https://doi.org/10.1093/nar/gkp427> PMID: 19465376
113. Felkel S, Vogl C, Rigler D, Dobretsberger V, Chowdhary BP, Distl O, et al. The horse Y chromosome as an informative marker for tracing sire lines. *Sci Rep*. 2019 Dec; 9(1):6095. <https://doi.org/10.1038/s41598-019-42640-w> PMID: 30988347
114. Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol*. 2015 Oct; 25(19):2577–83. <https://doi.org/10.1016/j.cub.2015.08.032> PMID: 26412128
115. Orlando L, Librado P. Origin and Evolution of Deleterious Mutations in Horses. *Genes*. 2019 Aug 28; 10(9):649. <https://doi.org/10.3390/genes10090649> PMID: 31466279
116. Schubert M, Ermini L, Sarkissian CD, Jónsson H, Ginolhac A, Schaefer R, et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 2014 May; 9(5):1056–82. <https://doi.org/10.1038/nprot.2014.063> PMID: 24722405
117. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 2014 Dec; 15(1):356. <https://doi.org/10.1186/s12859-014-0356-4> PMID: 25420514