A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism

Brandon D. Velie1*, Kim Jäderkvist Fegraeus1, Marina Solé1, Maria K. Rosengren1, Knut H. Røed2, Carl-Fredrik Ihler3, Eric Strand3 and Gabriella Lindgren1,4

Abstract

Background: Although harness racing is of high economic importance to the global equine industry, significant genomic resources have yet to be applied to mapping harness racing success. To identify genomic regions associated with harness racing success, the current study performs genome-wide association analyses with three racing performance traits in the Norwegian-Swedish Coldblooded Trotter using the 670 K Axiom Equine Genotyping Array.

Results: Following quality control, 613 horses and 359,635 SNPs were retained for further analysis. After strict Bonferroni correction, nine genome-wide significant SNPs were identified for career earnings. No genome-wide significant SNPs were identified for number of gallops or best km time. However, four suggestive genome-wide significant SNPs were identified for number of gallops, while 19 were identified for best km time. Multiple genes related to intelligence, energy metabolism, and immune function were identified as potential candidate genes for harness racing success.

Conclusions: Apart from the physiological requirements needed for a harness racing horse to be successful, the results of the current study also advocate learning ability and memory as important elements for harness racing success. Further exploration into the mental capacity required for a horse to achieve racing success is likely warranted.

Keywords: Genetic, Horse, Performance, Racehorse

Background

Regardless of horseracing discipline, speed, or perhaps more appropriately, unparalleled speed, is the “holy grail” of almost every horse owner, trainer, and breeder. However, speed alone does not necessarily equate to success on the racecourse. The manner in which a horse demonstrates speed is critical to its racing success [1–3]. For example, while the ability to gallop fast may result in a champion Thoroughbred (TB) or Quarter Horse (QH), the same ability in a Standardbred (SB) or Coldblooded Trotter (CT) is of little value. In SB and CT racing, horses undoubtedly require speed, but galloping, a four-beat gait, results in disqualification [1–3]. Thus, speed in these breeds must be demonstrated at trot, a contralateral two-beat gait, or pace, an ipsilateral two-beat gait [1–4]. Consequently, racing success in SBs, CTs, and other harness racing breeds depends not on an individual’s capacity for speed, but on an individual’s capacity for speed in a specific gait.

To date, significant genomic resources have been applied in studies attempting to map speed and racing success in TBs and QHs [4–18]. While these studies have proven to be of great value for gallop racing breeds, their applicability to harness racing breeds has been limited [19–21]. In fact, a genomic study exploring locomotion pattern in Icelandic horses is arguably the most influential

* Correspondence: brandon.velie@slu.se

© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
study to date when it comes to racing success in harness racing breeds [22]. The study discovered that a premature stop codon in the doublesex and mab-3 related transcription factor 3 (DMRT3) gene had a major effect on the pattern of locomotion in the horse, resulting in some horses possessing the ability to trot or pace at high speed without transitioning into gallop [22]. Follow-up studies were then able to demonstrate that for many harness racing breeds the DMRT3 mutation was associated with racing success (e.g. increased earned prize money) [20, 23–26]. However, the mutation appears to only account for between 0 and 6.3% of the phenotypic variation in traits widely used to evaluate racing success [24, 26]. Considering that heritability estimates for some of these traits are as high as 0.38, the likelihood of other genes playing a significant role in a harness racing horse’s success is high [20, 23, 27–38]. Nevertheless, genome-wide association (GWA) studies with harness racing performance are lacking [23].

Despite the fact that harness racing success is of high economic importance to the global equine industry, genome scans for performance appear to predominantly target Thoroughbreds and sport horses (e.g. Warmbloods) [4–6, 8–14, 39–41]. Most studies involving harness racing breeds tend to focus on detecting genes underlying certain conditions and diseases [42–47]. While disease studies are undoubtedly important for improved animal well-being, a greater awareness of the genes, and by extension the underlying biological mechanisms, involved in racing success are also likely to prove highly valuable. A deeper understanding of the biology underpinning success in a racehorse ultimately can lead to more targeted medical treatments as well as a heightened awareness of the physical limitations (e.g. lack of speed at trot) some horses will inevitably possess.

Motivated by these facts, we conducted a GWA study to identify genomic regions and genes associated with harness racing success using the Norwegian-Swedish Coldblooded Trotter (NSCT). Norwegian-Swedish Coldblooded Trotters are ideally suited for GWA analyses of harness racing performance as their small population is not only likely to correspond with low within breed genetic variation, but the limited region in which NSCTs are eligible to race is also likely to reduce environmental variation. Thus, a more accurate assessment of the relationship between genomic regions and harness racing performance is achievable.

Methods

Data collection

Pedigree information and performance data on all raced and unraced NSCTs born between 1 January 2000 and 31 December 2009 were provided by the trotting associations in both Norway and Sweden (Norsk Rikstoto and Svensk Travsport). Pedigree information included horse name, horse id, date of birth, country of birth, sex, breeder, sire id, and dam id. Performance data, as of 8 February 2017, was presented per race per horse (i.e. data included individual race records with each record corresponding to a given horse’s specific performance in the race). This data included non-competitive premie and qualification races, with each record containing information on horse id, race date, race track, race type, race distance, trainer, owner, driver, finish position, prize money earned, gallop status, and average km time [48, 49].

Sample acquisition

In order to reflect the raced population as accurately as possible, a list of raced individuals was randomly generated from the data described above using the statistical software R [50]. In addition to having raced, two requirements were set for inclusion in the study: 1. Hair and/or blood samples had to be readily accessible from the pedigree registration authorities in either Norway (Department of Basic Sciences and Aquatic Medicine, Norwegian University of Life Sciences) or Sweden (Animal Genetics Laboratory, Swedish University of Agricultural Sciences) 2. Sufficient sample material had to be available to ensure DNA quality standards would likely be achieved. The first 661 horses on the list that met these criteria were included in the study. Average relatedness within the selected group, estimated using the genetic software Contribution, Inbreeding, Coancestry, was 0.16 (interquartile range [IQR] 0.11–0.18) [51].

DNA isolation

Deoxyribonucleic acid was prepared from the hair roots using a standard hair-preparation procedure. Briefly, 186 μL Chelex 100 Resin (Bio-Rad Laboratories, Hercules, CA) and 14 μL of proteinase K (20 mg/mL; Merck KgaA, Darmstadt, Germany) were added to the sample. The mix was incubated at 56 °C for 2 h and the proteinase K was inactivated for 10 min at 95 °C. For DNA preparation from blood samples, 350 μL of blood was used and isolated on the Qiashympson instrument using the Qiashympy DSP DNA mini kit (Qiagen, Hilden, Germany). A summary of the final horses selected for genotyping is shown in Table 1.

Genotyping and quality control

Prior to quality control (QC) the data set consisted of individuals genotyped using the 670 K Axiom Equine Genotyping Array (n = 570) and the 670 K+ Axiom Equine Genotyping Array (n = 91). The data from the two arrays were subsequently merged based on SNP name, chromosome, and position, yielding a combined SNP data set of 611,888 SNPs for 661 horses. Iterative QC was then performed with the GenABEL package in R to remove poorly genotyped and noisy data using the
following thresholds: minor allele frequency (MAF) (< 0.5%), missing genotypes per single nucleotide polymorphism (SNP) (> 5%), missing SNPs per sample (> 15%), and Hardy-Weinberg equilibrium (HWE) (first QC \( p < 1 \times 10^{-10} \); second QC FDR < 0.2) [50].

Phenotyping
Pedigree and performance data for the 661 genotyped horses were structured for analyses using custom scripts written in Perl v5.20.1. Sex classifications were based on the official sex of the horse at the end of the study. Number of career starts was defined as the total number of competitive races for each individual (i.e. premie and qualification races were excluded). Number of gallops (NG) was defined as the total number of competitive races in which the horse was recorded as galloping at some point during the race. Career earnings (CE) were calculated as the total amount of prize money won for each horse as of 8 February 2017. Prize money won in Sweden (SEK) was converted to Norwegian currency (NOK) based on the average exchange rate for the year in which the race occurred [52]. Horses having participated in only premie or qualification races were given a value for career earnings of −1 NOK in order to distinguish them from horses that had zero career earnings, despite having raced competitively. Best km time (BT), independent of starting method, was defined as the fastest average km time for a competitive race in which the horse was not recorded as galloping during the race (i.e. races in which a horse galloped, regardless of if the horse was disqualified, were excluded).

Genome-wide association analyses
Genome-wide association (GWA) analyses were performed using the GenABEL package in R [50]. The package was used to compute an autosomal genomic kinship matrix as well as perform standard K-means clustering. K-means clustering with \( K = \{1,2,\ldots,10\} \) were executed to determine the number of clusters (subpopulations). For each iteration, the sum of within-cluster sum of squares (\( \sum WCSS \)) was calculated and subsequently plotted vs. \( K \). To define the subpopulations, the number of clusters corresponding with the first inflection point (\( K = 3 \)) was chosen [53]. The multidimensional scaling (MDS) plot yielded no apparent outliers and a visualization of the genomic-kinship matrix and subpopulations using MDS can be seen in Fig. 1.

To account for any population stratification, genome-wide association analyses of CE, BT, and NG were performed using a mixed model-structured association approach ("mmscore" function with the "strata" option in GenABEL). Preliminary analyses indicated significant effects of sex, birth year, and number of career starts on all three traits. Country of birth was also shown to influence CE. As a result, sex, birth year, number of career starts, and country of birth were included as co-variants in the final analyses accordingly. Genome-wide significance for each of the analyses was determined by Bonferroni correction (\( p < 1.39 \times 10^{-7} \), corrected for total number of SNPs post QC) while a "suggestive" genome-wide significance threshold was also set at \( 1.0 \times 10^{-5} \). Manhattan and quantile-quantile plots were generated in R while the extent of linkage disequilibrium (LD) was estimated by calculating \( r^2 \) values between all pairs of SNPs with inter-SNP distances of less than 1 Mb using PLINK v1.09 (http://zzz.bwh.harvard.edu/plink/). The effB in the GenABEL result was regarded as the allele substitution effect and the proportion of phenotypic variance explained by each significant SNP was estimated as follows:

\[
VAR(\%) = \frac{2pq\beta^2}{S^2} \times 100
\]

Where \( p \) and \( q \) are the allele frequencies, \( \beta \) is the estimated allele substitution effect, and \( S^2 \) is the sample phenotypic variance. Stepwise regressions were then performed to estimate the total proportion of phenotypic variation explained by the multiple genome-wide associated SNPs (before and after pruning at \( r^2 > 0.2 \); PLINK

| Table 1 Summary information for the genotyped and the final sample of horses |
|-------------------------------------------------|
| Genotyped Horses | Final GWA Horses |
| N | N |
| **Sex** | | |
| Intact males | 70 | 62 |
| Females | 271 | 247 |
| Geldings | 320 | 304 |
| **Country of birth** | | |
| Norway | 360 | 312 |
| Sweden | 301 | 301 |
| **Year of birth** | | |
| 2000 | 45 | 36 |
| 2001 | 73 | 66 |
| 2002 | 84 | 80 |
| 2003 | 69 | 64 |
| 2004 | 56 | 49 |
| 2005 | 62 | 55 |
| 2006 | 70 | 67 |
| 2007 | 71 | 70 |
| 2008 | 63 | 59 |
| 2009 | 68 | 67 |
| **Total** | 661 | 613 |
command –indep-pairwaise 100 25 0.2) for each trait using the lm function in R [50]. The bioinformatics database Ensembl (http://www.ensembl.org/) was used for candidate gene screening. Genomic coordinates of genome-wide significant and suggestive genome-wide significant SNPs +/- 500 kb were used as inputs to generate a list of annotated genes using the Ensembl Bio-mart function. The PANTHER Classification system was then used to obtain an overview of the biological processes, molecular functions, and pathways known to be affected by these genes [54, 55].

Results
Following QC, 359,635 autosomal SNPs and 642 horses were available for association analyses. Of these individuals, 29 had only participated in non-competitive premie and/or qualification races. Initial association analyses with CE, combined with the vast array of reasons known to prevent a horse from competitive racing suggested the exclusion of these horses would help to reduce noise in the final analyses. As a result, only 613 horses, representing 120 sires (interquartile range [IQR] 1–5) and 547 dams (IQR 1–1), were included in the final GWA analyses (Table 1). Descriptive statistics for CE, BT, and NG in the final sample are presented in Table 2. Both CE and NG were not normally distributed and were subsequently log transformed for the GWA analyses. The extent of LD decayed faster across the final sample of horses with mean $r^2$ dropping below 0.20 by 3 kb (Additional file 1). The GWA analysis of CE yielded multiple genome-wide significant SNPs ($p < 1.39 \times 10^{-7}$), with the majority of these SNPs residing on *Equus caballus* chromosome (ECA) 6 (Fig. 2; Table 3; Additional file 2). Analyses of both BT and NG failed to result in genome-wide significant SNPs. However, two regions of interest were apparent based on the presence of slight peaks on ECA17 and ECA23 in the resulting Manhattan plots (Fig. 3; Additional file 2). Genome-wide significant SNPs for CE, suggestive genome-wide significant SNPs for BT and NG, and the nearest genes are shown in Table 3. The most significant SNP was detected at the 20,006,740 position on chromosome 28 (AX-104828170, $p = 9.01E-10$).
and presented an estimated allele substitution effect of $-7079.46$ NOK. The favourable allele appeared to be the T allele, with each C allele resulting in a negative effect on CE. Despite the fact that the frequency of the T allele was 98.2% and the frequency of the C allele was only 1.8%, the percentage of phenotypic variance explained by this SNP was 3.85%. The 32 SNPs in total were estimated to explain 18.34%, 18.71% and 33.17% of the variation for CE, NG and BT, respectively, in the population studied. Consequently, after LD pruning, only 17 SNPs remained – explaining 14.17%, 18.38% and 33.13% of the variation for CE, NG and BT, respectively (Table 3). Overall, 378, 144, and 23 candidate genes identified were associated with known biological processes, molecular functions, and pathways, respectively (Figs. 4, 5 and 6; Additional file 3).

**Discussion**

Knowing where, why, and how genes and athletic prowess intersect in a racehorse has long been the goal of countless researchers, veterinarians, breeders, trainers, and owners [4–38]. While great strides in this area have recently been made for gallop racing horses, similar advancements for harness racing horses have been limited [4–26]. Using the NSCT, the current study explored the genetic background for athletic prowess in a harness racing horse by performing GWA analyses and functional classification for three traits associated with harness racing success. These analyses resulted in a total of 32 SNPs of interest with 9 demonstrating genome-wide significance and 13 residing in genes. Subsequent functional classifications went on to provide further support of the complexity of harness racing success with several candidate genes involved in neurological, metabolic, and musculoskeletal regulation identified. Since a gene can be declared as a candidate gene if at least 1 out of the 4 following characteristics are present: 1) the gene has a known physiological role in the phenotype of interest, 2) the gene affects the trait in question based on studies of knockouts, mutations, or transgenics in other species, 3) The gene is preferentially expressed in organs related to the quantitative trait, or 4) the gene is preferentially expressed during developmental stages related to the phenotype, a large fraction of the genes identified in the current study can plausibly be considered as candidate genes.

| Table 2 Descriptive statistics of CE, BT, NG for the final sample of horses (n = 613) |
|----------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Career earnings (NOK) | Min  | 0  | 36,321 | 124,625 | 302,506 | 327,493 | 4,216,554 |
| Best km time (s) | 78.9  | 86.8 | 89.3  | 89.6  | 92.4  | 105.7  |
| Number of gallops | 0  | 5  | 11  | 14.4 | 21  | 123  |

CE career earnings, BT Best km time, NG number of gallops

![Fig. 2](image) Manhattan plot of the genome-wide association analysis of career earnings. The red horizontal line indicates the genome-wide significance level and the black horizontal line indicates the suggestive genome-wide significance level. Uncorrected $\lambda = 1.0532$. 

![Table 2](image) Descriptive statistics of CE, BT, NG for the final sample of horses (n = 613)
As a result, the following discussion prioritizes genes that contained variants with significant or suggestive associations with our traits of interest. The rationale for this prioritization is simply that for associated variants that reside outside of annotated genes, it is in general more difficult to determine which gene(s) the variants act on.

Two genome-wide significant SNPs associated with career earnings were located in the **GRIN2B** gene, a gene also identified in a previous study exploring pacing ability in Icelandic horses [57]. The gene has been shown to be involved in neural regulations in humans and...
laboratory species with mutations in the gene having been associated with neurodevelopmental disorders [58–60]. Considered to be an important factor for learning and memory, one can only speculate as to its association with career earnings in a harness racing horse. However, horses with a greater capacity to learn and adapt to the highly variable nuances of harness racing would conceivably be more likely to achieve racing success. On the other hand, the gene’s association with attention-deficit/hyperactivity disorder in humans suggests that perhaps certain horses lack the ability to focus on racing and training, thereby preventing or at least hindering their racing performance [61].

**ATPase copper transporting beta (ATP7B)**

A single suggestive genome-wide significant SNP associated with best time was located in the ATP7B gene on ECA17. The gene encodes a protein that functions as a monomer, exporting copper out of cells. Excess copper can cause serious toxicity with the process of excess copper disposal relying heavily on ATP7B [62, 63]. Over 500 mutations have been identified in the gene, 380 of which are considered to be disease causing mutations [64]. Elevated levels of copper in the body often result in muscle stiffness with acute muscle stiffness prior to a race having the potential to affect individual performance and chronic muscle stiffness likely to impact a horse’s conditioning, trainability, and overall capacity for speed [62, 63].

**Potassium channel regulator (KCNRG)**

The KCNRG gene on ECA17 was also identified by two SNPs demonstrating suggestive genome-wide significance for best time. The gene encodes a protein which regulates the activity of voltage-gated potassium channels with a study using songbirds suggesting potassium channels to be lineage specific [64, 65]. The same study also revealed that apart from broad expression in the

---

**Fig. 3** Manhattan plots of the genome-wide association analyses of best km time and number of gallops. The red horizontal lines indicate the genome-wide significance levels and the black horizontal lines indicate the suggestive genome-wide significance levels. Top panel: Best km time analysis, uncorrected $\lambda = 1.0902$. Bottom panel: Number of gallops analysis, uncorrected $\lambda = 1.0256$. 

Velie et al. BMC Genetics (2018) 19:80
brain a subset of potassium channel genes are selectively expressed, with the authors hypothesizing that the *KCNRG* gene may be associated with learning [65]. Although no previous studies of racehorse performance have specifically identified *KCNRG* as important for racing success, a large conserved haplotype on ECA17 has been advocated to have selective importance in Thoroughbreds and closely related breeds [4, 66].

Also of note is the role voltage-gate potassium channels play in Hyperkalemic periodic paralysis (HYPP), a
The condition, caused by a mutation in the sodium voltage-gated channel alpha subunit 4 (SCN4A) gene, manifests intermittently with clinical signs ranging from muscle fasciculation to signs of paresis [67–70]. Hyperkalemia, a term used to describe abnormally high levels of potassium in the blood, is often seen during or immediately after an attack. Voltage-gated potassium channels are thought to remain open, allowing continual potassium efflux, thereby promoting an open sodium channel configuration. As a result, an HYPP attack can be triggered or an already occurring attack can increase in severity [67, 70–72]. Horses with HYPP also tend to possess hypertrophic muscles; however, they have been reported as having a reduced tolerance to exercise with relatively more lactate being produced during exercise [73, 74].

Phosphatidylinositol-4-phosphate 5-kinase type 1 beta (PIPK1B)
A single suggestive genome-wide significant SNP associated with best time was also located in the PIP5K1B gene on ECA23. Three widely expressed isoforms of PIP5K1 are responsible for the regulation of the major pools of cellular phosphatidylinositols in mammalian tissues, with PIP5K1B negatively regulated in response to oxidative stress [75, 76]. Neurite outgrowth, a critical process for neuronal development, has also been shown as negatively regulated by PIP5K1A [77]. Since the current study is the first to suggest an association between PIP5K1B and racing success, understanding the roles PIPS1 isoforms have on a horse’s capacity for speed remains a task for future studies. However, genes that influence cell differentiation processes, such as endocytosis, assuredly contribute in some way or another to the physical limitations and overall performance of any racehorse.

Dedicator of cytokinesis 8 (DOCK8)
Perhaps the most obvious candidate gene for harness racing success in the current study was DOCK8. Five suggestive genome-wide significant SNPs indicated the importance of DOCK8 to a horse’s best time, with 4 of the SNPs located in the gene. Mutations in DOCK8 result in a form of hyper-IgE syndrome; however, loss or mutations of DOCK8 have also been associated with intelligence and motor retardation [78–83]. While the importance of intelligence in a racehorse has been briefly discussed above, in the case of DOCK8 the significance of the gene may lie with its link to motor skills. DOCK8 is not only located on ECA23, the same chromosome as DMRT3, but multiple studies have hypothesized some sort of commonality or overlap between DMRT-(1,2,3) gene effects and DOCK8 [22, 82, 84]. Despite the established association between DMRT3 and harness racing performance, additional research of DMRT3 in horses strongly suggest that the mutation is unlikely to be the single cause of gaiting ability [22, 53, 85–88]. Therefore,
it is conceivable that DOCK8 also significantly contributes to harness racing horse’s propensity to exhibit speed at trot or pace.

**Phosphodiesterase 3A (PDE3A)**
The protein encoded by the PDE3A gene, a gene on ECA6 in which a single genome-wide significant SNP associated with career earnings is located, plays a critical role in cardiovascular function [89–91]. The encoded protein regulates vascular smooth muscle contraction and relaxation and has been linked to familial hypertension, cardiovascular disease, and fertility [89–92]. Healthy cardiovascular function is important for racing success as the act of racing undeniably requires a higher than resting-level of oxygen to support the horse’s increased muscle activity. A mutation in the PDE3A gene that ultimately alters cardiovascular function could potentially prevent a horse from meeting the higher metabolic demands of racing, thus decreasing his/her chances of winning and limiting his/her career earnings. On the contrary, an advantageous mutation in the gene could allow some horses to perform at an even greater cardiovascular level, increasing their likelihood of winning races and earning more prize money.

**Inositol polyphosphate-5-phosphatase D (INPP5D) & SRY-box 5 (SOX5)**
Also identified by single genome-wide significant SNPs on ECA6 were the INPP5D and the SOX5 genes. The INPP5D gene is an important regulator of immune cell signaling, while the SOX5 gene is involved in embryonic development and has been associated with multiple human diseases and disorders [93–99]. Moreover, both genes have been suggested as important in B cell activity indicating that their association with career earnings in the current study may be rooted in the immune response of a horse [95, 100]. However, mutations in SOX5 have also been theorized to disrupt neuronal development and function [101, 102].

**Other candidate genes**
Regions on ECA1, ECA7, and ECA16 have also previously been described as important for endurance performance traits, while regions on ECA14 and ECA18 associated with galloping in other studies do not appear to play a significant role in harness racing [4–8, 10–15, 19–21, 39, 66]. This likely suggest a greater demand for endurance in harness racing compared to galloping and is perhaps a sign of the different physiological demands for speed in trot versus speed in gallop. Candidate genes for harness racing success in the current study were also identified on ECA1, ECA2, ECA6, ECA7, ECA16, ECA17, ECA23, ECA25, ECA28, ECA29, and ECA31. However, it is important to note that the MAF threshold applied in the current study is slightly lower than is generally accepted. Although this may have inadvertently resulted in some SNP associations being simply by chance, it is also plausible that the lower MAF threshold allowed for the capture of candidate genes/regions that are perhaps the difference between an elite horse and a very, very good horse. Racing performance is undoubtedly complex and the unique history of the NSCT, being a blend of draught horse and racehorse, means that rare variants cannot be ruled out purely because they are rare – particularly when one considers the rarity of an elite racehorse. While not all candidate variants GENES are discussed above, the results of the current analyses clearly suggest that different molecular and cellular events mediate adaptive processes in the neuromusculoskeletal system in response to exercise. High intensity exercise (e.g. racing) is known to be associated with significant physiological adaptations in the neuromuscular system in equine athletes with prolonged and intense exercise potentially resulting in oxidative damage to cellular constituents [103]. Moreover, the importance of the central nervous system (CNS) as a critical “central governing” factor in sporting performance has been previously documented in endurance horses with exercise shown to induce several biological processes that regulate neurological functions that help to maintain good mental health [104]. Our results add to this line of thought, providing further evidence that genes involved in neural regulations (e.g. GRIN2B) likely play an important role in controlling the fundamental biological processes underlying adaptation to equine athletic performance.

**Conclusions**
After strict Bonferroni correction, 9 genome-wide significant and 23 suggestive genome-wide significant SNPs associated with harness racing success were identified. These SNPs were located on ECA1, ECA2, ECA6, ECA7, ECA16, ECA17, ECA23, ECA25, ECA28, ECA29, and ECA31 with eight genes (GRN2B, DOCK8, ATP7B, KCNRG, PIP5K1B, PDE3A, INPP5D, SRY5) suggested as strong candidate genes for harness racing success. Apart from the physical attributes required to achieve racing success, multiple candidate genes identified in the current study also advocate learning ability and memory as critical to success. However, further analyses of these genes based on additional genetic and functional studies are required to explore this notion in greater detail. Moreover, future studies should also consider a validation study with an independent population as well as sequencing of candidate genes to better identify causal alleles.
Additional files

**Additional file 1:** Extent of linkage disequilibrium of the final sample of horses. Pairwise r² was calculated between each SNP within 1 Mb. (TIF 15316 kb)

**Additional file 2:** QQ plots for earnings, best km time, and number of gallops analyses. Top panel - corrected QQ plot for earnings analysis (Uncorrected $\lambda = 1.0532$); Middle panel - corrected QQ plot for best km time analysis ($\lambda = 1.0902$); Bottom panel - corrected QQ plot for number of gallops analysis ($\lambda = 1.0256$). (TIF 111 kb)

**Additional file 3:** Functional classification gene list from PANTHER analysis. (XLSX 27 kb)

Abbreviations

ATP7B: ATPase copper transporting beta; BT: Best km time; CE: Career earnings; CT: Coldblooded trotter; DMRT3: Doublesex and mab-3 related transcription factor 3; DOCX8: Deducator of cytokinesis, ECA: Equus caballus chromosome; GRIN2B: Glutamate ionotropic receptor NMDA type subunit 2B; GWA: Genome-wide association; INPP5D: Inositol polyphosphate-5-phosphatase D; KCNRG: Potassium channel regulator; NG: Number of gallops; NSCT: Norwegian-Swedish Coldblooded Trotter; PDE3A: Phosphodiesterase 3A; PIP5K1B: Phosphatidylinositol-4-phosphate 5-kine type 1 beta; QH: Quarter horse; SB: Standardbred trotter; SOX5: SRY-box 5; TB: Thoroughbred

Acknowledgements

We want to thank Christina Olsson, Annica Edberg, Knut Larsen, Channeli Lilja Lawrence, Thorvaldur Arnason, the Swedish Trotting Association, and the Norwegian Trotting Association for providing their support and the data/samples for the study.

Funding

This work was supported by the Swedish-Norwegian Foundation for Equine Research, H-15-47-075, http://hastforskning.se (GL, ES). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The data that support the findings of this study are available from the Swedish Trotting Association (Stockholm, Sweden) and the Norwegian Trotting Association (Oslo, Norway), 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 authors upon reasonable request and with permission of the Swedish Trotting Association (Stockholm, Sweden) and the Norwegian Trotting Association (Oslo, Norway).

Authors’ contributions

BDV, KJF, CI, ES, and GL conceived and designed the experiments; KJF, MKR, and KHR contributed to sampling. GL and ES contributed the reagents and resources; BDV and MS analyzed the data and drafted the manuscript; BDV, KJF, CI, ES, and GL discussed and contributed to data analysis; All authors read and approved the final manuscript.

Ethics approval and consent to participate

All experimental procedures and sample collection methods were approved by the Ethics Committee for Animal Experiments in Uppsala, Sweden [Number: C 121/14]. Samples used in the study were already available at either the Animal Genetics Laboratory at SLU in Uppsala, Sweden or the Department of Basic Sciences and Aquatic Medicine at the Norwegian University of Life Sciences in Oslo, Norway as they previously had been used for parentage testing. Permission to use the samples was granted from the Swedish Trotting Association and the Norwegian Trotting Association (the owners of the samples per the rules/guidelines of the industry).

Consent for publication

Not applicable.

Competing interests

The authors have the following interest: GL is a co-inventor on a granted patent concerning commercial testing of the DMRT3 mutation: A method to predict the pattern of locomotion in horses. PCT EP 12747875.8. European patent registration date: 2011-05-05, US patent registration date: 2011-08-03. There are no further patents, products in development, or marketed products to declare.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

1Department of Animal Breeding & Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. 2Department of Basic Sciences and Aquatic Medicine, Norwegian University of Life Sciences, Oslo, Norway. 3Department of Companion Animal Clinical Sciences, Norwegian School of Veterinary Science, Oslo, Norway. 4Department of Biosystems, KU Leuven, 3001 Leuven, Belgium.

Received: 3 February 2018 Accepted: 20 August 2018

References

1. United States Trotting Association. Charter, Bylaws, Rules and regulations. In: Rule book. United states trotting association. 2018. http://www.ustrotting.com/pdf/USTRuleBook.pdf. Accessed 10 Jan 2018.
2. Harness Racing Australia. Australian harness racing rules. Harness racing Australia Inc. 2017. http://www.harness.org.au/rules/AHRRules.pdf. Accessed 10 Jan 2018.
3. Union Européenne du Trot. International Agreement On Trotting Races. Union Européenne du Trot. 2018. http://www.uet-trot.eu/images/pdf/uet/en/publications/international_agreement_on_trotting_races.pdf. Accessed 10 Jan 2018.
4. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, et al. Genome-wide analysis reveals selection for optimum traits in domestic horse breeds. PLoS Genet. 2013;9(10):3211.
5. Hill EW, McGivney BA, Gu J, Whiston R, MacHugh DE. A genome-wide SNP-association study confirms a sequence variant (g.6649373C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for thoroughbred racehorses. BMC Genomics. 2010;11:552.
6. Petersen JL, Valberg SJ, Mickelson JR, McCue ME. HaploTivity diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. Anim Genet. 2014;45:827–35.
7. Tozaki T, Hill EW, Hirota K, Kako H, Gamahara H, Miyake T, Sugita S, Hasegawa T, Ishida N, Nakano Y, Kurosawa M. A cohort study of racing performance in Japanese thoroughbred racehorses using genome information on ECA18. Anim Genet. 2011;43:542–52.
8. Shin DH, Lee JW, Park JE, Choi HY, Oh HS, Kim HJ, Kim H. Multiple genes related to muscle identified through a joint analysis of a two-stage genome-wide association study for racing performance of 1,156 thoroughbreds. Asian Australas J Anim Sci. 2015;28:771–81.
9. Moon S, Lee JW, Shin D, Shin KY, Kim J, Choi KY, Kim J, Kim H. A genome-wide scan for selective sweeps in racing horses. Asian Australas J Anim Sci. 2015;28:1525–31.
10. Tozaki T, Miyake T, Kako H, Gawahara H, Sugita S, Hasegawa T, Ishida N, Hirota K, Nakano Y. A genome-wide association study for racing performances in thoroughbreds clarifies a candidate region near the MSTN gene. Anim Genet. 2010;41:26–35.
11. Hill EW, Gu J, Elvers SS, Fonseca RG, McGivney BA, Govindarajan P, Orr N, Katz LM, MacHugh DE. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. PLoS One. 2010;5:e8645.
12. Binns MM, Boehler DA, Lambert DH. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the thoroughbred horse in the USA. Anim Genet. 2010;41:154–8.
13. Hill EW, Fonseca RG, McGivney BA, Gu J, MacHugh DE, Katz LM. MSTN genotype (g.6649373C>T) association with speed indices in thoroughbred racehorses. J Appl Physiol. 2012;86–90.
14. Tozaki T, Sato F, Hill EW, Miyake T, Endo Y, Kako H, Gawahara H, Hirota K, Nakano Y, Nambo Y, Kurosawa M. Sequence variants at the myostatin gene

Additional files

**Additional file 1:** Extent of linkage disequilibrium of the final sample of horses. Pairwise r² was calculated between each SNP within 1 Mb. (TIF 15316 kb)

**Additional file 2:** QQ plots for earnings, best km time, and number of gallops analyses. Top panel - corrected QQ plot for earnings analysis (Uncorrected $\lambda = 1.0532$); Middle panel - corrected QQ plot for best km time analysis ($\lambda = 1.0902$); Bottom panel - corrected QQ plot for number of gallops analysis ($\lambda = 1.0256$). (TIF 111 kb)

**Additional file 3:** Functional classification gene list from PANTHER analysis. (XLSX 27 kb)
35. Velie BD, Hamilton NA, Wade CM. Heritability of racing performance in the Australian thoroughbred racing population. Anim Genet. 2015;46:23–9.

36. Tozaki T, Miyake T, Kakoi H, Gawahara H, Hirot A, Nakane Y, Kurosawa M. Heritability estimates for racing performance in Japanese thoroughbred racehorses using linear and non-linear model analyses. J Anim Breed Genet. 2012;129:402–8.

37. Velie BD, Hamilton NA, Wade CM. Performance selection for thoroughbreds in Hong Kong. Equine Vet J. 2015;47:43–7.

38. Sharrman P, Wilson AJ. Racehorses are getting faster. Biol Lett. 2015;11:20150310.

39. Ricard A, Robert C, Blouin C, Baste F, Torquet G, Morgensthal C, Rivière J, Mach N, Mata X, Schibler L, Barye E. Endurance exercise ability in the horses: a trait with complex polygenic determination. Front Genet. 2017;8:89.

40. Frischnecht M, Singer-Hasler H, Leeb T, Rieder S, Neuditschko M. Genome-wide association studies based on sequence-derived genotypes reveal new QTL associated with conformation and performance traits in the Franches-Montagnes horse breed. Anim Genet. 2016;47:227–9.

41. Schröder W, Klostermann A, Stock KF, Distl O. A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses. Anim Genet. 2012;43:392–400.

42. Lykkjøn S, Dolvik NI, McCue ME, Rendahl AK, Mickelson JR, Røed KH. Equine developmental orthopaedic diseases—a genome-wide association study of first phalangeal plantar osteochondral fragments in Standardbred trotters. Anim Genet. 2013;44:766–9.

43. Tyssöö S, Dupuis MC, Guérin G, Schibler L, Denoix JM, Elsen JM, Ricard A. Genome-wide association studies for osteochondrosis in French trotter horses. J Anim Sci. 2013;90:451–50.

44. McCoy AM, Beeson SK, Sliney BK, Lykkjøn S, Ralston SL, Mickelson JR, McCue ME. Identification and validation of risk loci for osteochondrosis in Standardbreds. BMC Genomics. 2016;17:41.

45. Lykkjøn S, Dolvik NI, McCue ME, Rendahl AK, Mickelson JR, Røed KH. Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. J Anim Sci. 2010;89:447–53.

46. Dupuis MC, Zhang Z, Druet T, Denoix JM, Charlier C, Lekeux P, Georges M. Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. Mamm Genome. 2011;22:613–20.

47. Dupuis MC, Zhang Z, Durlet T, Denoix JM, Charlier C, Lekeux P, Georges M. Detection of copy number variant loci in the horse genome and examination of their association with recurrent laryngeal neuropathy. Anim Genet. 2013;44:206–8.

48. Svensk Travsport: Kvalificerings- & premielopp. https://www.travsport.se/artikel/kvalificerings-_-_premielopp (2017). Accessed 1 Nov 2017.

49. Det Norske Travskap: Lover & Reglement. https://www.travskap.no (2017). Accessed 1 Nov 2017.

50. R Development Core Team: R: A Language and Environment for Statistical Computing. https://www.r-project.org/ (2018). Accessed 13 Jan 2018.

51. Sargolzaei M, Iwaisaki H, Colleau JJ. CFC, a tool for monitoring genetic diversity. In: Proceedings of the 8th world congress on genetics applied to livestock production, vol. 8. Brazil: Belo Horizonte; 2009. p. 1551

52. OFX: Historical Exchange Rates. https://www.ofx.com/en-au/forex-news/historical-exchange-rates/ (2017). Accessed 1 Nov 2017.

53. Everitt BS, Hothorn TA. Handbook of statistical analyses using R Chapman & Hall/CRC; 2010.

54. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from gene ontology and Reactome pathways, and data analysis tool enhancements. Nucl Acids Res. 2016; https://doi.org/10.1093/nar/gkw1138.

55. Mi H, Muruganujan A, Casasgande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8:1551–66.

56. Weller JI. Quantitative trait loci analysis in animals. 2nd ed. Oxfordshire: CABI; 2009.

57. Jäderkvist Fegraeus K, Hirschberg I, Arnason T, Andersson L, Velie BD, Andersson LS, Lindgren G. To pace or not to pace: a pilot study of four- and five-gaited Icelandic horses homzygous for the DMRT3 ‘Gait keeper’ mutation. Anim Genet. 2017;48:694–7.

58. Schröder W, Klostermann A, Stock KF, Distl O. A genome-wide association study for qualitative trait loci of show-jumping in Hanoverian warmblood horses. Anim Genet. 2012;43:392–400.
61. Kim JI, Yoo JH, Kim D, Jeong B, Kim BN. The effects of GRN2B and DRO4 gene variants on local functional connectivity in attention-deficit/ hyperactivity disorder. Brain Imaging Behav. 2017; https://doi.org/10.1007/ s11868-017-9680-2.

62. Concelli M, Iacobacci S, Chesi G, Carissimo A, Polischuk R. A systems biology approach reveals new endoplasmic reticulum-associated targets for the correction of the ATP7B mutant causing Wilson disease. Metallomics. 2016;8:290–30.

63. Kieffer CA, Medici V. Wilson disease: at the crossroads between genetics and epigenetics—a review of the evidence. Liver Res. 2017;1:121–30.

64. Usman H, Mathew MK. Potassium channel regulator KCNRG regulates surface expression of shaker-type potassium channels. Biochim Biophys Acta. 2010;1801:1301–5.

65. Lovell PV, Carleton JB, Mello CV. Genomics analysis of potassium channel genes in songbirds reveals molecular specializations of brain circuits for the maintenance and production of learned vocalizations. BMC Genomics. 2013; 14:470.

66. Gu J, Oh N, Park SD, Katz LM, Sulimova G, MacHugh DE, Hill EW. A genome scan for positive selection in thoroughbred horses. PLoS One. 2009;4:e5767.

67. Liu T, Lee SY. Phosphatidylinositol 4-phosphate 5-kinase Iβ mediates an autocrine loop that controls cell proliferation. Mol Med. 2013;45:e16.

68. Bowling AT, Byrns G, Spier S. Evidence for a single pedigree source of the Morgan and American curly horses. J Am Vet Med Assoc. 1992;200:240–3.

69. Hinchliff KW, Kaneps AJ, Geor RJ. Musculoskeletal System I. In: Equine sports medicine and surgery: Saunders Elsevier; 2014. p. 104–93.

70. Cannon SC, Brown RH Jr, Corey DP. A sodium channel defect in hyperkalemic periodic paralysis. Vet Clin North Am Equine Pract. 1997;13:129–44.

71. Steele D, Naylor JM. Hyperkalemic periodic paralysis, plasma lactate and exercise tolerance. J Equine Vet Sci. 1996;202:933.

72. Ricker K, Camacho LM, Grafe P, Lehmann-Horn F, Rueld R. Adynamia episodica congenita associated with exercise tolerance. J Exerc Rehabil. 2017;13:617–20.

73. Campos-Chillon F, Farmerie TA, Bouna GI, Clay CM, Camavele EM. Effects of aging on gene expression and mitochondrial DNA in the equine ococyte and follicle cells. Reprod Fertil Dev. 2015;27:925–33.

74. Hamilton MJ, Halvorsen EC, LeParde ND, Bosiljic M, Ho HW, Lam V, Baniñí J, Bennewit KL, Krystal G. SHR presses lung inflammation and inhibits marmoset tumor metastasis in BALB/c mice. Oncotarget. 2016;7:3677–91.

75. Somasundaram R, Fernandes S, Deering JJ, de Haar C, Kuipers EJ, Vogelar L, Middleton FA, van der Woude CJ, Peppelenbosch MP, Kerr WG, Fueller GM. Analysis of SHP1 expression and activity in Crohn’s disease patients. PLoS One. 2017;12:e0182308.

76. van der Buit J, Jones DR, Shah ZH, Halstead JR, Shah ZH, Halstead JR, McKeown S, Mohammed S, Moore S, Robinson J, et al. Worldwide frequency distribution of the ‘Gait keeper’ mutation in the DMRT3 gene. Anim Genet. 2014;45:274–82.

77. Jürgens K, Kasans N, Anderson LS, Lindgren G. Gaiteedness is associated with the DMRT3 Gait keeper mutation in Morgan and American curly horses. Anim Genet. 2014;45:808–9.

78. Jürgens K, Kast H, Holm N, Imsland F, Amason T, Andersonsson L, Anderson LS, Lindgren G. The effect of the DMRT3 Gait keeper mutation on riding traits and gaits in Standardbred and Icelandic horses. Livest Sci. 2015;176:33–9.

79. Regatieri IC, Eberth JE, Sarver F, Lear TL, Bailey E. Comparison of DMRT3 genotypes among American Saddlebred horses with reference to gait. Anim Genet. 2016;47:603–5.

80. Promovina M, Anderson LS, Jurut P, Renedo MC, Reissmann M, Tozaki T, et al. Worldwide frequency distribution of the ‘Gait keeper’ mutation in the DMRT3 gene. Anim Genet. 2014;45:274–82.

81. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. JAMA. 2009;302:168–78.

82. Netherton JJ, Jimmo SL, Palmer D, Tilley DG, Dunkerley HA, Raymond DR, Russell JC, Asbher PM, Sage EH, Vernon RB, Maurice DH. Altered phosphodiesterase 3-mediating AMP hydrolysis contributes to a hypermetabolic phenotype in obese JCR:LA-cp rat aortic vascular smooth muscle cells: implications for diabetes-associated cardiovascular disease. Diabetes. 2002;51:1194–200.

83. Boda H, Uchida H, Takao K, Ouchi Y, Fujita N, Kuno A, Hata T, Nagatani A, Funamoto Y, Miyata M, Yoshikawa T, Inagaki H. A PDE3A mutation in familial hypertension and brachydactyly syndrome. J Hum Genet. 2016;61:701–3.

84. Campos-Chillon F, Farmerie TA, Bouna GI, Clay CM, Camavele EM. Effects of aging on gene expression and mitochondrial DNA in the equine ococyte and follicle cells. Reprod Fertil Dev. 2015;27:925–33.

85. Hamilton MJ, Halvorsen EC, LeParde ND, Bosiljic M, Ho HW, Lam V, Baníñí J, Bennewit KL, Krystal G. SHR presses lung inflammation and inhibits marmoset tumor metastasis in BALB/c mice. Oncotarget. 2016;7:3677–91.

86. Somasundaram R, Fernandes S, Deering JJ, de Haar C, Kuipers EJ, Vogelar L, Middleton FA, van der Woude CJ, Peppelenbosch MP, Kerr WG, Fueller GM. Analysis of SHP1 expression and activity in Crohn’s disease patients. PLoS One. 2017;12:e0182308.

87. van der Buit J, Jones DR, Shah ZH, Halstead JR, Shah ZH, Halstead JR, McKeown S, Mohammed S, Moore S, Robinson J, et al. Worldwide frequency distribution of the ‘Gait keeper’ mutation in the DMRT3 gene. Anim Genet. 2014;45:274–82.

88. Rakhmanov M, Sic H, Kienzler AK, Fischer B, Rizzi M, Seidl M, et al. High activity disorder. Brain Imaging Behav. 2017; https://doi.org/10.1007/s11682-017-9690-2.