Title
Risk of false positive genetic associations in complex traits with underlying population structure: A case study

Permalink
https://escholarship.org/uc/item/9nx02033

Journal
The Veterinary Journal, 202(3)

ISSN
1090-0233

Authors
Finno, Carrie J
Aleman, Monica
Higgins, Robert J
et al.

Publication Date
2014-12-01

DOI
10.1016/j.tvjl.2014.09.013

Peer reviewed

Risk of false positive genetic associations in complex traits with underlying population structure: A case study

Carrie J. Finno [a,*], Monica Aleman [b], Robert J. Higgins [c], John E. Madigan [b], Danika L. Bannasch [a]

[a] Department of Population Health and Reproduction, University of California, Davis, CA 95616, USA
[b] Department of Medicine and Epidemiology, University of California, Davis, CA 95616, USA
[c] Department of Pathology, Microbiology and Immunology, University of California, Davis, CA 95616, USA

ARTICLE INFO

ABSTRACT

Genome-wide association (GWA) studies are widely used to investigate the genetic etiology of diseases in domestic animals. In the horse, GWA studies using 40–50,000 single nucleotide polymorphisms (SNPs) in sample sizes of 30–40 individuals, consisting of only 6–14 affected horses, have led to the discovery of genetic mutations for simple monogenic traits. Equine neuroaxonal dystrophy is a common inherited neurological disorder characterized by symmetric ataxia. A case-control GWA study was performed using genotypes from 42,819 SNP marker loci distributed across the genome in 99 clinically phenotyped Quarter horses (37 affected, 62 unaffected).

A significant GWA was not achieved although a suggestive association was uncovered when only the most stringently phenotyped NAD-affected horses ($n = 10$) were included (chromosome 8:62130605 and 62134644 [$\log(1/P) = 5.56$]). Candidate genes (*PIK3C3*, *RIT2*, and *SYT4*) within the associated region were excluded through sequencing, association testing of uncovered variants and quantitative RT-PCR. It was concluded that variants in *PIK3C3*, *RIT2*, and *SYT4* are not responsible for equine neuroaxonal dystrophy. This study demonstrates the risk of false positive associations when performing GWA studies on complex traits and underlying population structure when using 40–50,000 SNP markers and small sample size.

Published by Elsevier Ltd.

Introduction

Over the past 25 years, two approaches have been applied to uncover genes contributing to specific diseases in domestic animals. The first approach targets candidate genes selected based on their role in comparative diseases in other species, while the second involves mapping disease traits of interest to a chromosomal location using genetic markers (Andersson and Georges, 2004). The recent availability of large panels of single nucleotide polymorphisms (SNPs) in domestic animal species has led to an expansion in the search for genomic regions associated with genetic diseases through the use of genome-wide association (GWA) studies.

In the horse, the first-generation SNP array (Illumina EquineSNP50 Beadchip) contains 54,602 SNPs (McCue et al., 2012). This array was used to identify chromosomal regions, leading to mutation discovery, for various monogenic traits in the horse using relatively small sample sizes (Brooks et al., 2010; Fox-Clipsham et al., 2011; Andersson et al., 2012). In addition to simple monogenic traits, GWA studies using the equine SNP chips have identified quantitative trait loci for more complex traits (Dupuis et al., 2011; Corbin et al., 2012; Kulbrock et al., 2013). Despite the apparent success in mapping more complex traits and diseases, underlying genetic mutations have not been uncovered for many of these traits. The purpose of the present study was to demonstrate the risks of false positive associations when performing a GWA study using small sample sizes in populations with complex traits and underlying population structure.

Equine neuroaxonal dystrophy (NAD) is characterized by a symmetric ataxia and proprioceptive deficits, developing between 6 and 12 months of age with no sex predilection (Beech and Haskins, 1987; Aleman et al., 2011; Finno et al., 2013). Equine degenerative myeloencephalopathy (EDM) is considered a more pathologically advanced form of NAD, in which evidence of histological damage extends to the white matter of the spinal cord. Because we have previously established that cases of NAD and EDM can occur within families with the same underlying environmental risk factors and that EDM appears to be a pathologically more extensive form of NAD (Aleman et al., 2011), the disease has subsequently been termed NAD/EDM.

Currently there is no means to establish an ante-mortem diagnosis of NAD/EDM. Vitamin E plays an important role in the development of NAD/EDM in genetically predisposed foals (Blythe et al., 1991) but low serum vitamin E is not consistently reported

in all cases (Dill et al., 1990). There is very strong evidence that, in susceptible families, dietary vitamin E in the susceptible foal modifies the severity of the phenotype (Aleman et al., 2011). Definitive diagnosis, for the purposes of a genetic investigation, requires identification of characteristic lesions in the brainstem and spinal cord at post-mortem.

We have previously demonstrated a complex mode of inheritance for NAD/EDM and excluded putative variants in a strong candidate gene, α-tocopherol transfer protein (*TTPA*), as causative for NAD/EDM (Finno et al., 2013). Based on comparative knowledge of the clinical and histological presentation of NAD/EDM, there are no additional candidate genes to evaluate and either a linkage or association study is required to further investigate the etiology. Once a candidate region is discovered, prioritized genes for further evaluation will include those genes involved in synaptic function because studies have demonstrated accumulation of synaptic proteins in EDM-affected horses (Siso et al., 2003).

A GWA study was performed using two populations: (1) 37 clinically affected NAD/EDM Quarter horses (QHs) and 62 clinically unaffected QHs, and (2) a subset of the previous population that included only the 10 affected NAD/EDM QHs in which the disease was confirmed at post-mortem plus the same 62 clinically phenotyped unaffected QHs. Candidate genes were prioritized within the region(s) of significant association for further investigation. We hypothesized that, although NAD/EDM appears to be a complex disease trait, a major gene involved in synaptic transmission would be identified through the GWA study and a variant within that gene would be significantly associated with NAD/EDM. This study highlights the importance of adequate sample size and evaluation of underlying population structure when performing GWA studies to identify true regions of association using low coverage SNP arrays in complex traits.

Materials and methods

Horses

DNA was collected from 99 clinically phenotyped QHs, including 37 clinically NAD/EDM-affected and 62 unaffected QHs. Horses were defined as clinically affected with a mean ataxia score ≥2 and unaffected with a mean ataxia score of 0 as previously described (Aleman et al., 2011). Ten of the NAD/EDM-affected and four unaffected horses were confirmed at post-mortem examination, with histological findings as previously described (Aleman et al., 2011).

Of the population used for the GWA study, 33/37 affected and 53/62 control horses overlapped with horses used in a previous genetic study (Finno et al., 2013). Unaffected horses were ≥1 year of age and 1-year-old horses were re-examined at 3 years of age to confirm an unaffected phenotype, because cases of NAD/EDM have been reported up to 3 years of age (Beech and Haskins, 1987). Horses under 1 year of age were not included in this group because clinical signs of NAD/EDM may not be apparent during the first year of life. Of the post-mortem confirmed affected cases, five were by the same sire and five were unrelated within three generations.

Unaffected horses were sampled from the same farm as affected cases and 47/62 of these were unrelated to NAD/EDM-affected horses within three generations. Fifteen unaffected horses were related to NAD/EDM-affected horses within two generations but were raised under identical environmental conditions as NAD/EDM-affected horses and demonstrated no evidence of neurological disease. Therefore, these horses served as ideal controls to balance relatedness between NAD/EDM-affected and unaffected cases.

All protocols were approved by the University of California, Davis Institutional Animal Care and Use Committee (Protocol 15963).

Genome wide association (GWA)

Horses were genotyped for 54,602 SNPs using the Equine SNP50 genotyping array (Illumina). SNPs were selected that passed quality control settings (minor allele frequency >1% and genotyping across individuals >90%). A case/control standard allelic GWA study was performed and population stratification assessed by estimating the genomic control inflation factor (λ) using GenABEL (Aulchenko et al., 2007b), implemented in the R program (R Development Core Team, 2013). When λ = 1, there is no population stratification and association results should not be influenced by population structure (Wu et al., 2011).

To account for the population substructure and relatedness in this population of horses, a linear mixed model was implemented, using two distinct algorithms. The first program utilized an approximation method to perform the linear mixed model. Genome-wide rapid association using mixed model and regression (GRAMMAR), implemented in the R package GenABEL, first estimates the residuals from the linear mixed model under the null model and then treats these residuals as phenotypes for further genome-wide analysis by a standard linear model (Aulchenko et al., 2007a). The second program used an exact method to perform the linear mixed model, thereby avoiding repeatedly estimating variance components when performing each test, and was implemented through GEMMA (Zhou and Stephens, 2012). Both algorithms perform linear mixed models based on clustering that accounts for both population substructure and relatedness through use of a kinship matrix estimated from identical by descent distances.

Significance thresholds

A Bonferroni correction for 42,819 tests (corresponding to the number of usable SNPs), defined by significant $P_{genome-wide} = 0.05$, was determined, yielding a threshold of 1.17×10^{-6} (significant, log [$1/P$] = 5.93). As Bonferroni corrections to control type I error in genetic association studies are highly conservative and have been shown to 'overcorrect' SNPs that are not truly independent, it has been recommended to apply both significant and suggestive P-value thresholds to properly control for type I error (Duggal et al., 2008). There is a wide lack of consistency in suggestive P-values applied in GWA studies using 30–50,000 SNP markers, both within and across species (Dupuis et al., 2011; Do et al., 2014; Zhang et al., 2014). False discovery rate (FDR) has lower incidence of type II error (Verhoeven et al., 2005) and setting the FDR at 0.10 and 0.05 has been recommended as criteria for suggestive and significant linkage, respectively (Benjamini and Yekutieli, 2005). Therefore, to define a suggestive association using an FDR set at 0.10, empirical P-values (P_{emp}) from the m independent tests were ranked from $P_1 \ldots P_m$ for each locus ($i-m$) and then tested against the partial inequality $P_i \leq \alpha i/m$, where $\alpha = 0.10$ and i was the rank of that test based on P_{emp} in ascending order. Where $(P_{emp})_i \leq \alpha i/m$, the null hypothesis and those with lower P_i were rejected (Verhoeven et al., 2005).

Haplotype analysis

For any SNP with a suggestive association, haplotypes were reconstructed on that particular chromosome using Haploview (Barrett et al., 2005). Association testing of both single markers and haplotypes was performed with the number of permutations based on the number of markers on that particular chromosome, with the adjusted haplotype-wide significance threshold then set at $P_{corrected} = 0.05$.

Candidate gene sequencing

The highest genome-wide suggestive regions, upon correction for genomic inflation, were screened for candidate genes using the equine reference database, available at UCSC.[1] Within 1 Mb of the candidate region surrounding the two most highly associated SNPs on ECA8, there were three genes, all of which demonstrate expression in the central nervous system, *PIK3C3*, *RIT2* and *SYT4*. Two post-mortem-confirmed NAD/EDM affected and one post-mortem-confirmed unaffected horses (from GWA study) were selected for sequencing. *PIKC3* and *RIT2* were sequenced in both genomic and cDNA, while *SYT4* was sequenced in cDNA only.

The equine orthologs to human genes *PIK3C3* (NM_002647) and *RIT2* (NM_002930) were identified by the equine BLAT search[1] and exons identified within the September 2007 2.0 draft assembly of the domestic horse (*Equus caballus*)[2]. In addition to exons and ≥200 base pairs of flanking intronic sequence, both *PIK3C3* and *RIT2* were evaluated in the February 2009 human assembly[3] for promoter-associated sequence and variants in the 5′ untranslated region (UTR) and splice sites using Ensembl[4]. Sequences were compared to the horse assembly by an equine BLAT search and regions were included for sequencing. Primers flanking each region were designed[5] (Rozen and Skaletsky, 2000) and PCR performed using primer-specific melting temperatures (Appendix: Supplementary Table S1). Sequences were scanned for variants and the equine reference sequence was used as an additional unaffected sample.

Fine structure mapping

A custom-designed SNP genotyping platform was created using the 88 variants (84 SNPs, four insertions/deletions) uncovered from sequencing of *PIK3C3* and *RIT2* only as non-synonymous variants were not uncovered through sequencing of

---

SYT4 in cDNA. Genotyping was performed using a custom genotyping system (MassArray iPlex Gold, Sequenom) on the same 10 affected and 62 unaffected cases that had been genotyped on the EquineSNP50 platform. Primers are listed in Appendix: Supplementary Table S2. After quality filtration, 70 variants remained (12 excluded for minor allele frequency [maf] <1% and six excluded for genotyping <90%). A linear mixed model analysis, using GEMMA, was performed on the 10 affected and 62 unaffected cases with these 70 total variants. A Bonferroni adjustment was used to account for multiple testing (0.05/70 variants) with an adjusted $P = 7.14 \times 10^{-4}$.

Reverse transcription PCR (RT-PCR)

PIK3C3, *RIT2* and *SYT4* are expressed in the central nervous system (Ullrich et al., 1994; Lee et al., 1996; Zhou et al., 2010). Samples from the caudal brainstem, at the region of the obex, of the same NAD affected ($n = 2$) and unaffected horses ($n = 2$; original from genomic sequencing plus an additional unaffected QH) were flash-frozen immediately at necropsy and used to prepare mRNA (FastTrack 2.0 mRNA Isolation kit, Invitrogen). The equine cDNA ortholog to human gene SYT4 (NM_020783) was identified by the equine BLAT search. Primers are included in Appendix: Supplementary Table S1. PCR products were sequenced and evaluated for variants.

Quantitative reverse transcription PCR (q-RT-PCR)

For the three genes of interest – *PIK3C3*, *RIT2* and *SYT4* – and three reference genes – *ACTB*, *GAPDH*, *HPRT1* – a qRT-PCR assay was designed using the Universal Probe Library system assay design (Roche Diagnostics) with gene-specific primers (Appendix: Supplementary Table S2). Primers were designed to cover exons that were included in all of the reported transcript variants (*PIK3C3* exon 24–25, *RIT2* exon 4, *SYT4* exon 2). Real time RT-PCR reactions were carried out on a 7900HT Fast Real-time PCR System (Life Technologies). Standard curves were constructed amplifying twofold serial dilutions of the same cDNA, which was used as calibrator for gene expression analyses. For each sample, the C_t (threshold cycle) determined the relative amount of target gene; each measurement was made in duplicate, and normalized to the average of the reference genes, which were also measured in duplicate. Because efficiencies ranged from 0.98 to 1.04 across all genes, indicating that the three housekeeping genes performed equally well, C_t was averaged across the housekeeping genes.

Relative quantification of gene expression was calculated by comparative threshold cycle method ($2^{-\Delta\Delta Ct}$), where $2^{-\Delta\Delta Ct}$ was calculated for each gene of interest using the average C_t of the three housekeeping genes. Data were analyzed using non-parametric statistical methods (Mann–Whitney test) because of the limited

number of cases in the study. A Bonferroni corrected *P*-value <0.01 ($\alpha = 0.05$/ number of tests) for multiple testing was selected as the level of significance. Results are reported as $2^{-\Delta\Delta Ct}$ values and 95% confidence intervals (CI).

Results

Genome wide association (GWA) study: Population 1 (clinical case/control)

After quality filtration, 42,819 SNPs remained (6441 excluded for minor allele frequency [maf] <1% and 5342 excluded for genotyping <90%). The first allelic case/control GWA study, using 37 clinically NAD/EDM-affected horses (median 2 years; range 1–8 years; 20 females, 17 males) and 62 unaffected QHs (median 2 years; range 1–34 years; 40 females, 22 males), had a genomic inflation of $\lambda = 1.44$. Three SNPs within two regions achieved genome-wide significance (ECA8:68127182 [$\log(1/P) = 6.99$] and ECA8:62130605 and 62134644 [$\log(1/P) = 6.82$]) and one additional SNP demonstrated a suggestive association (ECA5:6862666 [$\log(1/P) = 5.21$]) (Fig. 1a). Because of the elevated genomic inflation, two mixed model analyses were implemented in GenABLE and GEMMA. With relatedness accounted for through a kinship matrix, λ was reduced to 1.05. Through both the GenABLE and GEMMA analyses, no SNPs achieved significant or suggestive associations (Fig. 1b).

Genome wide association (GWA) study: Population 2 (confirmed case/control)

Using the same set of 42,819 SNPs, the second allelic case/ control GWAS was performed using 10 post-mortem confirmed affected (1 year of age [$n = 5$], 2 [2], 3 [1], 4 [1], 7 [1]; four females and six males; two classified histologically as NAD and eight as EDM) and the same 62 unaffected QHs ($\lambda = 1.13$). Three SNPs within two regions achieved genome-wide significance (ECA8:62130605 and
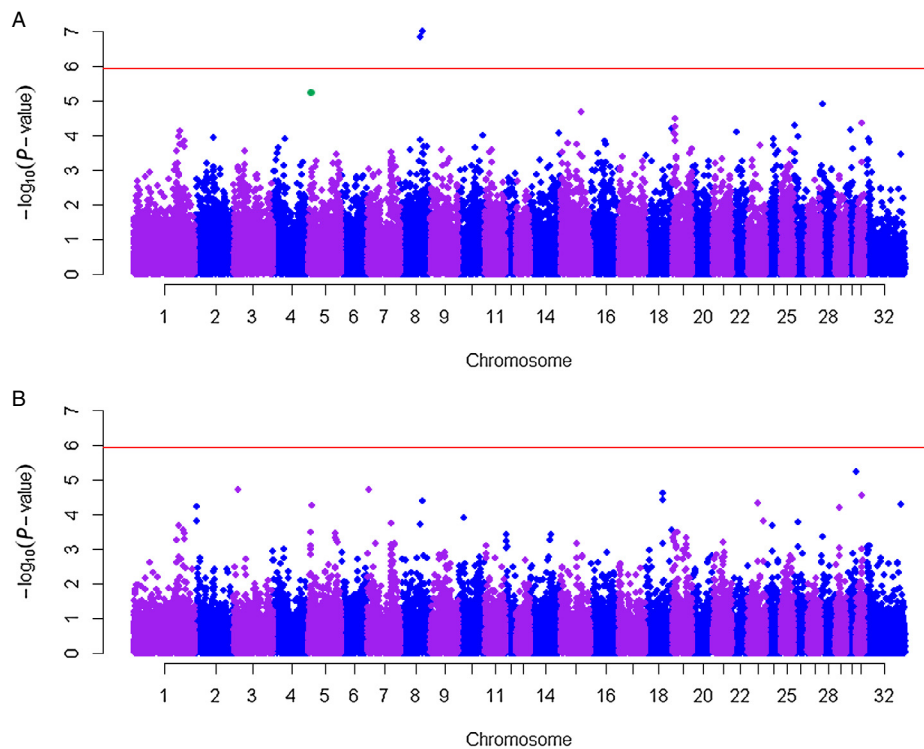

Fig. 1. (a) Unadjusted case-control association ($\lambda = 1.44$) and (b) linear mixed model ($\lambda = 1.05$) of 37 NAD/EDM clinically affected and 62 unaffected Quarter horses. Bonferroni correction for 42,691 tests (corresponding to the number of usable SNPs), defined by significant $P_{\text{genome-wide}} = 0.05$, yielded a respective thresholds of 1.17×10^{-6} (significant, log $[1/P] = 5.93$; red line). Suggestive associations, as defined by a false discovery rate at 0.10, are denoted in green.
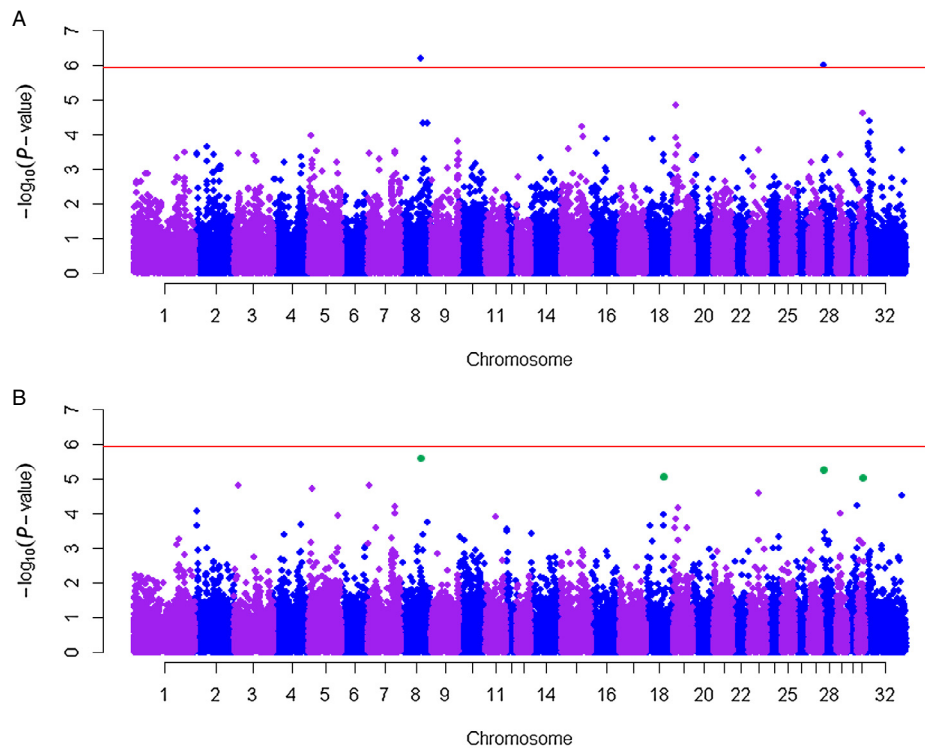
Fig. 2. (a) Unadjusted case-control association ($\lambda = 1.13$) and (b) linear mixed model ($\lambda = 1.07$) of 10 NAD/EDM post-mortem confirmed affected and 62 unaffected Quarter horses. Bonferroni correction for 42,691 tests (corresponding to the number of usable SNPs), defined by significant $P_{genome-wide} = 0.05$, yielded a respective thresholds of 1.17×10^{-6} (significant, log $[1/P] = 5.93$; red line). Suggestive associations, as defined by a false discovery rate at 0.10, are denoted in green. Suggestive associations were discovered for five SNPs (ECA8:62130705 and ECA8:62134644 [log($1/P$) = 5.57], ECA28:5813320[log($1/P$) = 5.22], ECA18_54203725 [log($1/P$) = 5.04] and ECA31:16561584[log($1/P$) = 4.98]).

62134644 [log($1/P$) = 6.18] and ECA28:5813320 [log($1/P$) = 6.01]) and there were no additional SNPs that demonstrated a suggestive association (Fig. 2a). Because of the elevated genomic inflation ($\lambda = 1.13$), two mixed model analyses were implemented in GenABLE and GEMMA. With relatedness accounted for through a kinship matrix, λ was reduced to 1.07.

Through the GenABLE analysis, the same two SNPs within one region on ECA8 yielded the highest suggestive evidence of an association (ECA8:62130605 and 62134644 [log($1/P$) = 5.57]). Additionally, three other SNPs yielded suggestive evidence of an association (ECA28:5813320 [log($1/P$) = 5.22], ECA18:54203725 [log($1/P$) = 5.04], and ECA31:16561584 [log($1/P$) = 4.98]) (Fig. 2b). Using GEMMA, the SNP at ECA28:5813320 yielded the highest suggestive association (log[$1/P$] = 5.09), followed by the SNPs at ECA8:62130605 and 62134644 (log[$1/P$] = 4.89).

Haplotype analysis

Haplotype analysis of ECA8, 28, 18 and 31 revealed a significant haplotype block ($P_{corrected} = 0.0011$) on ECA8, containing only the two associated SNPs (62130605 and 62134644). There were no significant haplotypes found on ECA28, 18 or 31.

Candidate region evaluation

Based on the single marker and haplotype association analyses, the candidate regions surrounding the associated SNPs on ECA8 (62130605 and 62134644) and ECA28:5813320 were further evaluated. Within 1 Mb of this candidate region surrounding the two most associated SNPs on ECA8, there were three genes which are expressed in the central nervous system, *PIK3C3*, *RIT2* and *SYT4*. *PIK3C3* is involved in synaptic function and mouse models have

demonstrated that defects in *PIK3C3* can lead to progressive forebrain degeneration (Wang et al., 2011). *RIT2*, also known as *RIN*, is only expressed in neural tissue (Lee et al., 1996) and is involved in synaptic endocytic trafficking (Navaroli et al., 2011). *STY4* encodes for synaptotagmin 4, a vesicle protein implicated in neurotransmitter release from neural and neuroendocrine tissues (Thomas and Elferink, 1998). The SNP at ECA28:5813320 is located within intron 11–12 of *NAV3*, a gene belonging to the neuron navigator family and has been implicated in cancer progression (Carlsson et al., 2013).

Based on the highest GWA, our knowledge of the pathophysiology of NAD/EDM and a study demonstrating abnormal accumulation of synaptic proteins, suggestive of severe axonal transport impairment in EDM-affected horses (Siso et al., 2003), we prioritized the candidate genes involved in synaptic transmission around the ECA8 SNPs for further evaluation. Because *PIK3C3* and *RIT2* were closest to the region of GWA, these genes were sequenced in both genomic and cDNA, whereas *SYT4* was only sequenced in cDNA.

Genomic DNA sequencing: PIK3C3 and RIT2

Genomic DNA from two post-mortem-confirmed NAD/EDM affected horses (age 2 years) and one post-mortem-confirmed unaffected horse (age 28 years) was used for genomic sequencing. Sequence was obtained ≥320 bp 5′ of exon 1 and encompassed the putative promoter and 5′UTR variants identified in the human assembly. Sequence was also obtained at putative splice sites, as identified in the human assembly, and included exon/intron boundaries (≥100 bp of each exon/intron boundary) and ≥340 bp 3′ to the 3′UTR. Within *PIK3C3*, a total of 77 variants (72 non-coding SNPs, one coding synonymous SNP and four insertions/deletions) were identified in the QH DNA samples relative to the published equine

genome sequence. Within *RIT2*, a total of 11 non-coding SNPs were identified (Appendix: Supplementary Table S3).

cDNA sequencing: PIK3C3, RIT2 and SYT4

There was no evidence of alternative splicing upon sequencing of cDNA for *PIK3C3*, *RIT2* and *SYT4*. The variant found within exon 22 of *PIK3C3* was confirmed in cDNA. Sequencing of cDNA for *SYT4* uncovered one synonymous SNP in exon 2 (Appendix: Supplementary Table S3). As NAD/EDM appears to be inherited as a complex trait (Finno et al., 2013), an association analysis was performed using the 88 variants discovered from sequencing *PIK3C3* and *RIT2* combined with the original GWA study genotypes. The synonymous variant in exon 2 of *SYT4* was not included in the association analysis.

Fine structure mapping

None of the identified variants within *PIK3C3* or *RIT2* achieved significance in the mixed linear model analysis (Appendix: Supplementary Table S4).

Real-time reverse transcription-PCR (q-RT-PCR)

q-RT-PCR of mRNA from caudal brainstem from six NAD-affected (1 year of age[$n = 4$], 2 years [2]) and six unaffected (1 year of age [2], 6 [1], 28 [1] and 34 [1] years of age) was performed to evaluate gene expression of *PIK3C3*, *RIT2* and *SYT4*. There were no significant changes in expression of *PIKC3* ($2^{-\Delta\Delta Ct}$ 0.95, CI 0.79–1.13; $P = 0.05$), *RIT2* ($2^{-\Delta\Delta Ct}$ 0.91; CI 0.49–1.7; $P = 0.75$) or *STY4* ($2^{-\Delta\Delta Ct}$ 1.03, CI −1.25–1.16; $P = 0.96$).

Discussion

Until recently, the majority of genetic mutations for which a molecular basis has been identified in domestic animals have been discovered using a homologous gene approach. With the sequencing of complete genomes in many domestic animals, polymorphic SNP markers have been identified and developed into SNP arrays to perform association studies for traits of interest. Based on the draft assembly of the equine genome and whole genome shotgun sequencing reads of additional horses, a SNP map of approximately one million SNPs was generated (Wade et al., 2009).

The first-generation Equine SNP50 Beadchip (Illumina) contains 54,602 SNPs (McCue et al., 2012). This SNP array was used to identify chromosomal regions containing strong candidate genes for various monogenic traits. Lavender foal syndrome was mapped to a candidate region using only six affected and 30 control horses and subsequent sequencing discovered the genetic mutation responsible for the disease (Brooks et al., 2010). The SNP50 Beadchip was also used to identify associations and led to the subsequent identification of genetic mutations for foal immunodeficiency syndrome using 14 affected, 17 known carriers and 10 ponies of unknown carrier status (Fox-Clipsham et al., 2011) and mapping of a mutation that is permissive for gaitedness in the horse using 30 'control' four-gaited horses and 40 'affected' five-gaited horses (Andersson et al., 2012).

In addition to simple monogenic traits, GWA studies using the equine SNP chips have identified quantitative trait loci for further investigation in osteochondritis dissecans (Corbin et al., 2012), recurrent laryngeal neuropathy (Dupuis et al., 2011), body size (Makvandi-Nejad et al., 2012) guttural pouch tympany (Metzger et al., 2012), equine uveitis (Kulbrock et al., 2013) and insect bite hypersensitivity (Schurink et al., 2012). Despite the apparent success in mapping more complex traits and diseases, underlying genetic mutations have not been uncovered for many of these traits.

In the original manuscript describing the sequencing of the reference horse, power estimates based on the length of linkage disequilibrium (LD; level of association between markers) in the horse, the number of haplotypes (i.e. combination of adjacent DNA sequences on a chromosome) within haplotype blocks and the polymorphism rate suggested that a more than 100,000 SNPs would be required to map traits within and across breeds (Wade et al., 2009). Based on the extent of LD in breeds such as the Quarter horse and Mongolian horse, it has been recommended that more markers are required for effective mapping in ancient breeds and those with a large effective population size (McCue et al., 2012). Therefore, the first-generation Equine SNP50 Beadchip represents about one-half of the estimated marker density required for adequately powered association studies in breeds with an average or high degree of LD.

Newer methods exist whereby linkage disequilibrium structure can be incorporated into the genomic association analysis in order to estimate the equivalent number of independent tests to allow for more exact correction for multiple testing (Zhang and Wagener, 2008). These calculations, however, are computationally intensive and were not performed in this study. With this number of markers, the effects of sample size, complexity of the trait and underlying population structure become limiting factors and may lead to false positive associations, as demonstrated in this report. Genome wide association analysis for NAD/EDM in this population of QHs revealed a region of association on ECA8 adjacent to three strong candidate genes, all expressed in the central nervous system and involved in synaptic function. Based upon direct sequencing and quantitative expression analyses, it is unlikely that uncovered variants in *PIK3C3*, *RIT2* or *STY4* are causative for NAD/EDM in the QH.

False positive associations can occur if the case and control samples arise from an admixed population (Lander and Schork, 1994) or there exists unequal familial relatedness between cases and controls (Yu et al., 2006). In this particular GWA, both population admixture and relatedness were likely causing the inflated λ of 1.13. The QH breed has a shorter LD than many other breeds and admixture has resulted in significant population substructure (McCue et al., 2012). Additionally, unequal familial relatedness was present in this population because 5 of the 10 post-mortem-confirmed NAD/EDM cases were by the same sire. To account for this stratification, two types of linear mixed models were applied, both of which incorporate a kinship matrix and phylogenetic control into the analysis. The second analysis, performed using GEMMA, was deemed necessary because recent studies have demonstrated that approximation by GRAMMAR can lead to an underestimation of all *P* values, especially in data sets in which individuals are closely related (Zhou and Stephens, 2012).

Both models improved the level of population stratification in this population (λ = 1.07) and validated the candidate regions identified with the allelic case-control analysis; however, the genome-wide associated SNPs now qualified as suggestive associations. Of note, in the GEMMA analysis the associated SNP on ECA28 was more highly associated than the two SNPs on ECA8; however, neither region achieved genome-wide significance.

Additional NAD/EDM-affected horses may have provided the power necessary to achieve genome-wide significance in the linear mixed model but the definitive phenotyping of NAD/EDM makes large sample sets of affected horses difficult to obtain. As we have demonstrated that NAD/EDM is inherited as a complex trait, there is the potential that some of the unaffected horses in this cohort had the mutated allele but did not demonstrate any clinical signs of neurological disease. All horses in this study were from one farm and were raised under identical conditions. Additionally, a previous random sampling of this group of horses demonstrated a

widespread vitamin E deficiency, as measured by serum α-tocopherol concentrations (Aleman et al., 2011). Based on these identical environmental risk factors between affected and unaffected horses, we have aimed to minimize the chances of unaffected horses containing the mutated allele with no apparent neurological phenotype.

Prioritization of candidate genes in any study requires knowledge of the pathophysiology of the disease. Axonal spheroids are the hallmark lesion of NAD/EDM and are likely to be the result of a disruption in axonal transport (Muller and Goss-Sampson, 1990). In horses with EDM, dystrophic axons demonstrated synaptophysin, synaptosomal-associated protein of 25 kDa, syntaxin-1 and α-synuclein immunoreactivity, suggesting a severe disruption of axonal transport and accumulation of these proteins in the swollen axons (Siso et al., 2003). Based on our knowledge of the pathophysiology of NAD/EDM, prioritized candidate genes should include genes involved in synaptic transmission or vitamin E and lipid transport. We have previously excluded *TTPA* as a candidate gene for NAD/EDM (Finno et al., 2013). The SNP at ECA28:5813320 is located within intron 11–12 of *NAV3*, a gene that has primarily been associated with central and peripheral nervous system tumors (Carlsson et al., 2013). Although there is no evidence that *NAV3* is involved in synaptic transmission or vitamin E and lipid transport, this particular gene may warrant further investigation in NAD/EDM.

It remains a possibility that the mutation causative for NAD/EDM resides in the associated region on ECA8 in the intergenic region between *PIK3C3* and *RIT2*, which spans approximately 740 kb. Regions of duplicated genes (Salmon Hillbertz et al., 2007) and SINE insertions (Karlsson et al., 2007) have been associated with disease phenotypes in other species. Re-sequencing of this intergenic region is necessary to completely exclude this possibility. It is more likely, however, that the region of association on ECA8 may be a false positive association due to the residual degree of genomic inflation present ($\lambda = 1.07$) (Tsepilov et al., 2013) and lack of genome-wide significance.

Conclusions

Variants in *PIK3C3*, *RIT2* and *STY4* uncovered in this study are not causative for NAD/EDM. This study highlights the importance of adequate sample size and evaluation of underlying population structure when performing GWA studies using low coverage SNP arrays in complex traits in order to identify true regions of association.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

Acknowledgements

Appendix: Supplementary material

Supplementary data to this article can be found online at doi: 10.1016/j.tvjl.2014.09.013.

References

Aleman, M., Finno, C.J., Higgins, R.J., Puschner, B., Gericota, B., Gohil, K., LeCouteur, R.A., Madigan, J.E., 2011. Evaluation of epidemiological, clinical, and pathological features of neuroaxonal dystrophy in Quarter horses. Journal of the American Veterinary Medical Association 239, 823–833.

Andersson, L., Georges, M., 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. Nature Reviews Genetics 5, 202–212.

Andersson, L.S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.J., Patra, K., Arnason, T., Wellbring, L., Hjalm, G., et al., 2012. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature 488, 642–646.

Aulchenko, Y.S., de Koning, D.J., Haley, C., 2007a. Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177, 577–585.

Aulchenko, Y.S., Ripke, S., Isaacs, A., van Duijn, C.M., 2007b. GenABEL: An R library for genome-wide association analysis. Bioinformatics (Oxford, England) 23, 1294–1296.

Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England) 21, 263–265.

Beech, J., Haskins, M., 1987. Genetic studies of neuraxonal dystrophy in the Morgan. American Journal of Veterinary Research 48, 109–113.

Benjamini, Y., Yekutieli, D., 2005. Quantitative trait Loci analysis using the false discovery rate. Genetics 171, 783–790.

Blythe, L.L., Hultgren, B.D., Craig, A.M., Appell, L.H., Lassen, E.D., Mattson, D.E., Duffield, D., 1991. Clinical, viral, and genetic evaluation of equine degenerative myeloencephalopathy in a family of Appaloosas. Journal of the American Veterinary Medical Association 198, 1005–1013.

Brooks, S.A., Gabreski, N., Miller, D., Brisbin, A., Brown, H.E., Streeter, C., Mezey, J., Cook, D., Antczak, D.F., 2010. Whole-genome SNP association in the horse: Identification of a deletion in myosin Va responsible for lavender foal syndrome. PLoS Genetics 6, e1000909.

Carlsson, E., Krohn, K., Ovaska, K., Lindberg, P., Hayry, V., Maliniemi, P., Lintulahti, A., Korja, M., Kivisaari, R., Hussein, S., et al., 2013. Neuron navigator 3 alterations in nervous system tumors associate with tumor malignancy grade and prognosis. Genes, Chromosomes and Cancer 52, 191–201.

Corbin, L.J., Blott, S.C., Swinburne, J.E., Sibbons, C., Fox-Clipsham, L.Y., Helwegen, M., Parkin, T.D., Newton, J.R., Bramlage, L.R., McIlwraith, C.W., et al., 2012. A genome-wide association study of osteochondritis dissecans in the Thoroughbred. Mammalian Genome 23, 294–303.

Dill, S.G., Correa, M.T., Erb, H.N., deLahunta, A., Kallfelz, F.A., Waldron, C., 1990. Factors associated with the development of equine degenerative myeloencephalopathy. American Journal of Veterinary Research 51, 1300–1305.

Do, D.N., Ostersen, T., Strathe, A.B., Mark, T., Jensen, J., Kadarmideen, H.N., 2014. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genetics 15, 27.

Duggal, P., Gillanders, E.M., Holmes, T.N., Bailey-Wilson, J.E., 2008. Establishing an adjusted P-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics 9, 516.

Dupuis, M.C., Zhang, Z., Druet, T., Denoix, J.M., Charlier, C., Lekeux, P., Georges, M., 2011. Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. Mammalian Genome 22, 613–620.

Finno, C.J., Famula, T., Aleman, M., Higgins, R.J., Madigan, J.E., Bannasch, D.L., 2013. Pedigree analysis and exclusion of alpha-tocopherol transfer protein (TTPA) as a candidate gene for neuroaxonal dystrophy in the American quarter horse. Journal of Veterinary Internal Medicine 27, 177–185.

Fox-Clipsham, L.Y., Carter, S.D., Goodhead, I., Hall, N., Knottenbelt, D.C., May, P.D., Ollier, W.E., Swinburne, J.E., 2011. Identification of a mutation associated with fatal foal immunodeficiency syndrome in the Fell and Dales pony. PLoS Genetics 7, e1002133.

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas, E.J., 3rd, et al., 2007. Efficient mapping of Mendelian traits in dogs through genome-wide association. Nature Genetics 39, 1321–1328.

Kulbrock, M., Lehner, S., Metzger, J., Ohnesorge, B., Distl, O., 2013. A genome-wide association study identifies risk loci to equine recurrent uveitis in German warmblood horses. PLoS ONE 8, e71619.

Lander, E.S., Schork, N.J., 1994. Genetic dissection of complex traits. Science 265, 2037–2048.

Lee, C.H., Della, N.G., Chew, C.E., Zack, D.J., 1996. Rin, a neuron-specific and calmodulin-binding small G-protein, and Rit define a novel subfamily of ras proteins. Journal of Neuroscience 16, 6784–6794.

Makvandi-Nejad, S., Hoffman, G.E., Allen, J.J., Chu, E., Gu, E., Chandler, A.M., Loredo, A.I., Bellone, R.R., Mezey, J.G., Brooks, S.A., et al., 2012. Four loci explain 83% of size variation in the horse. PLoS ONE 7, e39929.

McCue, M.E., Bannasch, D.L., Petersen, J.L., Gurr, J., Bailey, E., Binns, M.M., Distl, O., Guerin, G., Hasegawa, T., Hill, E.W., et al., 2012. A high density SNP array for the domestic horse and extant Perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genetics 8, e1002451.

Metzger, J., Ohnesorge, B., Distl, O., 2012. Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. PLoS ONE 7, e41640.

Muller, D.P., Goss-Sampson, M.A., 1990. Neurochemical, neurophysiological, and neuropathological studies in vitamin E deficiency. Critical Reviews in Neurobiology 5, 239–263.

Navaroli, D.M., Stevens, Z.H., Uzelac, Z., Gabriel, L., King, M.J., Lifshitz, L.M., Sitte, H.H., Melikian, H.E., 2011. The plasma membrane-associated GTPase Rin interacts with the dopamine transporter and is required for protein kinase C-regulated dopamine transporter trafficking. Journal of Neuroscience 31, 13758–13770.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.

Salmon Hillbertz, N.H., Isaksson, M., Karlsson, E.K., Hellmen, E., Pielberg, G.R., Savolainen, P., Wade, C.M., von Euler, H., Gustafson, U., Hedhammar, A., et al., 2007. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nature Genetics 39, 1318–1320.

Schurink, A., Wolc, A., Ducro, B.J., Frankena, K., Garrick, D.J., Dekkers, J.C., van Arendonk, J.A., 2012. Genome-wide association study of insect bite hypersensitivity in two horse populations in the Netherlands. Genetics, Selection, Evolution 44, 31.

Siso, S., Ferrer, I., Pumarola, M., 2003. Abnormal synaptic protein expression in two Arabian horses with equine degenerative myeloencephalopathy. The Veterinary Journal 166, 238–243.

Thomas, D.M., Elferink, L.A., 1998. Functional analysis of the C2A domain of synaptotagmin 1: Implications for calcium-regulated secretion. Journal of Neuroscience 18, 3511–3520.

Tsepilov, Y.A., Ried, J.S., Strauch, K., Grallert, H., van Duijn, C.M., Axenovich, T.I., Aulchenko, Y.S., 2013. Development and application of genomic control methods for genome-wide association studies using non-additive models. PLoS ONE 8, e81431.

Ullrich, B., Li, C., Zhang, J.Z., McMahon, H., Anderson, R.G., Geppert, M., Sudhof, T.C., 1994. Functional properties of multiple synaptotagmins in brain. Neuron 13, 1281–1291.

Verhoeven, K., Simonsen, K., McIntyre, L., 2005. Implementing false discovery control; increasing your power. Oikos 108, 643–647.

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al., 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326, 865–867.

Wang, L., Budolfson, K., Wang, F., 2011. Pik3c3 deletion in pyramidal neurons results in loss of synapses, extensive gliosis and progressive neurodegeneration. Neuroscience 172, 427–442.

Wu, C., DeWan, A., Hoh, J., Wang, Z., 2011. A comparison of association methods correcting for population stratification in case-control studies. Annals of Human Genetics 75, 418–427.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38, 203–208.

Zhang, F., Wagener, D., 2008. An approach to incorporate linkage disequilbrium structure into genomic association analysis. Journal of Genetics and Genomics 35, 381–385.

Zhang, F., Zhang, Z., Yan, X., Chen, H., Zhang, W., Hong, Y., Huang, L., 2014. Genome-wide association studies for hematological traits in Chinese Sutai pigs. BMC Genetics 15, 41.

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821–824.

Zhou, X., Wang, L., Hasegawa, H., Amin, P., Han, B.X., Kaneko, S., He, Y., Wang, F., 2010. Deletion of PIK3C3/Vps34 in sensory neurons causes rapid neurodegeneration by disrupting the endosomal but not the autophagic pathway. Proceedings of the National Academy of Sciences of the United States of America 107, 9424–9429.