Analysis of genetic variation contributing to measured speed in Thoroughbreds identifies genomic regions involved in the transcriptional response to exercise

9 authors, including:

Gabriella Farries
University College Dublin
12 PUBLICATIONS   **121** CITATIONS

SEE PROFILE

Katie F Gough
University College Dublin
32 PUBLICATIONS   **193** CITATIONS

SEE PROFILE

Andrew C Parnell
Maynooth University
128 PUBLICATIONS   **7,377** CITATIONS

SEE PROFILE

Beatrice A Mcgivney
University College Dublin
55 PUBLICATIONS   **1,173** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Genome-Wide microRNA Binding Site Variation between Extinct Wild Aurochs and Modern Cattle Identifies Candidate microRNA-Regulated Domestication Genes View project

New models and methods for stable isotope ecology View project

Analysis of genetic variation contributing to measured speed in Thoroughbreds identifies genomic regions involved in the transcriptional response to exercise

G. Farries* [iD], K. F. Gough*, A. C. Parnell[†], B. A. McGivney*[‡], C. L. McGivney*, P. A. McGettigan*, D. E. MacHugh*[§], L. M. Katz[¶] and E. W. Hill*[‡]

*UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin D04 V1W8, Ireland. [†]Insight Centre for Data Analytics, Hamilton Institute, Maynooth University, Kildare W23 F2H6, Ireland. [‡]Plusvital Ltd, Dun Laoghaire Industrial Estate, Pottery Road, Dublin A96 KW29, Ireland. [§]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin D04 V1W8, Ireland. [¶]UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin D04 V1W8, Ireland.

Summary

Despite strong selection for athletic traits in Thoroughbred horses, there is marked variation in speed and aptitude for racing performance within the breed. Using global positioning system monitoring during exercise training, we measured speed variables and temporal changes in speed with age to derive phenotypes for GWAS. The aim of the study was to test the hypothesis that genetic variation contributes to variation in end-point physiological traits, in this case galloping speed measured during field exercise tests. Standardisation of field-measured phenotypes was attempted by assessing horses exercised on the same gallop track and managed under similar conditions by a single trainer. PCA of six key speed indices captured 73.9% of the variation with principal component 1 (PC1). Verifying the utility of the phenotype, we observed that PC1 (median) in 2-year-old horses was significantly different among elite, non-elite and unraced horses ($P < 0.001$) and the temporal change with age in PC1 varied among horses with different *myostatin* (*MSTN*) g.66493737C>T SNP genotypes. A GWAS for PC1 in 2-year-old horses ($n = 122$) identified four SNPs reaching the suggestive threshold for association ($P < 4.80 \times 10^{-5}$), defining a 1.09 Mb candidate region on ECA8 containing the *myosin XVIIIB* (*MYO18B*) gene. In a GWAS for temporal change in PC1 with age ($n = 168$), five SNPs reached the suggestive threshold for association and defined candidate regions on ECA2 and ECA11. Both regions contained genes that are significantly differentially expressed in equine skeletal muscle in response to acute exercise and training stimuli, including *MYO18A*. As *MYO18A* plays a regulatory role in the skeletal muscle response to exercise, the identified genomic variation proximal to the myosin family genes may be important for the regulation of the response to exercise and training.

Keywords exercise, functional genomics, GWAS, global positioning system, horse, SNP, skeletal muscle

Introduction

Over 300 years of strong artificial selection for athletic performance in the Thoroughbred horse has resulted in a high level of functional adaptation to exercise (Constantinopol *et al.* 1989; Evans *et al.* 1993). The aerobic capacity of horses is superior to that of other domestic species of similar size (Jones *et al.* 1989; Jones & Lindstedt 1993). The large aerobic capacity of a Thoroughbred is achieved through adaptations within the skeletal muscle, respiratory and cardiovascular systems allowing efficient delivery and utilisation of oxygen (Constantinopol *et al.* 1989). In terms of power output, Thoroughbreds have a much greater proportion of muscle mass to body weight (~55%) compared with less athletic horse breeds (~42%) and most other mammalian species (30–40%; Gunn 1987). A small founding population, in which it is estimated as few as 30 horses contributed over 75% of the genetic variation,

Address for correspondence

E. W. Hill, UCD School of Agriculture and Food Science, University College Dublin, Belfield, Dublin D04 V1W8, Ireland.
E-mail: emmeline.hill@ucd.ie

coupled with restricted external gene flow, has led to a small effective population size ($N_e \approx 100$) and a high inbreeding coefficient (0.12; Cunningham *et al.* 2001; Corbin *et al.* 2010). Changes in breeding practices, such as the shuttling of stallions across geographic regions, have led to an increase in inbreeding in the Thoroughbred over time (Binns *et al.* 2012). Despite this apparent lack of genetic diversity and high level of adaptation to exercise, there remains significant variation in terms of speed and aptitude for racing in the Thoroughbred population.

The genetic contribution to horseracing performance has been an area of scientific interest for some time (Field & Cunningham 1976; Gaffney & Cunningham 1988). The heritability of racing performance has previously been estimated from racing times (Oki *et al.* 1995; Mota *et al.* 2005), race earnings (Tozaki *et al.* 2012; Velie *et al.* 2015) and performance ratings (Gaffney & Cunningham 1988; Tozaki *et al.* 2012). The heritability of racing times is highly dependent on the race distance, with estimates higher for shorter distance (0.25–0.29) than longer distance races (0.05–0.08; Oki *et al.* 1995; Mota *et al.* 2005).

Whereas racetrack performance is the ultimate purpose for the preparation and training of racehorses, using racing phenotypes to understand the genetics of performance may be complicated by the disproportionate influence of environmental factors, such as race tactics, surface conditions and jockey. In order to reduce the influence of these external factors, experimental studies of physiological parameters such as maximal oxygen consumption (VO_{2max}) have been undertaken on high-speed treadmills, whereby highly controlled measurements can be taken (Allen *et al.* 2016). However, performing standardised exercise tests on a high-speed treadmill presents difficulties in the context of active racing Thoroughbreds and does not reproduce conditions normally encountered during training. In order to increase sample sizes and better capture the normal training environment, field-based exercise measurements have been developed. For example, the use of global positioning system (GPS) and simultaneous heart rate (HR) recording has been used to quantify exercise load and performance during training (Gramkow & Evans 2006; Kingston *et al.* 2006; Vermeulen & Evans 2006; Fonseca *et al.* 2010). The two measures most commonly derived from these data are the velocity at which HR is 200 beats per minute (bpm; V_{200}) and the velocity at which HR reaches its maximum (V_{HRmax}). V_{HRmax} has been associated with increased earnings per racing start (Gramkow & Evans 2006), with both V_{200} and V_{HRmax} shown to increase after a period of training (Kobayashi *et al.* 1999; Vermeulen & Evans 2006). Similarly, indices derived from HR measurements and blood lactate concentrations in standardised treadmill and field exercise tests have shown these measurements to correlate with racetrack performance (Evans *et al.* 1993; Harkins *et al.* 1993; Gramkow & Evans 2006; Vermeulen & Evans 2006). Therefore, the use of measured

physiological parameters in a field setting may provide a more refined approach to measuring performance capacity in racehorses.

Exercise performance is dependent on the optimisation of central and peripheral components that may be improved with training. Exercise training results initially in central adaptations such as increased stroke volume and cardiac output (Evans & Rose 1988). Peripheral adaptations such as increased mitochondrial biogenesis (Adhihetty *et al.* 2003), fat utilisation (Holloszy *et al.* 1977), protein synthesis (Yarasheski *et al.* 1993; Hartman *et al.* 2006), muscle hypertrophy (Foreman *et al.* 1990) and muscle fibre type switching occur later (Serrano *et al.* 2000; Rivero *et al.* 2007). Numerous equine studies have confirmed an increase in VO_{2max} and oxidative enzyme activity following training (Roneus *et al.* 1992; Roneus 1993; Katz *et al.* 1999; Serrano *et al.* 2000; Hinchcliff *et al.* 2002; McGowan *et al.* 2002), which are probably achieved by increased mitochondrial volume and improved metabolic and mechanical efficiency of the muscle. Anaerobic capacity, speed and force production have been observed to increase following high-intensity (~80–150% VO_{2max}) exercise training (Hinchcliff *et al.* 2002; Rivero *et al.* 2002; Yamano *et al.* 2002; Eto *et al.* 2004).

The racing speed of Thoroughbreds increases with age until approximately 4–5 years of age (Gramm & Marksteiner 2010; Takahashi 2015). In American Thoroughbreds, racing speed measured using Beyer speed figures (Beyer 1993) was found to peak at 4.45 years old (Gramm & Marksteiner 2010). In a much larger cohort of Japanese Thoroughbreds ($n = 17\,071$–$46\,251$ horses, depending on race distance category), average racing speed (metres per second) was used to examine how speed changed with age and throughout each racing season (Takahashi 2015). Average racing speed increased with age for all distance categories, with the observation maintained despite an increase in the average handicap weight carried. Average racing speed plateaued at the beginning of the 4-year-old season. Although previous studies have examined how racing speed changes on a population level, changes in racing speed with age on an individual basis have not been reported.

Data derived from measured speed indices have yet to be fully utilised for genetic analyses. In a cohort of racing Thoroughbreds, speed measured by GPS was recorded and used in a candidate gene study at the *myostatin* (*MSTN*) gene (Hill *et al.* 2012). Variation at the *MSTN* gene is associated with race distance aptitude (Hill *et al.* 2010a,b, 2019), body mass to height ratio (Tozaki *et al.* 2011), muscle fibre-type proportions (Petersen *et al.* 2014) and precocity (Farries *et al.* 2018) in Thoroughbreds. Most of these traits were associated with the g.66493737C>T SNP within intron 1 of *MSTN*. This SNP has been shown to tag an ERE-1 SINE insertion polymorphism within the promotor region, which alters the transcription start site of *MSTN*,

which has been identified as the functional variant (Rooney *et al.* 2018). In terms of GPS measured speed, using five measured speed variables, speed was significantly greater for homozygous C/C horses compared with C/T and T/T horses ($P < 0.05$) genotyped for the g.66493737C>T SNP (Hill *et al.* 2012). However, whereas genetic variants have been shown to contribute to differences in measured speed, to date, no genome-wide investigation of heritable variation in measured speed during training has been reported to identify other genetic contributions to the trait.

The aim of this study was to use speed variables measured in a field-based setting as phenotypes for GWAS to test the hypothesis that there is underlying genetic variation in the Thoroughbred contributing to observed variation in speed during high-intensity exercise. As racing speed increases with age, we hypothesised that measured speed during training would also increase with age and that there are heritable components contributing to variation in this change.

Methods

Ethics statement

University College Dublin Animal Research Ethics Committee approval (AREC-P-12-55-Hill), a licence from the Department of Health (B100/3525) and informed owner consent were obtained for all horses.

Cohort

Data were recorded from 2900 GPS recordings for $n = 294$ horses between 2013 and 2016. Of these, $n = 191$ horses ($n = 104$ male, $n = 87$ female) met the criteria for at least one of the phenotypes used for GWAS. All horses were born between 2011 and 2014 and were maintained at a single flat racing training yard under the supervision of a single trainer. Training, management and nutritional regimens were similar for each horse. Horses used in GWAS were the progeny of 21 stallions, with four of the stallions siring over 20 horses each within the study cohort.

Training

Training commenced with the introduction of the horse to being ridden, a process known as 'breaking'. In general, for the study cohort, once accustomed to being ridden by a jockey, a period of preparatory training commenced, comprising a 600–1000 m walk, trot and slow canter undertaken six times per week. After this period, submaximal training commenced. Each submaximal exercise bout consisted of a warm-up where horses walked for 40–70 min prior to the jockey mounting, then walked a further 300 m mounted, followed by a 700 m trot and

slow canter down the incline of the track. This was followed by an 800–1000 m brisk canter up the incline of the track. After a period of submaximal training, at the discretion of the trainer, high-intensity sprint exercise bouts (work days) were incorporated into the submaximal training regime.

Work day

Work days (WD) were performed on a woodchip, 1500 m, uphill, all-weather gallop track, with the final 800 m straight set on a 2.7% incline (Fonseca *et al.* 2010).

On each work day, horses were walked on a horse walker for 30–60 min, followed by 5–10 min of walking in hand. Warm-up under saddle consisted of a 300 m walk followed by a 700 m trot and slow canter down the incline of the track. A short period of walk followed. The sprint portion of the work day consisted of the horses galloping at high intensity for 800–1000 m.

Experimental measurements

Velocity and distance were measured using a STATSports Viper Pod GPS monitoring system (STATSports Technologies Ltd.). The raw GPS data were smoothed using kernel smoothing implemented through '*ksmooth*' within the R environment (version: 3.4.1; R Core Team 2017). Speed indices originally described by Fonseca *et al.* (2010) were derived from the smoothed GPS output (Fig. 1). The speed indices (Table 1) were measured during the sprint portion of the work day, defined as from when the horse first exceeded 5 m/s until it reached peak velocity (V_{peak}).

Correlations among speed indices were determined using Pearson's correlation within R. PCA was performed using V_{peak}, acceleration (Acc), average sprint velocity (aveSpr), distance covered in the 6 s following V_{peak} (Dist6b), distance covered in the 6 s preceding V_{peak} (Dist6b) and distance covered in the 6 s before and after reaching V_{peak} (Dist6) as input variables, using '*princomp*' within the R environment (Fig. 2).

Table 1 Description of speed indices derived from raw global positioning system data.

| Speed index	Definition
HR_{peak}	Peak heart rate (bpm)
V_{peak}	Peak velocity (m/s)
Acc	Time taken (s) from when the horse first exceeded 5 m/s in the sprint period until V_{peak} was reached
aveSpr	Average velocity (m/s) during the sprint period
Dist6a	Distance (m) covered in the 6 s post-V_{peak}
Dist6b	Distance (m) covered in the 6 s preceding V_{peak}
Dist6	Distance (m) covered in the 6 s before and after reaching V_{peak}
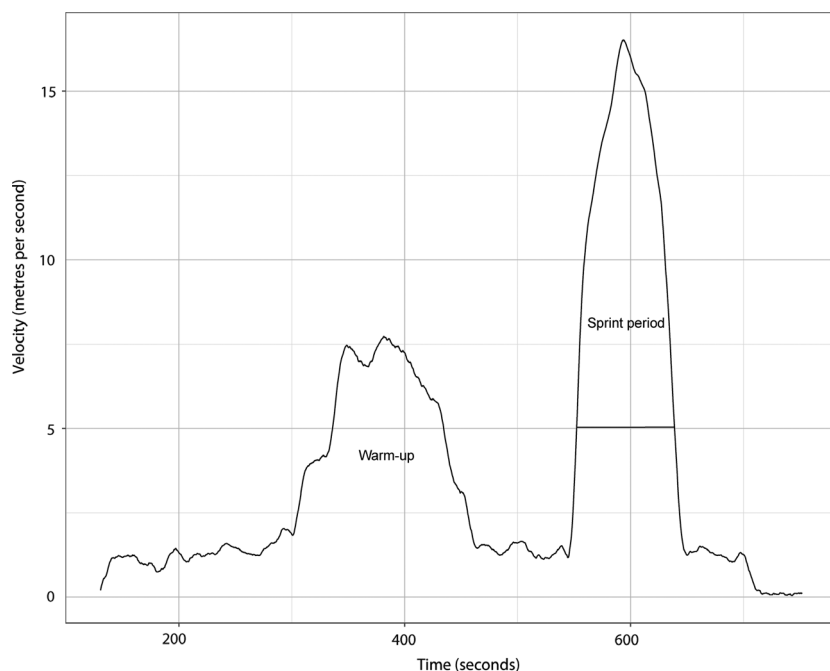
Figure 1 Example of smoothed global positioning system output from a recorded work day.

Metadata

Track condition was assessed by a single observer and graded from 1 (firmest) to 4 (softest). Although track condition was measured, it was not included in statistical modelling and used solely for descriptive purposes. For each recorded work day, the number of previous cumulative work days was also calculated, as well as the age in days of the horse on the day of recording.

For a subset of horses, whole blood was collected at rest and 5 min after the cessation of exercise into fluoride oxalate tubes. Blood samples were centrifuged and plasma lactate concentration measured using a YSI2300 STAT PLUS on-site bench-top autoanalyser (YSI UK Ltd). The plasma lactate concentrations were used to verify and describe the intensity of the exercise bouts performed.

†Vpeak: peak velocity, aveSpr: average sprint velocity, Dist6a: distance covered in 6s after peak velocity, Dist6b: distance covered in 6s before peak velocity and Dist6: distance covered in the 6s before and after peak velocity
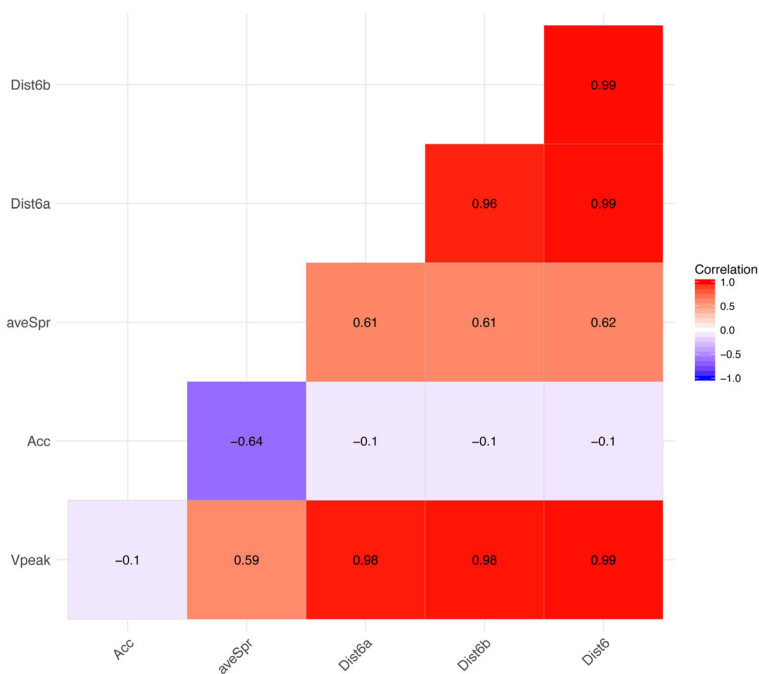
Figure 2 Correlations between speed indices included in the PCA.

Race records were collated for horses born between 2011 and 2013. Horses were categorised as elite (Thoroughbred elite; TBE) if they had won or placed at Listed or Group level, and other (Thoroughbred other; TBO) if they had competed in a race but never won or placed at Listed or Group level (these horses may have won or placed at non-Listed- or Group-level races). Horses that had never raced were categorised as unraced (UNR). The median value for PC1 was derived for 2-year-old horses with more than three GPS recordings, and compared among TBE, TBO and UNR horses using a one-way ANOVA. A linear model was used to examine the change in PC1 with age and cumulative WDs, as well as the change in PC1 with age and cumulative WDs among *MSTN* g.66493737C>T SNP genotypes.

Phenotypes

Two-year-old speed (2yo_PC1)

Principal component 1 median values were calculated from GPS data recorded during the 2-year-old year of each horse (2yo_PC1), for those horses with more than three recordings during this period.

Modelling of speed with age (ΔPC1)

The change in PC1 with age (ΔPC1) was modelled using a robust linear model (to reduce the impact of outliers) within the MASS package (version 7.3-47) (Venables & Ripley 2010). Horses had more than three recordings spanning a period of 3 months or greater. All horses were aged between 1.55 and 5 years of age.

Genotyping

Genomic DNA was extracted from whole blood using the Maxwell 16 automated DNA purification system (Promega). Horses were genotyped using medium- and high-density SNP genotyping arrays: Illumina Equine SNP70 BeadChip (Illumina), $n = 47$ and Axiom Equine Genotyping Array (Axiom MNEC670) (Affymetrix), $n = 144$. Concordant SNPs derived from the SNP70 and SNP670 arrays were used for the analysis. Individuals and SNPs were subject to a genotyping threshold of 95%. SNPs missing within an individual sample but passing the 95% call rate threshold were imputed using BEAGLE (version: 3.3.2) (Browning & Browning 2013) using a dataset of 1677 Thoroughbred horses genotyped on the SNP70 array in a separate study (Hill *et al.* 2019). A genetic sex check and minor allele frequency threshold of >0.05 were also included as quality control. Previously 10 horses were genotyped on both the SNP70 and SNP670 array and post-imputation concordance was found to be >99% (Farries *et al.* 2018). After quality control, 49 270 SNPs were suitable for downstream analyses.

All horses were genotyped for the *MSTN* g.66493737C>T SNP using a custom Taqman® assay (Life Technologies) and the genotype was included for the genomic analyses, resulting in a total of 49 271 SNPs.

A genomic relatedness matrix (GRM) among horses was constructed and PCA carried out using GCTA (version 1.42.2) (Yang *et al.* 2011). Results were visualised using GGPLOT2 (Wickham 2009).

Tests of genome-wide association

GWASs were performed using a mixed-model association test in GENABEL (version 1.8-0) (Aulchenko *et al.* 2007). Sex was included as a covariate. GWAS results were visualised with Manhattan plots generated using QQMAN (version 0.1.2) (Turner 2018). The threshold for genome-wide significance was determined using the Bonferroni correction based on the effective number of independent loci (M_e) using the GENETIC ERROR CALCULATOR (version 0.2) (Li *et al.* 2012), with the threshold for genome-wide significance set at $\frac{0.05}{M_e}$.

LD between associated SNPs and surrounding SNPs was determined using PLINK (version 1.09) (Purcell *et al.*, 2007). Haplotypes were also called using PLINK with a maximum pairwise distance of 20 Mb based on the work of Corbin *et al.* (2010). Candidate regions were mined for genes using the EquCab2 genome annotation and Ensembl Biomart (Wade *et al.* 2009; Smedley *et al.* 2015). Candidate genes were queried against gene sets identified by Bryan *et al.* (2017) to be differentially expressed in Thoroughbred skeletal muscle (gluteus medius) in response to an acute bout of sprint exercise (a work day; 3241 genes), or a 6 month period of training (3405 genes). The EquCab2 annotation was used to enable integration with the results of Bryan *et al.* (2017).

Results

Cohort

Owing to the overrepresentation of particular stallions and therefore the presence of a large number of half-siblings, population structure was investigated using PCA of the GRM, which showed a high degree of clustering by sire (Fig. S1).

Based on the $n = 191$ horses used in at least one of the phenotypes for GWAS, the M_e was estimated to be 20 842. This resulted in a genome-wide significance threshold of $P < 2.40 \times 10^{-6}$, and a suggestive threshold of $P < 4.80 \times 10^{-5}$. Of the 191 horses used in either phenotype for GWAS, 99 were included in GWAS for both 2yo_PC1 and ΔPC1. Six of these horses were also included in the transcriptomic investigation by Bryan *et al.* (2017).

Experimental measurements

For the 2900 GPS recordings from 294 horses, there was a mean of 9.9 recordings per horse (range 1–50). The mean

recorded V_{peak} was 16.36 m/s (range 14.23–17.63 m/s), and Dist6a and Dist6b had mean distances of 96.77 m (range 84.16–104.67 m) and 97.18 m (range 84.49–103.88 m) respectively (Table 2). The mean Dist6ab was 193.9 m (range 169.09–208.55). Mean resting and peak plasma lactate concentrations were 0.70 mM (range 0.12–4.54 mM) and 28.40 mM (range 19.15–42.40) respectively for 176 paired samples taken from 100 horses (Table 2). Track condition was recorded for 2594 of the WDs, with the majority taking place on a going with a score of 2 (Table S1). Of the 2900 GPS recordings, 1056 had simultaneous heart rate recordings, with a mean HR_{peak} of 219 bpm (range 182–237 bpm; Table 2).

Principal component 1 from the PCA of V_{peak}, Acc, aveSpr, Dist6a, Dist6b and Dist6 explained 73.9% of the variation in these speed indices (Fig. 3). PC1 was largely determined by V_{peak}, Dist6a, Dist6b and Dist6. PC2 explained 22.0% of the variance and was largely weighted by Acc. PC1 and PC2 collectively explained 95.9% of the variance in the six speed indices.

Comparison of elite performance status for horses phenotyped for 2yo_PC1 showed a significant difference in 2yo_PC1 between TBE ($n = 21$), TBO ($n = 67$) and UNR ($n = 14$) horses (Fig. 4). TBE horses had significantly higher 2yo_PC1 compared with TBO ($P_{FDR} = 6.97 \times 10^{-3}$) and UNR ($P_{FDR} = 3.94 \times 10^{-4}$) horses. The TBO cohort also had significantly higher 2yo_PC1 than UNR ($P_{FDR} = 0.02$). Horses within the TBE cohort had a mean of 21.67 recordings (range 8–50), TBO 15.19 recordings (range 4–38) and UNR 7.69 recordings (range 4–14). There was a significant difference in the number of recordings between TBE, TBO and UNR horses, with TBE horses having a

greater number of recordings than TBO and UNR horses ($P = 3.10 \times 10^{-3}$ and $P = 4.00 \times 10^{-6}$).

Of the 294 horses with a GPS recording, 212 had more than three recordings and were genotyped for the *MSTN* g.66493737C>T SNP ($n = 81$ C/C, $n = 117$ C/T and $n = 14$ T/T). These 212 horses were used to discern the relationships between PC1, age, cumulative work days and the *MSTN* g.66493737C>T SNP using a linear model. There was a significant relationship between PC1 and age ($P = 1.37 \times 10^{-8}$). An investigation of the relationship between PC1, age and *MSTN* g.66493737C>T SNP genotype revealed a significant difference in △PC1 among *MSTN* genotypes ($P = 6.48 \times 10^{-7}$; Fig. 5). This difference was seen between C/C and C/T horses ($P = 0.03$) but not C/C and T/T horses ($P = 0.47$). A significant association between PC1 and number of cumulative WDs was also observed ($P = 7.09 \times 10^{-11}$). A significant difference in the interaction of cumulative WDs with PC1 was observed between C/C and T/T horses ($P = 7.05 \times 10^{-3}$), but not C/T and T/T horses ($P = 0.05$; Fig. S2).

Tests of genome-wide association

In a GWAS using 2yo_PC1 as the phenotype, no SNP reached genome-wide significance ($P < 2.40 \times 10^{-6}$); however, four SNPs on ECA8 reached the suggestive threshold of $P < 4.80 \times 10^{-5}$ and defined a 1.09 Mb region (Fig. 6, Table 3). The highest ranked SNP g.10114168C>T ($P = 4.28 \times 10^{-6}$) approached the genome-wide significance threshold (Table 3). The closest gene to the g.10114168C>T SNP was *MYO18B*, with the SNP located 67 kb downstream of the transcription start site (TSS) (Table 3). This SNP was part of a 69.9 kb, 3 SNP haplotype spanning g.10044262–10114168 (Table S1). The third ranked SNP g.9023079G>A ($P = 2.94 \times 10^{-5}$) within the same candidate region was located 278 kb downstream of the *MN1 proto-oncogene* (*MN1*) gene (Table 3). This SNP was upstream of a 170.5 kb haplotype spanning g.9067048–9237581 (Table S1).

No SNP reached genome-wide significance in the GWAS for △PC1 ($P < 2.40 \times 10^{-6}$); however, five SNPs reached the suggestive threshold ($P < 4.80 \times 10^{-5}$), defining two candidate regions on ECA2 and ECA11 (Fig. 7). Three SNPs on ECA11 defined a 1.82 Mb region containing 24 genes (Table 4). The third ranked SNP in the GWAS for △PC1 was g.42938387G>A, located within the last intron of the *MYO18A* gene, which was in high LD with the top-ranked g.42471888C>T SNP ($r^2 = 0.72$) located 10.3 Mb downstream of *FOXN1* (Table 4). The g.42938387G>A SNP was part of a 10 SNP haplotype spanning 236.7 kb (g.42738710–42975477; Table S2). The top-ranked SNP, g.42471888C>T, was 32.9 kb upstream of a 62.2 kb haplotype comprising two SNPs (Tables S2 and 5).

Two SNPs on ECA2 reached the suggestive threshold for association with △PC1 (Table 5). These SNPs were located

Table 2 Mean experimental measurement values.[1]

Speed index	Mean	Standard deviation	Range
Peak velocity (V_{peak})	16.35 m/s	0.42 m/s	14.23–17.63 m/s
Acceleration time (Acc)	40.06 s	7.99 s	21.00–89.30 s
Average sprint velocity (aveSpr)	13.35 m/s	0.40 m/s	10.88–14.54 m/s
Distance covered in 6 s after peak velocity (Dist6a)	96.77 m	2.41 m	84.16–104.67 m
Distance covered in 6 s before peak velocity (Dist6b)	97.18 m	2.47 m	84.49–103.88 m
Distance covered in the 6 s before and after peak velocity (Dist6)	193.79 m	4.83 m	169.1–208.6 m
HR_{peak}	219 bpm	7.71 bpm	182–237 bpm
Resting plasma lactate concentration	0.70 mM	0.46 mM	0.12–4.54 mM
Peak plasma lactate concentration	28.40 mM	4.26 mM	19.15–42.40 mM

[1]Speed index results derived from 2900 GPS recordings from 294 horses. Peak heart rate (HR_{peak}) values from 1056 heart rate recordings. Plasma lactate concentrations taken from 176 paired samples.
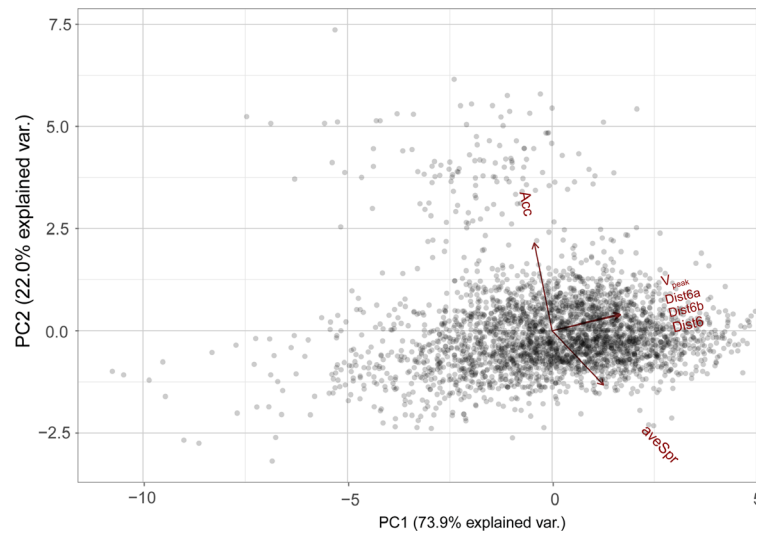
Figure 3 Biplot of the principal component analysis of peak velocity (V_{peak}), acceleration time (Acc) and distance covered in the 6 s preceding and proceeding V_{peak} (Dist6a, Dist6b and Dist6) measurements.
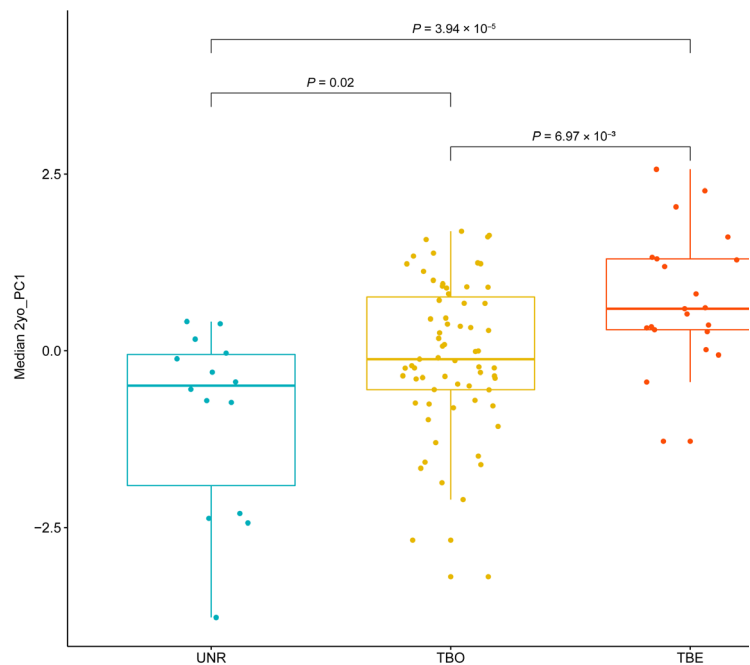

Figure 4 Comparison of median value for principal component 1 in 2-year old-elite, non-elite and unraced horses.

[1]Horses had >3 recordings. Elite (TBE, placed at Listed or Group level, $n = 21$), non-elite (TBO, unplaced at Listed or Group level, $n = 67$) and unraced (UNR, never had a racing s

30 kb apart, with no other SNPs between them. Both of these SNPs were within the *sodium channel and clathrin linker 1* (*SCLT1*) gene. The g.100260390T>C SNP was located at the end of a 382.0 kb haplotype spanning g.99878345–100260390 (8 SNPs).

Discussion

For the first time, we have applied GPS monitoring technology in horses to generate a composite speed index for use in a GWAS. All recordings took place on a work day as this was the highest-intensity exercise regularly undertaken by the horses that could be measured. This was reflected in our experimental measurements with horses achieving a mean V_{peak} of 16.36 m/s, mean HR_{peak} of 217 bpm and mean peak plasma lactate concentration of 29.92 mM (Table 2). Using PCA, we summarised 73.9% of the variation in five key speed indices used previously to quantify speed during training by a single measure, PC1 (Fonseca *et al.* 2010; Hill *et al.* 2012). 2yo_PC1 was significantly different among TBE, TBO and UNR horses (Fig. 4), validating this index as indicative of racing performance ability. However, it should be noted that there was a sampling bias with regard to TBE horses having greater numbers of recordings compared with TBO and UNR horses.
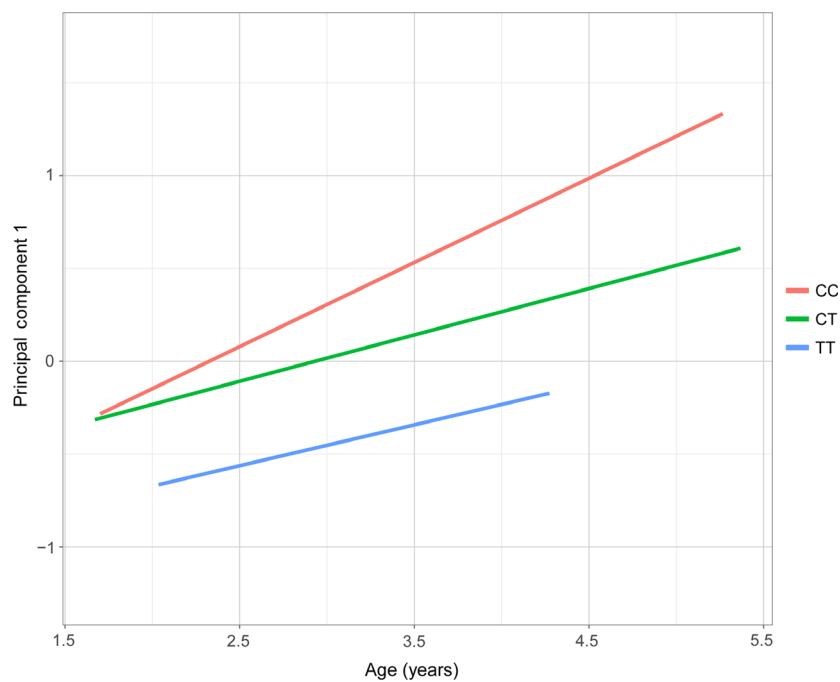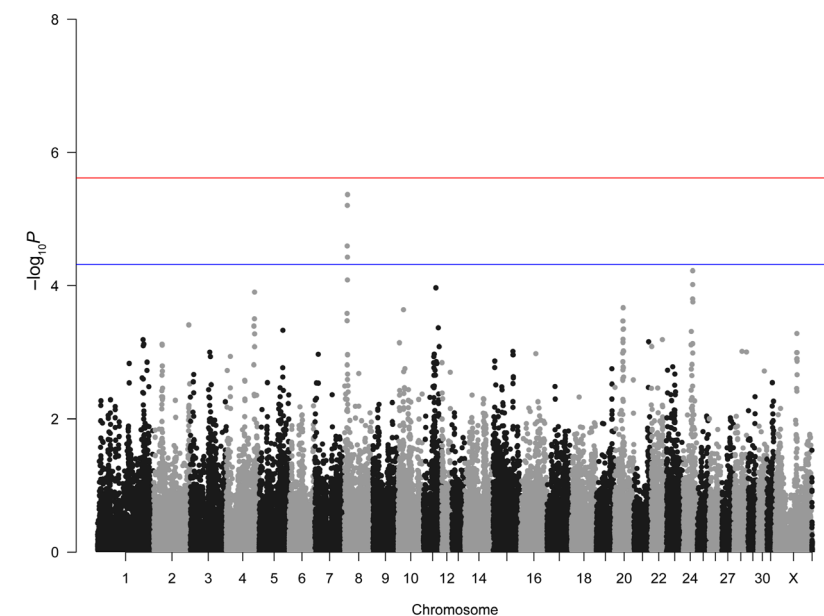
Figure 5 Linear model of measured speed (using principal component 1) with age (years) across the three *MSTN* g.66493737C>T SNP genotypes.

¹Using mixed model association in *n* = 122 Thoroughbreds. Red line indicates threshold for genome-wide significance (*P* < 2.40 × 10⁻⁶), blue line indicates suggestive threshold for association (*P* < 4.80 × 10⁻⁵).

Figure 6 Manhattan plot of association for median recorded 2-year-old speed, as quantified by principal component 1.

As a preliminary test, the relationship between age and speed was investigated among *MSTN* g.66493737C>T SNP genotypes. Although C/C and C/T horses started at a relatively similar speed and age, the rate of improvement in speed (PC1) with age for C/C horses was significantly greater than for C/T horses ($P = 0.03$). In comparison, T/T horses showed much lower initial speed, but the rate of improvement in speed was not significantly different from C/C or C/T horses ($P = 0.47$). Changes in speed with cumulative work days were also modelled, and a similar pattern of change to that observed for modelling speed with age was observed for C/C and C/T horses. Although T/T horses had lower speed on early work days, they had a much more rapid increase in speed with cumulative work days compared with C/C and C/T horses (Fig. S2). The observation of lower speed among T/T on early WDs is probably due to their being adapted to longer-distance exercise and having lower aptitude for speed, as was previously observed by Hill *et al*. (2012), showing T/T horses to have significantly lower recorded speed compared

Table 3 GWAS results for the candidate region identified on ECA8 for median 2-year-old speed (as measured by principal component 1).[1]

GWAS result		Nearest gene				
SNP position	P	Gene	TSS	Gene end	LD with top SNP (r^2)	Distance from TSS (kb)
9023079	2.94×10^{-5}	MN1	8744909	8786454	0.42	−278.17
9067048	0.25	MN1	8744909	8786454	0.04	−322.14
9155934	0.05	MN1	8744909	8786454	0.09	−411.03
9186481	0.05	MN1	8744909	8786454	0.06	−441.57
9194496	0.11	MN1	8744909	8786454	0.06	−449.59
9213062	0.03	MN1	8744909	8786454	0.07	−468.15
9237581	0.02	CRYBA4	9694831	9700087	0.07	457.25
9304992	7.69×10^{-3}	CRYBA4	9694831	9700087	0.09	389.84
9325552	8.16×10^{-3}	CRYBA4	9694831	9700087	0.10	369.28
9374074	0.31	CRYBA4	9694831	9700087	0.02	320.76
9379759	0.25	CRYBA4	9694831	9700087	0.03	315.07
9430846	0.85	CRYBA4	9694831	9700087	0.05	263.99
9466731	7.36×10^{-6}	CRYBA4	9694831	9700087	0.60	228.10
9597547	7.64×10^{-3}	CRYBA4	9694831	9700087	0.20	97.28
9606317	0.32	CRYBA4	9694831	9700087	0.01	88.51
9657903	0.06	CRYBA4	9694831	9700087	0.02	36.93
9736050	9.31×10^{-5}	TPST2	9762022	9775088	0.71	25.97
9738714	4.27×10^{-5}	TPST2	9762022	9775088	0.82	23.31
9759042	0.75	TPST2	9762022	9775088	0.02	2.98
9810461	2.78×10^{-3}	HPS4	9810007	9834106	0.16	−0.45
9870854	0.75	SEZ6L	9888560	9972812	0.02	17.71
9894359	0.03	SEZ6L	9888560	9972812	0.09	−5.80
9910682	0.03	SEZ6L	9888560	9972812	0.09	−22.12
9913455	0.75	SEZ6L	9888560	9972812	0.02	−24.90
10044262	1.20×10^{-3}	MYO18B	10181087	10411809	0.55	136.83
10093911	1.18×10^{-3}	MYO18B	10181087	10411809	0.53	87.18
10114168	5.08×10^{-6}	MYO18B	10181087	10411809	1.00	66.92

[1]Nearest gene to GWAS SNP as determined by distance from transcription start site (TSS). SNP highlighted in bold font reached the suggestive threshold of $P < 4.80 \times 10^{-5}$.
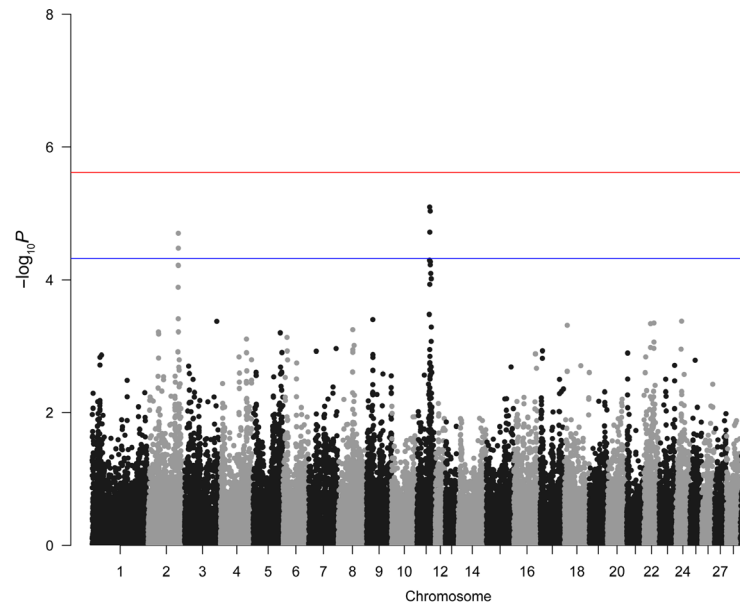

Figure 7 Manhattan plot of association for the change in measured speed (principal component 1) with age.

[1]Using mixed model association in $n = 168$ horses. Red line indicates threshold for genome-wide significance ($P < 2.40 \times 10^{-6}$), blue line indicates suggestive threshold for association ($P < 4.80 \times 10^{-5}$)

with C/C individuals. The steeper increase in speed with cumulative WDs among T/T horses may be a consequence of commencing WDs at a later age (Fig. 4) and participating in fewer WDs over a given period compared with C/C and C/T horses (Fig. S2). However, it may also reflect a greater response in speed to a given amount of training. To

Table 4 GWAS results for the candidate region identified on ECA11 for change in principal component 1 with age.[1]

SNP position	P	Gene	LD with top SNP (r^2)	Distance from TSS (kb)	log$_2$FC	P_{FDR}	log$_2$FC	P_{FDR}
42471888	7.97×10^{-6}	FOXN1	1	−10 263				
42504809	6.13×10^{-3}	UNC119	0.29	3374				
42566965	0.08	KIAA0100	0.21	8631				
42666135	1.83×10^{-3}	TRAF4	0.22	480	−1.06	6.90×10^{-13}	−0.58	8.25×10^{-4}
42683705	1.13×10^{-3}	FAM222B	0.11	10 376				
42738710	6.21×10^{-3}	ERAL1	0.22	−7256				
42746814	0.01	ERAL1	0.21	848				
42786705	1.17×10^{-4}	DHRS13	0.19	411	−0.79	6.78×10^{-9}		
42808655	0.03	PHF12	0.21	15 772				
42866856	0.16	SEZ6	7.19×10^{-5}	32 103				
42868734	0.02	SEZ6	0.17	33 981				
42868931	4.23×10^{-3}	SEZ6	0.21	34 178				
42870756	0.02	SEZ6	0.17	36 003				
42898983	0.1	PIPOX	0.02	−13 181				
42938387	1.91×10^{-5}	MYO18A	0.72	397	−0.37	2.43×10^{-9}	−0.41	2.60×10^{-4}
42975477	0.45	MYO18A	0.21	37 487				
42975749	0.67	MYO18A	4.88×10^{-3}	37 759				
42992877	2.27×10^{-3}	MYO18A	0.25	54 887				
43001964	2.27×10^{-3}	MYO18A	0.25	63 974				
43001984	0.49	MYO18A	0.12	63 994				
43073026	0.2	CRYBA1	2.74×10^{-3}	3251			0.35	4.29×10^{-3}
43079211	7.51×10^{-3}	NUFIP2	0.21	−4281	0.87	2.37×10^{-19}		
43137269	1.77×10^{-3}	NUFIP2	0.51	53 777				
43163727	0.12	TAOK1	3.60×10^{-3}	−77 922				
43220683	0.61	TAOK1	0.02	−20 966				
43277172	0.08	TAOK1	0.21	35 523				
43282773	1.42×10^{-3}	TAOK1	0.22	41 124				
43372894	0.04	ANKRD13B	0.21	−1124				
43410795	0.01	SSH2	0.17	13 572	0.77	7.45×10^{-6}		
43414226	0.3	SSH2	0.11	17 003				
43592444	3.13×10^{-3}	EFCAB5	0.42	−75 031			0.7	0.02
43596583	2.83×10^{-3}	EFCAB5	0.45	−70 892				
43669190	0.3	EFCAB5	0.11	1715				
43740401	0.08	EFCAB5	0.14	72 926				
43795573	0.5	SLC6A4	3.60×10^{-3}	−60 103	1.36	8.68×10^{-7}	1.02	9.44×10^{-3}
43839853	0.17	SLC6A4	7.92×10^{-3}	−15 823				
43874468	0.23	SLC6A4	8.82×10^{-3}	18 792				
43882694	0.06	BLMH	0.13	−16 223	−0.38	8.04×10^{-12}	−0.44	7.76×10^{-5}
43892849	0.02	BLMH	0.2	−6068				
43937486	0.45	TMIGD1	0.12	−24 413				
43992206	0.07	CPD	0.17	−18 977				
44036146	0.08	CPD	0.03	24 963				
44045687	0.01	CPD	0.22	34 504				
44155574	0.72	TUSC5	0.08	−6975				
44182205	0.5	TUSC5	6.23×10^{-3}	19 656				
44232578	0.39	U6	0.04	16 962				
44285589	0.65	ABR	0.02	−31 704				
44292705	9.20×10^{-6}	ABR	0.33	−24 588				

[1]Nearest gene to GWAS SNP as determined by distance from transcription start site (TSS). Genes identified as differentially expressed in response to acute exercise and training stimuli by Bryan *et al*. (2017) were reported with log$_2$ fold change (log$_2$FC) and false discovery rate adjusted *P*-value (P_{FDR}). SNP highlighted in bold font reached the suggestive threshold of $P < 4.80 \times 10^{-5}$.

determine this, the total submaximal training and the cumulative WDs would need to be incorporated to quantify total training load.

The increase in speed with age is probably a combined effect of horses both growing and strengthening with age, as well as a physiological response to training. In human

studies, the responses to given training stimuli have been shown to be highly variable among individuals, and highly heritable (Bouchard *et al.* 2011; Rankinen *et al.* 2012; Rice *et al.* 2012). Much of this work has been performed through the HERITAGE family study, where 551 sedentary individuals from 90 Caucasian families and 40 African-American

Table 5 ECA2 SNPs identified in GWAS for change in principal component 1 with age.[1]

GWAS result		Nearest gene				
SNP position	P	Gene	TSS	Gene end	LD with top SNP (r^2)	Distance from TSS (kb)
100260390	1.99×10^{-5}	*SCLT1*	100184394	100329118	1.00	−76.00
100290657	3.33×10^{-5}	*SCLT1*	100184394	100329118	0.98	−106.26

[1]From $n = 168$ horses. SNPs reached the suggestive threshold for association ($P < 4.80 \times 10^{-5}$). Both SNPs are located within the *SCLT1* gene and distance from the transcription start site (TSS) given.

families underwent a standardised 20-week training programme (Bouchard *et al.* 1995). Physiological measurements were taken before and after the training intervention and used to compare responses between individuals in response to the training. Thus far, heritable contributions to traits such as change in fat-free mass (Rice *et al.* 1997), maximum power output (Rico-Sanz *et al.* 2004) and gains in maximal oxygen uptake (Bouchard *et al.* 2011) have been reported. Consequently, for the present study, we examined how measured speed changed with both age and cumulative WDs. It should be noted that the use of cumulative WDs to measure training does not capture differences in duration of submaximal training or periods of time out of training. Including these additional data may improve the quantification of training load, and thus enable more detailed examination of the training response in individual horses. Likewise, use of concurrent HR monitoring data may be applied to quantify gains in aerobic fitness.

Tests of genome-wide association did not detect SNPs reaching genome-wide significance. However, this is not entirely unexpected for several reasons. Firstly, the small sample size for each phenotype means that there may be limited power to detect significant associations. However, in a study such as this where the horses were under highly controlled environmental conditions (all horses were housed together, trained on the same gallop track and were under similar management), there was a reduction in environmental variation compared with studies that utilise racetrack performance, where there are many external influences such as different trainers, jockeys, tracks, track conditions, race tactics and other horses. There is an inherent trade-off between increasing the sample size and reducing the ability to control environmental factors. Secondly, it is likely that traits such as measured speed and the change in speed with age are highly polygenic. Polygenic traits are influenced by many loci, each with a small effect size contributing to the overall trait variation, many of which would be too small to detect using stringent thresholds for genome-wide significance (Manolio *et al.* 2009). Whereas no individual locus reached genome-wide significance, several SNPs did reach the suggestive threshold for association ($P < 4.80 \times 10^{-5}$) and were used to define candidate genomic regions.

The GWAS for 2yo_PC1 identified a 1.09 Mb candidate region on ECA8 containing six genes (Table 3). The highest

ranked SNP, g.10114168C>T ($P = 4.28 \times 10^{-6}$) was located 67 kb downstream of the *MYO18B* gene. Myosin-18B protein acts as a stabiliser within contractile fibres and during formation of myosin II stacks (Jiu *et al.* 2019). A human *MYO18B* null mutation has been shown to significantly compromise sarcomere assembly and force production within skeletal muscle (Alazami *et al.* 2015; Berger *et al.* 2017). A similar loss of function was reported in cardiomyocytes of mice, with the lack of a functional *Myo18b* gene resulting in malformation of myofibrillar structures (Ajima *et al.* 2008). In a study of the transcriptional response to resistance exercise vs. aerobic exercise in human skeletal muscle, *MYO18B* was significantly differentially expressed in response to resistance exercise compared with aerobic exercise (Dickinson *et al.* 2018). However, *MYO18B* was not detected as differentially expressed in equine skeletal muscle in response to exercise or training (Bryan *et al.* 2017). The exercise test used in Bryan *et al.* (2017) had both resistance and aerobic components, which may explain *MYO18B* not being identified as differentially expressed in response to this exercise test. These results suggest a role for the Myosin-18B protein in skeletal muscle structure, as well as *MYO18B* within the transcriptional response to resistance exercise within muscle.

One of the SNPs reaching the suggestive threshold for association was located 278 kb downstream of the *MN1* gene. *MN1* has mostly been investigated in terms of its role in the pathogenesis of leukaemia in humans (Valk *et al.* 2004; Heuser *et al.* 2006; Carella *et al.* 2007). However, in terms of muscle physiology, *MN1* has been found to be overexpressed in human skeletal muscle compared with other human tissues (GTEx Consortium 2013), and in horses has been shown to be significantly downregulated in response to acute exercise ($\log_2 FC = -0.47$, $P_{FDR} = 4.87 \times 10^{-4}$; Bryan *et al.* 2017), suggesting that *MN1* may be important for skeletal muscle function. Other genes within this ECA8 candidate region include *Crystallin beta A4* (*CRYBA4*) and *Tyrosylprotein sulfotransferase 2* (*TPST2*). Crystallin proteins comprise a major structural component of the vertebrate optic lens, with mutations within *CRYBA4* associated with conditions of the eye such as myopia and cataracts (Billingsley *et al.* 2006; Ho *et al.* 2012). In the context of muscle, crystallin proteins were upregulated in response to mechanical stress in cultured

rabbit myocardium (Bupha-Intr *et al.* 2007). TPST-2 is an integral membrane glycoprotein acting as a tyrosine sulfurase within the trans-Golgi network and ubiquitously expressed across tissues (Beisswanger *et al.* 1998; Ouyang *et al.* 2002; Sasaki *et al.* 2007). A missense mutation within *TPST2* in mice is associated with the 'growth-retarded mouse' phenotype, showing significant dwarfism, impaired thyroid stimulating hormone response and associated thyroid hypoplasia (Sasaki *et al.* 2007).

The GWAS for ΔPC1 identified the 1.82 Mb candidate region on ECA11 containing 24 genes, as defined by three SNPs reaching the suggestive threshold of significance. The strongest association was with the g.42471888T>C SNP, the nearest gene is *Forkhead Box N1* (*FOXN1*) located 10.3 Mb downstream of the SNP (Table 4). *FOXN1* encodes a transcription factor regulating organogenesis of the thymus (Jackson *et al.* 2010). Absence of functional *FOXN1* results in the 'nude' phenotype, where there is lack of hair and immune deficiency owing to the lack of a functional thymus (Golson & Kaestner 2016). The second ranked SNP, g.44292705T>G, was located 24.6 Mb from the *ABR Activator of RhoGEF and GTPase* gene (*ABR*; Table 4). *ABR* interacts with Rac which is a master regulator of immune cell function (Gong *et al.* 2013). Ablation of *ABR* results in increased severity in asthma symptoms in a murine asthma model (Gong *et al.* 2013), increased susceptibility to endotoxemia and increased lung tissue injury after exposure to endotoxin (Cunnick *et al.* 2009).

The third suggestive SNP was located 397 kb from the TSS of the *MYO18A* gene, which is functionally closely related to *MYO18B*. Whereas *MYO18A* is expressed in many tissues, *MYO18B* appears to be predominately expressed in cardiac and skeletal muscle (Liang *et al.* 1999; Ajima *et al.* 2008). *MYO18A* has recently been identified as a key regulatory gene in the transcriptional response of equine skeletal muscle to exercise and training, ranking among the top 30 hub and bottleneck genes in an interaction network analysis using RNA-seq data (Bryan *et al.* 2017). Functional networks within the skeletal muscle transcriptome were reconstructed by combining known protein–protein interactions and co-expression of differentially expressed genes. In resting samples (i.e. no exercise stimulus), *MYO18A* was ranked ninth among all expressed genes for the greatest number of direct interactions with other genes (degree = 5) and number of co-expression interactions (betweenness = 158). There was a similar observation among transcripts derived from samples following exercise; *MYO18A* ranked eighth for degree (7) and tenth for betweenness (608). In the network, *MYO18A* was the centre of a cluster enriched for genes related to actin filament processes, cytoskeleton protein binding, insulin and platelet-derived growth factor signalling (Bryan *et al.* 2017). Platelet-derived growth factor signalling has been shown to play a role in angiogenesis in

skeletal muscle (Hellstrom *et al.* 1999; Naylor *et al.* 2014), with angiogenesis being a key response to training within the muscle (Booth *et al.* 2015). These results highlight *MYO18A* as a key regulator in the coordination of the transcriptional response to exercise and genetic variation proximal to the gene may be responsible for variation in the response.

MYO18A is just one of 11 genes present in the candidate region identified on ECA11 for ΔPC1 that are also transcriptionally responsive in muscle to exercise and training (Table 4). These results suggest that this genomic region is an important transcriptionally active region that responds to exercise and training stimuli in the muscle. Genomic variation in this region may alter the transcriptional response to exercise and training in skeletal muscle, and thus may impact end-point phenotypes relevant to adaptation to training, such as the temporal change in speed. Given that the regulation of gene expression has been proposed as the greatest determinant of phenotype, the identification of genomic loci contributing to variation in expression is particularly important (Gilad *et al.* 2008; Peck *et al.* 2015). Greater understanding of regulation in skeletal muscle gene expression and variation in gene expression between horses may be applied to inform trainers and optimise training regimes for individual horses. More than 85% of loci identified by GWAS are in non-coding regions, and an enrichment of gene expression quantitative trait loci (eQTL) has been identified in GWAS hits (Hindorff *et al.* 2009; Brown *et al.* 2013; Jansen *et al.* 2017; Zhernakova *et al.* 2017). The identification of eQTL is particularly important to inferring functional relevance for loci identified by GWAS, and as key contributors to end-point phenotypes themselves through their influence on gene regulation. Therefore, eQTL may be useful as markers for prediction and selection of physiological traits.

Another candidate region identified in the GWAS for ΔPC1 was on ECA2; the two SNPs reaching the suggestive threshold for association were within *SCLT1*. *SCLT1* has also been found to be significantly upregulated in skeletal muscle in response to exercise in Thoroughbreds ($\log_2 FC = 0.52$, $P_{FDR} = 2.98 \times 10^{-5}$; Table 6). *SCLT1* encodes an adaptor protein CAP-1a linking voltage-gated sodium channel $Na_v 1.8$ and clathrin, expressed primarily in dorsal root ganglia and hypothesised to be related to nociception (Liu *et al.* 2005). However, a recent study in *Sclt1* knockout mice found that they exhibited a range of phenotypes including polycystic kidneys, shortened intestine and increased fibrosis (Li *et al.* 2017). This is suspected to be a result of reduced ciliary function and generation (Tanos *et al.* 2013). *SCLT1* has also been implicated as a significant driver of breast cancer using transcriptomic signatures (Pongor *et al.* 2015). This may be due to *SCLT1* loss increasing PKA and subsequent ERK and STAT3 signalling, increasing cell proliferation (Li *et al.* 2017). This

role in cell proliferation may therefore have implications for cell growth and turnover which may be relevant to exercise.

Conclusion

Research into the genetic basis of exercise traits in a highly adapted animal model such as the Thoroughbred continues to provide insights into the genomics underlying the physiological and skeletal muscle transcriptional responses to exercise and training that subsequently impact athletic performance on the racecourse. Using GPS monitoring during training in a cohort of active racing Thoroughbreds, we derived a useful index of speed, PC1, that was significantly different among 2-year-old elite, non-elite and unraced horses. Using 2yo_PC1 as a phenotype in a GWAS, we identified a candidate region on ECA8 including the *MYO18B* gene, which may be related to sarcomere formation and function. A GWAS for ΔPC1 identified candidate regions on ECA2 and ECA11. The three ECA11 SNPs defined a 1.82 Mb region containing several genes involved in the transcriptional response to exercise and training, including *MYO18A*, a central component of the regulation of skeletal muscle gene expression in the horse (Bryan *et al.* 2017). We hypothesise, therefore, that genomic variation in this region may alter the transcriptional response to exercise, mediating variation in end-point phenotypes that may respond to training. This new knowledge of genomic variant contribution in speed variables, which appear to be central regulators of the transcriptional response to exercise, may make them useful predictors for the training response in Thoroughbreds. In turn, such genomic information may be applied in the early identification of horses with higher potential for speed, and more specifically by trainers to inform the optimisation of individual training regimes.

Acknowledgements

Conflicts of interest

EWH, DEM and LMK are named inventors on multiple international patents relating to the application of *MSTN* variation in the prediction of race distance performance.

Data availability

Data are available on request from the UCD Technology Transfer Office for researchers who meet the criteria for access to confidential data.

References

Adhihetty P.J., Irrcher I., Joseph A.M., Ljubicic V. & Hood D.A. (2003) Plasticity of skeletal muscle mitochondria in response to contractile activity. *Experimental Physiology* **88**, 99–107.

Ajima R., Akazawa H., Kodama M., Takeshita F., Otsuka A., Kohno T., Komuro I., Ochiya T. & Yokota J. (2008) Deficiency of Myo18B in mice results in embryonic lethality with cardiac myofibrillar aberrations. *Genes to Cells* **13**, 987–99.

Alazami A.M., Kentab A.Y., Faqeih E., Mohamed J.Y., Alkhalidi H., Hijazi H. & Alkuraya F.S. (2015) A novel syndrome of Klippel-Feil anomaly, myopathy, and characteristic facies is linked to a null mutation in MYO18B. *Journal of Medical Genetics* **52**, 400–4.

Allen K.J., Erck-Westergren E. & Franklin S.H. (2016) Exercise testing in the equine athlete. *Equine Veterinary Education* **28**, 89–98.

Aulchenko Y.S., Ripke S., Isaacs A. & Van Duijn C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–6.

Beisswanger R., Corbeil D., Vannier C., Thiele C., Dohrmann U., Kellner R., Ashman K., Niehrs C. & Huttner W.B. (1998) Existence of distinct tyrosylprotein sulfotransferase genes: molecular characterization of tyrosylprotein sulfotransferase-2. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11134–9.

Berger J., Berger S., Li M. & Currie P.D. (2017) Myo18b is essential for sarcomere assembly in fast skeletal muscle. *Human Molecular Genetics* **26**, 1146–56.

Beyer A. (1993) *Beyer on Speed*. Houghton Mifflin, New York.

Billingsley G., Santhiya S.T., Paterson A.D. *et al.* (2006) CRYBA4, a novel human cataract gene, is also involved in microphthalmia. *The American Journal of Human Genetics* **79**, 702–9.

Binns M.M., Boehler D.A., Bailey E., Lear T.L., Cardwell J.M. & Lambert D.H. (2012) Inbreeding in the Thoroughbred horse. *Animal Genetics* **43**, 340–2.

Booth F.W., Ruegsegger G.N., Toedebusch R.G. & Yan Z. (2015) Endurance exercise and the regulation of skeletal muscle metabolism. *Progress in Molecular Biology and Translational Science* **135**, 129–51. https://doi.org/10.1016/bs.pmbts.2015.07.016

Bouchard C., Leon A.S., Rao D.C., Skinner J.S., Wilmore J.H. & Gagnon J. (1995) The HERITAGE family study. Aims, design, and measurement protocol. *Medicine and Science in Sports and Exercise* **27**, 721–9.

Bouchard C., Sarzynski M.A., Rice T.K., Kraus W.E., Church T.S., Sung Y.J., Rao D.C. & Rankinen T. (2011) Genomic predictors of the maximal O-2 uptake response to standardized exercise training programs. *Journal of Applied Physiology* **110**, 1160–70.

Brown C.D., Mangravite L.M. & Engelhardt B.E. (2013) Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genetics* **9**, e1003649.

Browning B.L. & Browning S.R. (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–71.

Bryan K., McGivney B.A., Farries G., McGettigan P.A., McGivney C.L., Gough K.F., MacHugh D.E., Katz L.M. & Hill E.W. (2017) Equine skeletal muscle adaptations to exercise and training: evidence of differential regulation of autophagosomal and mitochondrial components. *BMC Genomics* **18**, 595.

Bupha-Intr T., Holmes J.W. & Janssen P.M. (2007) Induction of hypertrophy in vitro by mechanical loading in adult rabbit myocardium. *American Journal of Physiology. Heart and Circulatory Physiology* **293**, H3759–67.

Carella C., Bonten J., Sirma S. *et al.* (2007) MN1 overexpression is an important step in the development of inv(16) AML. *Leukemia* **21**, 1679–90.

Constantinopol M., Jones J.H., Weibel E.R., Taylor C.R., Lindholm A. & Karas R.H. (1989) Oxygen transport during exercise in large mammals. II. Oxygen uptake by the pulmonary gas exchanger. *Journal of Applied Physiology* **67**, 871–8.

Corbin L.J., Blott S.C., Swinburne J.E., Vaudin M., Bishop S.C. & Woolliams J.A. (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics* **41**(Suppl 2), 8–15.

Cunnick J.M., Schmidhuber S., Chen G. *et al.* (2009) Bcr and Abr cooperate in negatively regulating acute inflammatory responses. *Molecular and Cellular Biology* **29**, 5742–50.

Cunningham E.P., Dooley J.J., Splan R.K. & Bradley D.G. (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics* **32**, 360–4.

Dickinson J.M., D'Lugos A.C., Naymik M.A., Siniard A.L., Wolfe A.J., Curtis D.R., Huentelman M.J. & Carroll C.C. (2018) Transcriptome response of human skeletal muscle to divergent exercise stimuli. *Journal of Applied Physiology*, **124**, 1529–40.

Eto D., Yamano S., Mukai K., Sugiura T., Nasu T., Tokuriki M. & Miyata H. (2004) Effect of high intensity training on anaerobic capacity of middle gluteal muscle in Thoroughbred horses. *Research in Veterinary Science* **76**, 139–44.

Evans D.L. & Rose R.J. (1988) Cardiovascular and respiratory responses to submaximal exercise training in the thoroughbred horse. *Pflugers Archiv (European Journal of Physiology)* **411**, 316–21.

Evans D.L., Harris R.C. & Snow D.H. (1993) Correlation of racing performance with blood lactate and heart rate after exercise in thoroughbred horses. *Equine Veterinary Journal* **25**, 441–5.

Farries G., McGettigan P.A., Gough K.F., McGivney B.A., MacHugh D.E., Katz L.M. & Hill E.W. (2018) Genetic contributions to precocity traits in racing Thoroughbreds. *Animal Genetics* **49** 193–204. https://doi.org/10.1111/age.12622

Field J.K. & Cunningham E.P. (1976) A further study of the inheritance of racing performance in thoroughbred horses. *Journal of Heredity* **67**, 247–8.

Fonseca R.G., Kenny D.A., Hill E.W. & Katz L.M. (2010) The association of various speed indices to training responses in Thoroughbred flat racehorses measured with a global positioning and heart rate monitoring system. *Equine Veterinary Journal Supplement* **42**, 51–7. https://doi.org/10.1111/j.2042-3306.2010.00272.x

Foreman J.H., Bayly W.M., Allen J.R., Matoba H., Grant B.D. & Gollnick P.D. (1990) Muscle responses of thoroughbreds to conventional race training and detraining. *American Journal of Veterinary Research* **51**, 909–13.

Gaffney B. & Cunningham E.P. (1988) Estimation of genetic trend in racing performance of thoroughbred horses. *Nature* **332**, 722–4.

Gilad Y., Rifkin S.A. & Pritchard J.K. (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **24**, 408–15.

Golson M.L. & Kaestner K.H. (2016) Fox transcription factors: from development to disease. *Development* **143**, 4558–70.

Gong D., Fei F., Lim M., Yu M., Groffen J. & Heisterkamp N. (2013) Abr, a negative regulator of Rac, attenuates cockroach allergen-induced asthma in a mouse model. *Journal of Immunology* **191**, 4514–20.

Gramkow H.L. & Evans D.L. (2006) Correlation of race earnings with velocity at maximal heart rate during a field exercise test in thoroughbred racehorses. *Equine Veterinary Journal Supplement* 118–22.

Gramm M. & Marksteiner R. (2010) The effect of age on Thoroughbred racing performance. *Journal of Equine Science* **21**, 73–8.

GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. *Nature Genetics* **45**, 580–5.

Gunn H.M. (1987) Muscle, bone and fat proportions and muscle distribution of thoroughbreds and other horses. In: *Equine Exercise Physiology 2: Proceedings of the Second International Conference on Equine Exercise Physiology, San Diego, California, August 7–11, 1986* (Ed. by J.R. Gillespie & N.E. Robinson). San Diego, CA: ICEEP Publications.

Harkins J.D., Beadle R.E. & Kamerling S.G. (1993) The correlation of running ability and physiological variables in thoroughbred racehorses. *Equine Veterinary Journal* **25**, 53–60.

Hartman J.W., Moore D.R. & Phillips S.M. (2006) Resistance training reduces whole-body protein turnover and improves net protein retention in untrained young males. *Applied Physiology, Nutrition, and Metabolism* **31**, 557–64.

Hellstrom M., Kalen M., Lindahl P., Abramsson A. & Betsholtz C. (1999) Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–55.

Heuser M., Beutel G., Krauter J., Dohner K., von Neuhoff N., Schlegelberger B. & Ganser A. (2006) High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood* **108**, 3898–905.

Hill E.W., Gu J., Eivers S.S., Fonseca R.G., McGivney B.A., Govindarajan P., Orr N., Katz L.M. & MacHugh D.E. (2010a) A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* **5**, e8645.

Hill E.W., McGivney B.A., Gu J.J., Whiston R. & MacHugh D.E. (2010b) A genome-wide SNP-association study confirms a sequence variant (g.66493737C > T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics* **11**, 9.

Hill E.W., Fonseca R.G., McGivney B.A., Gu J.J., MacHugh D.E. & Katz L.M. (2012) MSTN genotype (g.66493737C/T) association with speed indices in Thoroughbred racehorses. *Journal of Applied Physiology* **112**, 86–90.

Hill E.W., McGivney B.A., Rooney M.F., Katz L.M., Parnell A. & MacHugh D.E. (2019) The contribution of myostatin (MSTN) and additional modifying genetic loci to race distance aptitude in Thoroughbred horses racing in different geographic regions. *Equine Veterinary Journal* **51**, 625–633. https://doi.org/10.1111/evj.13058

Hinchcliff K.W., Lauderdale M.A., Dutson J., Geor R.J., Lacombe V.A. & Taylor L.E. (2002) High intensity exercise conditioning increases accumulated oxygen deficit of horses. *Equine Veterinary Journal* **34**, 9–16.

Hindorff L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S. & Manolio T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9362–7.

Ho D.W., Yap M.K., Ng P.W., Fung W.Y. & Yip S.P. (2012) Association of high myopia with crystallin beta A4 (CRYBA4) gene polymorphisms in the linkage-identified MYP6 locus. *PLoS One* **7**, e40238.

Holloszy J.O., Rennie M.J., Hickson R.C., Conlee R.K. & Hagberg J.M. (1977) Physiological consequences of the biochemical adaptations to endurance exercise. *Annals of the New York Academy of Sciences* **301**, 440–50.

Jackson B.C., Carpenter C., Nebert D.W. & Vasiliou V. (2010) Update of human and mouse forkhead box (FOX) gene families. *Human Genomics* **4**, 345–52.

Jansen R., Hottenga J.J., Nivard M.G., Abdellaoui A., Laport B., de Geus E.J., Wright F.A., Penninx B. & Boomsma D.I. (2017) Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Human Molecular Genetics* **26**, 1444–51.

Jiu Y., Kumari R., Fenix A.M., Schaible N., Liu X., Varjosalo M., Krishnan R., Burnette D.T. & Lappalainen P. (2019) Myosin-18B promotes the assembly of myosin II stacks for maturation of contractile actomyosin bundles. *Current Biology* **29**, 81–92.e5.

Jones J.H. & Lindstedt S.L. (1993) Limits to maximal performance. *Annual Review of Physiology* **55**, 547–69.

Jones J.H., Longworth K.E., Lindholm A., Conley K.E., Karas R.H., Kayar S.R. & Taylor C.R. (1989) Oxygen transport during exercise in large mammals. I. Adaptive variation in oxygen demand. *Journal of Applied Physiology* **67**, 862–70.

Katz L.M., Bayly W.M., Hines M.T. & Sides R.H. (1999) Differences in the ventilatory responses of horses and ponies to exercise of varying intensities. *Equine Veterinary Journal Supplement* **30**, 49–51.

Kingston J.K., Soppet G.M., Rogers C.W. & Firth E.C. (2006) Use of a global positioning and heart rate monitoring system to assess training load in a group of Thoroughbred racehorses. *Equine Veterinary Journal* **38**, 106–9.

Kobayashi M., Kuribara K. & Amada A. (1999) Application of V200 values for evaluation of training effects in the young Thoroughbred under field conditions. *Equine Veterinary Journal* **31**, 159–62.

Li M.X., Yeung J.M., Cherny S.S. & Sham P.C. (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics* **131**, 747–56.

Li J., Lu D., Liu H., Williams B.O., Overbeek P.A., Lee B., Zheng L. & Yang T. (2017) Sclt1 deficiency causes cystic kidney by activating ERK and STAT3 signaling. *Human Molecular Genetics* **26**, 2949–60.

Liang Y., Wang A., Belyantseva I.A. *et al.* (1999) Characterization of the human and mouse unconventional myosin XV genes responsible for hereditary deafness DFNB3 and shaker 2. *Genomics* **61**, 243–58.

Liu C., Cummins T.R., Tyrrell L., Black J.A., Waxman S.G. & Dib-Hajj S.D. (2005) CAP-1A is a novel linker that binds clathrin and the voltage-gated sodium channel Na(v)1.8. *Molecular and Cellular Neurosciences* **28**, 636–49.

Manolio T.A., Collins F.S., Cox N.J. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747–53.

McGowan C.M., Golland L.C., Evans D.L., Hodgson D.R. & Rose R.J. (2002) Effects of prolonged training, overtraining and detraining on skeletal muscle metabolites and enzymes. *Equine Veterinary Journal Supplement* 257–63.

Mota M.D.S., Abrahao A.R. & Oliveira H.N. (2005) Genetic and environmental parameters for racing time at different distances in Brazilian Thoroughbreds. *Journal of Animal Breeding and Genetics* **122**, 393–9.

Naylor A.J., McGettrick H.M., Maynard W.D., May P., Barone F., Croft A.P., Egginton S. & Buckley C.D. (2014) A differential role for CD248 (Endosialin) in PDGF-mediated skeletal muscle angiogenesis. *PLoS One* **9**, e107146.

Oki H., Sasaki Y. & Willham R.L. (1995) Genetic parameter estimates for racing time by restricted maximum-likelihood in the Thoroughbred horse of Japan. *Journal of Animal Breeding and Genetics* **112**, 146–50.

Ouyang Y.B., Crawley J.T., Aston C.E. & Moore K.L. (2002) Reduced body weight and increased postimplantation fetal death in tyrosylprotein sulfotransferase-1-deficient mice. *Journal of Biological Chemistry* **277**, 23781–7.

Peck L.S., Thorne M.A., Hoffman J.I., Morley S.A. & Clark M.S. (2015) Variability among individuals is generated at the gene expression level. *Ecology* **96**, 2004–14.

Petersen J.L., Valberg S.J., Mickelson J.R. & McCue M.E. (2014) Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Animal Genetics* **45**, 827–35.

Pongor L., Kormos M., Hatzis C., Pusztai L., Szabo A. & Gyorffy B. (2015) A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome Medicine* **7**, 104.

Purcell S., Neale B., Todd-Brown K., *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

R Core Team (2017) *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rankinen T., Sung Y.J., Sarzynski M.A., Rice T.K., Rao D.C. & Bouchard C. (2012) Heritability of submaximal exercise heart rate response to exercise training is accounted for by nine SNPs. *Journal of Applied Physiology* **112**, 892–7.

Rice T., Daw E.W., Gagnon J., Bouchard C., Leon A.S., Skinner J.S., Wilmore J.H. & Rao D.C. (1997) Familial resemblance for body composition measures: the HERITAGE Family Study. *Obesity Research* **5**, 557–62.

Rice T.K., Sarzynski M.A., Sung Y.J., Argyropoulos G., Stutz A.M., Teran-Garcia M., Rao D.C., Bouchard C. & Rankinen T. (2012) Fine mapping of a QTL on chromosome 13 for submaximal exercise capacity training response: the HERITAGE Family Study. *European Journal of Applied Physiology* **112**, 2969–78.

Rico-Sanz J., Rankinen T., Rice T., Leon A.S., Skinner J.S., Wilmore J.H., Rao D.C. & Bouchard C. (2004) Quantitative trait loci for maximal exercise capacity phenotypes and their responses to training in the HERITAGE Family Study. *Physiological Genomics* **16**, 256–60.

Rivero J.L., Sporleder H.P., Quiroz-Rothe E., Vervuert I., Coenen M. & Harmeyer J. (2002) Oral L-carnitine combined with training promotes changes in skeletal muscle. *Equine Veterinary Journal Supplement* 269–74.

Rivero J.L., Ruz A., Marti-Korff S., Estepa J.C., Aguilera-Tejero E., Werkman J., Sobotta M. & Lindner A. (2007) Effects of intensity and duration of exercise on muscular responses to training of thoroughbred racehorses. *Journal of Applied Physiology* **102**, 1871–82.

Roneus M. (1993) Muscle characteristics in standardbreds of different ages and sexes. *Equine Veterinary Journal* **25**, 143–6.

Roneus M., Essen-Gustavsson B., Lindholm A. & Persson S.G. (1992) Skeletal muscle characteristics in young trained and untrained standardbred trotters. *Equine Veterinary Journal* **24**, 292–4.

Rooney M.F., Hill E.W., Kelly V.P. & Porter R.K. (2018) The "speed gene" effect of myostatin arises in Thoroughbred horses due to a promoter proximal SINE insertion. *PLoS One* **13**, e0205664.

Sasaki N., Hosoda Y., Nagata A., Ding M., Cheng J.M., Miyamoto T., Okano S., Asano A., Miyoshi I. & Agui T. (2007) A mutation in Tpst2 encoding tyrosylprotein sulfotransferase causes dwarfism associated with hypothyroidism. *Molecular Endocrinology* **21**, 1713–21.

Serrano A.L., Quiroz-Rothe E. & Rivero J.L. (2000) Early and long-term changes of equine skeletal muscle in response to endurance training and detraining. *Pflugers Archiv (European Journal of Physiology)* **441**, 263–74.

Smedley D., Haider S., Durinck S. *et al.* (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research* **43**, W589–98.

Takahashi T. (2015) The effect of age on the racing speed of Thoroughbred racehorses. *Journal of Equine Science* **26**, 43–8.

Tanos B.E., Yang H.J., Soni R., Wang W.J., Macaluso F.P., Asara J.M. & Tsou M.F. (2013) Centriole distal appendages promote membrane docking, leading to cilia initiation. *Genes and Development* **27**, 163–8.

Tozaki T., Sato F., Hill E.W. *et al.* (2011) Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. *Journal of Veterinary Medical Science* **73**, 1617–24.

Tozaki T., Miyake T., Kakoi H., Gawahara H., Hirota K., Nakano Y. & Kurosawa M. (2012) Heritability estimates for racing performance in Japanese Thoroughbred racehorses using linear and non-linear model analyses. *Journal of Animal Breeding and Genetics* **129**, 402–8.

Turner S.D. (2018) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **3**, 751.

Valk P.J., Verhaak R.G., Beijen M.A. *et al.* (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *New England Journal of Medicine* **350**, 1617–28.

Velie B.D., Hamilton N.A. & Wade C.M. (2015) Heritability of racing performance in the Australian Thoroughbred racing population. *Animal Genetics* **46**, 23–9.

Venables W.N. & Ripley B.D. (2010) *Modern Applied Statistics with S.* Springer Publishing Company, Incorporated, New York, NY.

Vermeulen A.D. & Evans D.L. (2006) Measurements of fitness in thoroughbred racehorses using field studies of heart rate and velocity with a global positioning system. *Equine Veterinary Journal Supplement* 113–7.

Wade C.M., Giulotto E., Sigurdsson S. *et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science (New York, N.Y.)* **326**, 865–7.

Wickham H. (2009) *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag, New York.

Yamano S., Eto D., Sugiura T., Kai M., Hiraga A., Tokuriki M. & Miyata H. (2002) Effect of growth and training on muscle adaptation in Thoroughbred horses. *American Journal of Veterinary Research* **63**, 1408–12.

Yang J.A., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82.

Yarasheski K.E., Zachwieja J.J. & Bier D.M. (1993) Acute effects of resistance exercise on muscle protein synthesis rate in young and elderly men and women. *American Journal of Physiology* **265**, E210–4.

Zhernakova D.V., Deelen P., Vermaat M. *et al.* (2017) Identification of context-dependent expression quantitative trait loci in whole blood. *Nature Genetics* **49**, 139–45.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 PCA of the genomic relationship matrix for horses used in GWAS.

Figure S2 Linear model of measured speed (using principal component 1) with cumulative work days for the three *MSTN* g.66493737C>T SNP genotypes.

Figure S3 Q–Q plot of *P*-values of association using linear vs. mixed model association for median recorded 2-year-old speed, as quantified by principal component 1 (2yo_PC1).

Figure S4 Q–Q plot of *P*-values of association using linear vs. mixed-model association for the change in measured speed (principal component 1) with age (ΔPC1).

Table S1 Number of recordings at each level of track condition.

Table S2 (a) Haplotypes identified in ECA8 candidate region for median 2-year-old speed (as measured by principal component 1, PC1). (b) Haplotypes identified in ECA1 candidate region for change in speed (as measured by PC1) with age.

Table S3 Allele frequencies for candidate SNPs reaching the suggestive threshold for association with median 2-year-old speed (2yo_PC1) and speed progression with age (principal component 1) between elite (won or placed at Listed or Group level) and non-elite populations (raced, but never won or placed at Listed or Group level).

Table S4 Allele frequencies for candidate SNPs reaching the suggestive threshold for association with median 2-year-old speed (2yo_PC1) and speed progression with age (ΔPC1) between raced and unraced populations.