Corbin, L. J., Blott, S. C., Swinburne, J. E., Sibbons, C., Fox-Clipsham, L. Y., Helwegen, M., ... Vaudin, M. (2012). A genome-wide association study of osteochondritis dissecans in the Thoroughbred. Mammalian Genome, 23(3-4), 294-303. DOI: 10.1007/s00335-011-9363-1

A genome-wide association study of osteochondritis dissecans in the Thoroughbred

Laura J. Corbin[1], Sarah C. Blott[2], June E. Swinburne[2], Charlene Sibbons[2], Laura Y. Fox-Clipsham[2], Maud Helwegen[2], Tim D. H. Parkin[3], J. Richard Newton[2], Lawrence R. Bramlage[4], C. Wayne McIlwraith[5], Stephen C. Bishop[1], John A. Woolliams[1] and Mark Vaudin[2].

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, Scotland, UK.

[2]Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk, CB8 7UU, UK.

[3]Boyd Orr Centre for Population and Ecosystem Health, Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, 464 Bearsden Road, Glasgow, G61 1QH, Scotland, UK (formerly [2]).

[4]Rood and Riddle Equine Hospital, 2150 Georgetown Road, Lexington, KY 40511, USA.

[5]College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO 80523, USA.

Running Head: GWAS of OCD

Key Words: equine, genome-wide association, osteochondrosis, quantitative trait loci, SNP, Thoroughbred.

Corresponding Author: Laura J. Corbin, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, UK. Phone: +44 131 651 9100, Fax: +44 131 651 9105, Email: laura.corbin@roslin.ed.ac.uk.

Abstract

Osteochondrosis is a developmental orthopaedic disease that occurs in horses, other livestock species, companion animal species and humans. The principal aim of this study was to identify quantitative trait loci (QTL) associated with osteochondritis dissecans (OCD) in the Thoroughbred using a genome-wide association study. A secondary objective was to test the effect of previously identified QTL in the current population. Over three hundred horses, classified as cases or controls according to clinical findings, were genotyped for the Illumina Equine SNP50 BeadChip. An animal model was first implemented in order to adjust each horse's phenotypic status for average relatedness among horses and other potentially confounding factors which featured in the data. The genome-wide association test was then conducted on the residuals from the animal model. A single SNP on chromosome 3 was found to be associated with OCD at a genome-wide level of significance, as determined by permutation. According to the current sequence annotation, the SNP is located in an intergenic region of the genome. The effects of 24 SNPs, representing QTL previously identified in a sample of Hanoverian Warmblood horses, were tested directly in the animal model. When fitted alongside the significant SNP on ECA3, two of these SNP were found to be associated with OCD. Confirmation of the putative QTL identified on ECA3 requires validation in an independent sample. The results of this study suggest that a significant challenge faced by equine researchers is the generation of sufficiently large datasets to effectively study such complex diseases as osteochondrosis.

Introduction

Osteochondrosis (OC) is a disease of the locomotory system which affects the joints of many animals, most frequently being observed in pigs, horses and dogs. Osteochondrosis can be described as a focal disturbance of endochondral ossification (Ytrehus et al. 2007) that occurs in young, growing individuals and as such has been classified as a developmental orthopaedic disease. Primary lesions, thought to be initiated by a failure of blood supply to the cartilage (Ytrehus et al. 2007), progress to form retained cores of cartilage eventually causing dissecting lesions on the joint surface (McIlwraith 2011). In its early stages, the condition has been referred to as dyschondroplasia or more recently osteochondrosis latens (Ytrehus et al. 2007) and is likely to be subclinical in nature. In the most serious cases, where cartilage or subchondral bone fragments become separated from the articular surface, introducing an inflammatory component, the disease may be referred to as osteochondritis dissecans (OCD). In such cases, typical clinical signs of the disease are synovitis and pain accompanied by varying degrees of lameness (McIlwraith 2011). In the horse, joints most commonly affected are the fetlock, hock, and stifle; within these joints specific predilection sites have been identified (McIlwraith 1993). Prevalence estimates for OC vary widely, ranging from 3% (stifle OC in Thoroughbreds (Oliver et al. 2008)) to 70% (estimates for all joints in Dutch Warmbloods (van Grevenhof et al. 2009a)). A large proportion of this variation is attributable to differences in the type and number of anatomical locations examined, differences in the specific manifestation of the disease considered and to breed differences (Philipsson et al. 1993; Pieramati et al. 2003; van Grevenhof et al. 2009a; Wittwer et al. 2006). A recent prevalence estimate of 25% for the Thoroughbred (Lepeule et al. 2009) appears

typical. This relatively high disease prevalence, along with the likely contribution of OC to the predominance of lameness as a cause of wastage in young horses (Olivier et al. 1997; Rossdale et al. 1985), makes OC a high priority for study.

Whilst there exists both experimental and anecdotal evidence of a genetic component to OC, the aetiopathogenesis of the disease is not fully understood (Ytrehus et al. 2007). The disease is considered multifactorial in origin with at least some evidence of both environmental factors, for example nutrition, and physiological factors, such as growth and body size, endocrine factors and conformation, which may themselves be mediated through genetics, playing a role in the condition (Lepeule et al. 2009; McIlwraith 2004; van Weeren et al. 1999). Low to moderate estimates of heritability for OC across a range of breeds and disease manifestations (Philipsson et al. 1993; Pieramati et al. 2003; Schougaard et al. 1990; van Grevenhof et al. 2009b; Wittwer et al. 2007a) together with between breed differences in prevalence (Lepeule et al. 2009) indicate that genetic variability exists in disease susceptibility. Typical values for OC scored as a single binary trait (all joints combined) are 0.10 to 0.20 (Pieramati et al. 2003; Wittwer et al. 2007a) but heritability estimates of up to 0.5 have been reported for individual joints (Grøndahl and Dolvik 1993).

The search for markers to explain the proposed genetic variance in susceptibility to OC began several years ago, with the intention both of enhancing our understanding of the condition and of enabling marker assisted selection. Early studies using primarily linkage based analyses (dependent on family data), to detect regions of the genome associated with OC in the horse have identified several putative quantitative trait loci (QTL) (Dierks et al. 2007; Wittwer et al. 2007b). As is typical for QTL

discovered using this approach, their effects are generally large but their locations are imprecise. Whilst several of these QTL have undergone further refinement, very few have been validated in independent data sets. Similar studies in pigs have revealed few (Andersson-Eklund et al. 2000) or no (Lee et al. 2003) QTL for osteochondrosis. These results illustrate the difficulty in identifying truly associated regions for complex traits using linkage analysis.

The opportunity for QTL studies in horses has recently been advanced by the publication of the equine genome sequence (Wade et al. 2009) together with the release of the Illumina Equine SNP50 BeadChip, which has allowed the implementation of genome-wide association studies (GWAS). In contrast to linkage analysis, GWAS rely on samples of individuals, which may be unrelated, genotyped at medium to high density. It is expected that this approach will allow the identification of common variants which could not be found using the traditional linkage based approach (Iles 2008). We are aware of four GWAS for OC that have been carried out in three different horse breeds to date: Lampe (2009) and Komm (2010) (using the same data), Teyssèdre et al. (2010) and Lykkjen et al. (2010). The number of QTL identified per study ranges from four (Lykkjen et al. 2010) to 18 (Lampe 2009) with the range likely at least partly attributable both to differences in significance thresholds used and to differing phenotype definitions. A single putative correspondence between QTL has been described (Lykkjen et al. 2010).

This study demonstrates the use of clinical observations as a source of data for use in genomic studies and is the first QTL mapping study for OC to be conducted in the Thoroughbred. A GWAS was performed on 348 samples using the Illumina Equine

SNP50 BeadChip to identify loci associated with OCD in the Thoroughbred. In addition, QTL for OC previously identified in a Hanoverian Warmblood (HWB) population were tested for their effect in the current data set.

Materials & Methods

Sample Collection

Blood samples were collected over two years (2007/2008) from 348 Thoroughbreds (159 males, 189 females) classified either as cases (169) or controls (179) for OC. Horses were admitted for surgery to the Rood and Riddle Equine Hospital, Lexington, Kentucky, at age nine to twelve months. Horses originated from one of 19 surrounding horse farms. The number of horses per farm ranged from two to 89, with approximately equal numbers of cases and controls sourced from each farm (see Figure 1). Management of the horses, including feeding, housing and exercise levels are expected to vary by farm. Due to the anonymity of samples, pedigree details for the horses were not available but the sample is expected to comprise a mixture of half sibs (by sire and dam since data was collected across two years) and more distantly related horses.

Osteochondrosis case samples (n=169; 90 males, 79 females) consisted of horses which were diagnosed as having OC requiring surgery in at least one joint from radiographic surveys performed by referring veterinarians (see Supplementary Material, Table S1, for further details). The diagnosis was then confirmed through repeat radiography of suspected OC affected regions on the admission of the horses to the equine hospital. In order to be considered for surgery, cartilage and/or bone fragments separated from the articular surface would have to be present, and therefore our cases can be considered as suffering specifically from osteochondritis dissecans (OCD). Subsequent arthroscopic surgeries were performed by L. R. Bramlage. A typical arthroscopic surgery involves the removal of all fragments and the

debridement of any separated articular cartilage and defective bone (McIlwraith 2002). Horses were affected in at least one of the following joints: fetlock (24.9%), hock (56.2%), stifle (29.6%), shoulder (0.6%). The total number of joints affected per horse ranged from 1 to 5.

Control samples (n=179; 69 males, 110 females) consisted of horses which were admitted to the hospital for surgical procedures other than OC, most commonly the insertion of a transphyseal bridge to address angular limb deformities (ALD), the arthroscopic removal of osteochondral fractures of the proximal (first) phalanx in the fetlock joint (fetlock chips) and the treatment of sesamoid fractures (see Supplementary Material, Document S1, for further details). Many case horses also underwent these procedures (see Table 1). All control horses were clear from signs of OC, as determined by a full radiographic survey (as in cases) prior to surgery.

Genotyping

Blood samples were collected in ethylenediaminetetraacetic acid and DNA extracted either by Tepnel (http://www.tepnel.com/dna-extraction-service.asp) or at the Animal Health Trust using Nucleon BACC DNA extraction kits (http://www.tepnel.com/dna-extraction-kits-blood-and-cell-culture.asp). A small dilution of each sample was prepared at 70ng/ul and submitted for genotyping to Cambridge Genomic Services (http://www.cgs.path.cam.ac.uk/services/snp-genotyping/services.html). The Illumina Equine SNP50 Genotyping BeadChip (www.illumina.com/documents/products/datasheets/datasheet_equine_snp50.pdf), which comprises 54,602 single nucleotide polymorphisms (SNP) located across all autosomes and the X chromosome, was used. These were selected from the database

of over one million SNP (http://www.broadinstitute.org/ftp/distribution/horse_snp_release/v2/) generated during the sequencing of the horse genome (http://www.broadinstitute.org/mammals/horse). Samples for this study were genotyped alongside samples for several other studies and the full genotyped dataset was inspected using the Illumina GenomeStudio genotyping module and a series of quality control metrics used to identify poorly performing SNP. Quality control (QC) at this stage led to the removal of 7.1% of the SNPs (n=3,895) due to poor genotyping quality (see Table S2 in Supporting Material). These SNPs were set to missing prior to the commencement of quality control for this study.

Quality Control (QC) for Data Analyses

Firstly, samples were checked for sex discrepancies (marker-based prediction of sex versus sample label) and intermediate X-chromosomal inbreeding ($0.2<F<0.8$), with exclusions being made on the basis of suspected sampling or genotyping errors. This process resulted in two exclusions due to sex discrepancy and 16 exclusions based on indeterminate sex as demonstrated by intermediate inbreeding, leaving 168 controls and 162 cases for further analysis.

For the GWAS (see below) the following thresholds were used for excluding data: minor allele frequency (<0.05), missing genotypes per SNP (>5%), missing SNP per sample (>5%) and differential proportions of missing SNPs between cases and controls ($p<0.05$). No exclusions were made on the basis of Hardy-Weinberg equilibrium (HWE).

For construction of a marker-based relationship matrix (see below), a subset of markers meeting more stringent QC was chosen as recommended by (Yang et al. 2011) with exclusions made as follows: minor allele frequency (<0.10), missing genotypes per SNP (>0.5%) missing genotypes per sample (>1%) and HWE ($p<0.05$).

Mixed Model Analysis

Binary case/control phenotypes were adjusted for fixed and random effects using the following linear mixed model in ASReml (Gilmour et al. 2009). A single categorical fixed effect was fitted which represents the division of samples into contemporary groups relating to the three most common reasons for surgery, other than OCD, listed in Table 1 (ALD, fetlock chip(s) and sesamoid fracture(s)) and sex, resulting in 2^3 x 2 = 16 classes in total, 11 of which contained observations in the final analysis (see Table S3 in Supplementary Material). A single random effect, animal, was fitted generating an individual animal model (Henderson 1975) in which the pedigree relationship matrix was replaced with a marker-based relationship matrix (G-matrix) in order to adjust for average allele sharing among sampled horses. Autosomal markers remaining after QC were used to generate the G-matrix as follows:

$f_{i,j} = \frac{1}{N}\sum_k \frac{(x_{i,k} - p_k)(x_{j,k} - p_k)}{(p_k(1 - p_k))}$, where summation is across SNPs ($k=1,N$), x_{ik} is a genotype of the i^{th} horse at the k^{th} SNP coded as 0, ½, 1 and p_k is the frequency of the allele that is homozygous for the genotype coded as 1 (Aulchenko et al. 2007). On the diagonal, $f_{i,i} = 0.5(1 + f_i)$, where f_i is the loss (or gain) of heterozygosity relative to the expectation. The relationship matrix describes the average relatedness between individuals and therefore controls for genetic stratification likely to be present in the sample. The transformation of the G-matrix into a distance matrix followed by a

multi-dimensional scaling (MDS) analysis (Cailliez 1983; Cox and Cox 1994; R Development Core Team 2009), also allowed data to be inspected for the presence of outliers and substructure. MDS plots based on the first two principal components were considered with respect to farm of origin, sex and contemporary group. Following the implementation of the mixed model, a vector of approximately normally distributed (N~(0,1)) residual errors replaced our binary (0,1) observation as the phenotype for testing in the GWAS.

Genome-Wide Association Study

GWAS was performed in GenABEL (Aulchenko et al. 2007) using a score test for a Gaussian distributed trait and no covariates (Schaid et al. 2002). A genome-wide significance level was calculated by performing 10,000 permutations of the residual phenotypes against genotypes. Permutations were carried out within sex, and the 5% significance level empirically determined. Confirmation of the effects of SNPs found to be significant by this approach was carried out by fitting all such SNP genotypes (coded as 0, 1, 2) simultaneously as fixed effects in the original mixed model.

Testing Previously Published QTL

SNPs selected to represent OC QTL detected in other studies were also tested by fitting them simultaneously as fixed effects in the mixed model. The QTL regions tested were based primarily on GWAS results published in Lampe (2009) and Komm (2010). These studies were performed on samples from HWB horses and it has been shown in a reference sample of more than 150,000 horses that the Thoroughbred contributes nearly 35% of this breed's genes (Hamann and Distl 2008). Whilst these studies examined a range of OC phenotypes, we tested only QTL relevant to OC or

OCD with fetlock and hock cases combined, as here (see Table S4 in Supplementary Material for a list of QTL). Where SNP names or precise SNP locations were provided, the exact SNP was fitted in the mixed model with the exception of one case where the SNP was not typed in our sample, in this case the closest SNP was fitted in the mixed model (type A in Table S4). In cases where only an approximate location was given, i.e. to the nearest 0.1Mb, current GWAS results for the region 1Mb upstream and 1Mb downstream were examined and the SNP with the smallest p-value fitted in the mixed model (type B in Table S4). Finally, in cases where several SNPs within a region were listed as being significant, the same range was searched in the current GWAS analysis and the SNP with the smallest p-value fitted in the mixed model (type C and D in Table S4). In order to assess their ability to enhance our model, all SNPs representing QTL were fitted simultaneously alongside contemporary group, SNPs found to be significant in the current GWAS and the G-matrix in the mixed model.

Results

Mixed Model Analysis

The genomic relationship matrix was calculated based on 30,554 autosomal SNPs that passed the stringent QC thresholds. The distribution of genomic relationships between individuals in the sample is shown in Figure 2. MDS plots revealed no obvious outliers or any genetic substructure relating to factors such as farm or contemporary group (data not shown). The fitting of the mixed model resulted in an extremely small estimated genetic variance component ($<10^{-7}$) making it impossible to estimate trait heritability with any precision; estimates of random animal effects (estimated additive breeding values) were correspondingly small (-5.8×10^{-8} to 6.5×10^{-8}). Therefore, the residuals generated for testing in the association study were influenced primarily by contemporary group. The distribution of residuals can be seen in Figure 3.

Genome-Wide Association Study

Following QC, 40,180 SNPs were tested for association; the mean minor allele frequency of remaining SNP was 0.28 and the distribution of minor allele frequencies was approximately uniform. Based on empirical genome-wide significance ($p<2.91 \times 10^{-6}$), a single SNP was found to be significantly associated with OCD as tested using residuals from the mixed model. This was SNP BIEC2-799865 located at 88,493,417bps on ECA3; this SNP has alleles C and T with a minor allele frequency (T) of 0.4 and conforms to a HWE genotype distribution (see Table 2 for genotype frequencies). Figure 4 shows a Manhattan plot of SNPs on ECA3. A haplotype block analysis of the region containing BIEC2-799865 revealed somewhat erratic linkage

disequilibrium (LD) structure surrounding the SNP making the definition of an associated QTL region problematical (Figure 5). The apparent deviation from the expectation of decreasing LD with increasing distance between markers exhibited by BIEC2-799865 and its neighbours goes someway to explaining why this SNP stands apart from surrounding SNP in Figure 4. With SNPs exhibiting r^2 (Purcell 2009; Purcell et al. 2007) with BIEC2-799865 of greater than 0.10 at distances up to 10Mb, we extended our search for other potentially associated SNPs within this range. A further four SNPs within 10Mb of BIEC2-799865 had $p<0.001$; two of these SNPs had r^2 of 0.45 – 0.55 with and were within 3 SNPs of BIEC2-799865 (see Figure 5), with the remainder being >4Mb away and having $r^2 < 0.10$. All four SNPs were located to the right of BIEC2-799865.

Fitting BIEC2-799865 as an additional covariate in the mixed model resulted in an estimated additive effect of -0.16 (±0.03), i.e. for every T allele an individual carries at the locus, that individual's probability of OCD is decreased by 0.16. This allows us to make a crude estimate of the contribution of this SNP to the overall phenotypic variance. Under the assumption of no dominance or interaction effects and using $V_A = 2p(1-p)\alpha^2$ (Falconer and Mackay 1996) where p is allele frequency at the locus and α is the estimated SNP effect, BIEC2-799865 explains ~5% of the variance of OCD. The effect of BIEC2-799865 remained significant even when contemporary group was removed from the mixed model. Fitting the additional four SNPs with $p<0.001$ alongside contemporary group and BIEC2-799865 resulted in both BIEC2-799865 and one of the more distant SNPs (BIEC2-802230), having regression coefficients significantly different from zero.

Testing Previously Published QTL

For each of the 24 QTL regions listed in Table S4, a representative SNP was added to the mixed model containing contemporary group, BIEC2-799865 and the random effect of animal so that all SNPs were analysed simultaneously. This analysis resulted in only two of the 24 SNPs having a significant association with OCD. These SNPs were BIEC2-859811 on ECA4 (39,852,072), representing a QTL at 39.26Mb (Table S4, QTL no. 8) (Komm 2010) and BIEC2-410967 on ECA18 (36,772,271), representing a QTL between 36,408,881 and 38,738,316 (Lampe 2009) (Table 4, QTL no. 16). BIEC2-799865 remained significant when fitted alongside the 24 QTL SNPs albeit with a slightly reduced size of effect (-0.11).

Discussion

This GWAS in the Thoroughbred revealed a single SNP, BIEC2-799865 on ECA3, to be associated with OCD at a genome-wide level of significance when tested using the residuals from a mixed model analysis. Population genetics theory allows us to predict that, assuming the heritability for OCD is 0.15, this QTL accounts for ~34% of the genetic variation of the trait. However, effect estimates based on primary GWAS data have been shown to be upwardly biased, often to a large degree (Göring et al. 2001) and so a majority of the genetic variance underlying OCD remains to be captured. Two neighbouring SNPs showed an association with OCD which approached significance ($p<0.001$); the relatively lower MAF of these SNPs (0.27 and 0.25) compared to that of BIEC2-799865 (0.4) may explain their failure to reach genome-wide significance. The lack of haplotype block structure around BIEC2-799865, means that the much sought after and characteristic GWAS peak is not observed in this case. Whilst the implication of this on the validity of the association is not clear, it does impact on our ability to precisely define a corresponding QTL region for further evaluation. Although it would have been desirable to fit haplotypes representing the associated region in our model, the low LD in the region hindered our ability to accurately infer phase. For the purposes of candidate gene discovery we chose to examine the region 1Mb either side of the SNP.

The 2Mb window surrounding BIEC2-799865 contained 22 labelled genes, 21 of which are described as protein coding and one of which is labelled as a pseudo gene. Whilst, according to the current annotation BIEC2-799865 lies between genes, LOC100064680 located at 88,494,283 - 88,511,285 bps contains (within an intron)

BIEC2-799867, the SNP which is both adjacent to and most highly correlated with BIEC2-799865. This gene is described as being similar to basic kruppel like factor and studies in mice and *C.elegans* show orthologues to this gene, *kruppel-like factor 3 (basic)* (*KLF3*), to be involved in adipogenesis (Sue et al. 2008; Zhang et al. 2009). More generally, KLFs have been described as DNA binding transcriptional regulators that play diverse roles during differentiation and development (Bieker 2001). Whilst the likely function of *KLF3* does not preclude its relevance, there is no evidence of a direct role for this gene in OC. This was true of most of the genes located within the QTL region defined, with the exception of *UDP-glucose dehydrogenase* (*UGDH*).

The *UGDH* gene (located at 87,818,121 – 87,843,937 bps) appears to function in the regulation of glycosaminoglycan (GAG) synthesis in cells lining the articular cartilage surface (Clarkin et al. 2011). These GAG are involved in extra-cellular matrix integrity, playing a crucial role in chondrogenesis, homeostasis and compressive resilience (Clarkin et al. 2011). A potential link between GAG and osteochondrosis has been demonstrated by the observation of differential levels of GAG in osteochondritic lesions versus healthy cartilage (Kuroki et al. 2002; Lillich et al. 1997). However, the direction of causality is not clear and several other studies have observed no significant difference (Bertone et al. 2005; de Grauw et al. 2006). Two SNPs located within introns of *UGDH* were not significantly associated with OCD ($0.05<p<0.10$). One of these SNPs did, however, show moderate LD ($r^2 = 0.1 – 0.2$) with BIEC2-799865 and the two neighbouring SNPs mentioned above (see Figure 5); as before, the relatively lower MAF of this SNP (0.34) may have prevented it from appearing above the background in terms of significance. The second SNP in *UGDH* had a MAF of 0.06 and therefore provides little information about either association

or LD. Whilst the distance of this gene from BIEC2-799865 and its relatively low LD with the SNP question its relevance, there are likely to be many untyped variants in this region, some of which could plausibly have stronger LD with BIEC2-799865.

Three previous GWAS for OC in the horse have also identified QTL on ECA3 (Komm 2010; Lampe 2009; Teyssèdre et al. 2010). The closest to BIEC2-799865 was presented recently in a pre-print version of a study carried out in French Trotters and is located at 100-110Mb (Teyssèdre et al. 2011). The relatively close proximity of the two QTL represents some correspondence between studies. However, with average LD at this distance (~12Mb) being r^2<0.02 (Corbin et al. 2010), it is also possible that these QTL represent two different underlying genetic variants.

Adding SNPs to represent previously identified QTL to our model (which included BIEC2-799865) resulted in two out of 24 SNPs tested having regression coefficients significantly different from zero ($p<0.05$) and therefore showing the potential to enhance the fit of the model. On ECA4, BIEC2-859811 (39,852,072) had a regression coefficient of -0.102 (±0.049). Komm (2010) identified six candidate genes located between 37.1Mb and 44.7Mb. On ECA18, BIEC2-410967 (36,772,271) had an estimated effect size of -0.085 (±0.042). Lampe (2009) identified three candidate genes in the vicinity of the QTL corresponding to this SNP. These apparent validations should however be viewed with caution since adjustments to the mixed model, for example the removal of BIEC2-799865, lead to different QTL being significant and we were therefore unable to unambiguously confirm any of the previous QTL in the current dataset.

There are several reasons for the poor correspondence between QTL studies of OC in the horse. Firstly, the QTL which have been identified to date may be false positives (McCarthy et al. 2008). Alternatively, subsequent studies may have been underpowered to detect them. In this case, such results may be due to, for example, differences in phenotypic definition or population ancestry. Ideally, replication studies should involve precisely the same allele or haplotype, the same phenotype and the same genetic model as the original signal (Weedon et al. 2008). In this study, by testing only the QTL regions associated with OC under the combined phenotype definition (hock and fetlock) used by Lampe (2009) and Komm (2010), the difference in phenotypic definition between the three studies was minimised.

Another reason for the lack of correspondence may be breed differences. Hamann et al. (2008) estimated that 35% of the HWB genes came from Thoroughbred lines, but it is not known what the proportion was in the Komm (2010) and Lampe (2009) sample of 154 foals. Assuming the same QTL are controlling the genetic predisposition to OC in both breeds, differences in allele and haplotype frequencies between breeds will impact on the proportion of variance the QTL explain and therefore on our ability to detect them. Furthermore, with no standardised method either for reporting QTL or for carrying out validation studies, the approach taken here to select SNPs for testing in the mixed model was largely subjective and we may have missed more appropriate SNPs.

Despite being one of the largest GWAS of OC in horses performed to date, the principal limitation of this study remains a lack of power. This lack of power is evidenced by both the low number of genome-wide significant SNP and the very

small estimated genetic component. Whilst disappointing, our inability to estimate heritability in this sample is perhaps not surprising given the relatively large standard errors which accompany some of the heritability estimates for OC to date (Pieramati et al. 2003; Wittwer et al. 2007a). Furthermore, our findings do not necessarily rule out a non-zero heritability, rather more data is needed to produce a reliable estimate.

The explanation for the apparent low power of this study is likely to be multifaceted. Firstly, since power is directly related to sample size, the relatively small number of horses genotyped for this study will have limited the number of identifiable QTL, as shown by power calculations of, for example, Wang et al. (2005). Secondly, phenotypic definition can play an important role in determining the power of GWAS of complex diseases. Optimal phenotypic definitions are those with strict inclusion criteria, with minimising genetic heterogeneity between cases being a useful way of increasing study power (McCarthy et al. 2008). Unfortunately, OC represents a clinically complex phenotype, affecting multiple joints and predilection sites within joints, as well as appearing in a variety of different forms. Just as prevalence and heritability estimates for OC have been affected by this problem, so we can expect QTL mapping studies to be. In this study, by considering exclusively those cases with fragments present (OCD), the genetic heterogeneity of the cases has been reduced and we are also following recommendations by van Grevenhof et al. (2009b) that flattened bone contours and fragments should be evaluated as statistically different disorders.

Several studies to date have considered further subdivision of OC cases by joint affected, resulting in different QTL being identified for each subgroup (Dierks et al. 2007; Wittwer et al. 2007b). This is appealing given the apparent low correlation

among the occurrence of lesions of OC in different body locations (Jorgensen and Andersen 2000; Jorgensen et al. 1995; van Grevenhof et al. 2009b) and the corresponding idea that OC is in fact a localised disease (Ytrehus et al. 2007). However, subdividing cases in this way represents a significant loss of power. Furthermore, testing several manifestations of the disease serves to exacerbate the already serious problem of multiple testing. For this reason and from a practical selection perspective, expressing OC as a single trait is more appealing, and should enable the identification of QTL controlling more generalised factors.

In this study, model complexity due to the presence of horses suffering from conditions other than OC in our cohort may have reduced the power of our association test. The uneven representation of cases and controls across the contemporary groups describing the presence or absence of ALD, fetlock chips and sesamoid fractures in our samples, represented a potential cause of bias in the sample and therefore had to be fitted in the model. In the event that none of these conditions are related to OC or have a hereditary component, our adjustment for contemporary group represents a loss of power through the reduction in the number of degrees of freedom of the model. However, in the case where one or more of these diseases has a hereditary component (of which there is some evidence (Philipsson et al. 1993; Wittwer et al. 2007a)), the exclusion of contemporary group from the model would result in severe confounding. Since the latter is by far the more serious case, we chose to fit contemporary group in the mixed model.

However, there is seemingly a trade-off to be made. Whilst the use of clinical data in this case added complexity and potentially noise to the data, it also gave us increased

confidence in our phenotypic classifications of OCD. In this study, all of our cases underwent arthroscopy, the so-called 'gold standard' of diagnosis of cartilage defects (McIlwraith 2010) and so we can be confident of high specificity. All of the controls had OC ruled out through a comprehensive radiographic survey of predilection sites and the evaluation of radiographs by a specialist in the field (LRB) significantly reduced the chance of OC going undiagnosed.

In this GWAS we identified a SNP associated with OCD in a sample of 330 Thoroughbreds. This association requires validation in an independent dataset in order to rule out the possibility that it represents a false positive association. In the event that the SNP is validated, further fine-mapping and re-sequencing of the region will be needed in order to elucidate the causal mutation behind this association. The likely issue of poor power to detect QTL in this study illustrates the challenge faced by members of the equine genetics community in collecting and genotyping sufficiently large samples for effective GWAS to be carried out. Here we have demonstrated the potential for clinical data to be utilised as a source of samples for the future.

Acknowledgments: LJC would like to thank A. Tenesa and R. Pong-Wong for helpful discussions and S. Miller for help with data preparation. Samples from OCD cases and controls were provided by LRB. LJC would like to thank two anonymous referees for their helpful comments. LJC, JAW and S. C. Bishop are financially supported by the British Equestrian Federation, the Biosciences Knowledge Transfer Network and the Biotechnology and Biological Sciences Research Council (BBSRC).

S. C. Blott, JES, CS, LYF-C, MH, TDHP and the genotyping were funded by the Horserace Betting Levy Board and the Thoroughbred Breeders' Association.

References

Andersson-Eklund L, Uhlhorn H, Lundeheim N, Dalin G, Andersson L (2000) Mapping quantitative trait loci for principal components of bone measurements and osteochondrosis scores in a wild boar x Large White intercross. Genet Res 75, 223-230

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23, 1294-1296

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265

Bertone AL, Bramlage LR, McIlwraith CW, Malemud CL (2005) Comparison of proteoglycan and collagen in articular cartlage of horses with naturally developing ostechondrosis and healing osteochondral fragments of experimentally induced fractures. Am J Vet Res 66, 1881-1890

Bieker JJ (2001) Krüppel-like Factors: Three Fingers in Many Pies. J Biol Chem 276, 34355-34358

Cailliez F (1983) The analytical solution of the additive constant problem. Psychometrika 48, 305-308

Clarkin CE, Allen S, Kuiper NJ, Wheeler BT, Wheeler-Jones CP, Pitsillides AA (2011) Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and release, control sulfated GAG synthesis, and promote chondrogenesis. J Cell Physiol 226, 749-761

Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Anim Genet 41, 8-15

Cox TF, Cox MAA (1994) Multidimensional Scaling. (London: Chapman and Hall)

de Grauw JC, Brama PA, Wiemer P, Brommer H, van de Lest CA, van Weeren PR (2006) Cartilage-derived biomarkers and lipid mediators of inflammation in horses with osteochondritis dissecans of the distal intermediate ridge of the tibia. Am J Vet Res 67, 1156-1162

Dierks C, Lohring K, Lampe V, Wittwer C, Drogemuller C, Distl O (2007) Genome-wide search for markers associated with osteochondrosis in Hanoverian warmblood horses. Mamm. Genome 18, 739-747

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4 ed. Malaysia: Pearson Education Limited, pp.126

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296, 2225-2229

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide Release 3.0. (Hemel Hempstead, UK: VSN International Ltd.)

Göring HHH, Terwilliger JD, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. Am. J. Hum. Genet. 69, 1357-1369

Grøndahl AM, Dolvik NI (1993) Heritability estimation of osteochondrosis in the tibiotarsal joint and of bony fragments in the palmar/plantar portion of the metacarpo- and metatarsophalangeal joints of horses. J Am Vet Med Assoc 203, 101-104

Hamann H, Distl O (2008) Genetic variability in Hanoverian warmblood horses using pedigree analysis. J Anim Sci 86, 1503-1513

Henderson CR (1975) Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics 31, 423-447

Iles MM (2008) What can genome-wide association studies tell us about the genetics of common disease? PLoS Genet 4, e33

Jorgensen B, Andersen S (2000) Genetic parameters for osteochondrosis in Danish Landrace and Yorkshire boars and correlations with leg weakness and production traits. Animal Science 71, 427-434

Jorgensen B, Arnbjerg J, Aaslyng M (1995) Pathological and radiological investigations on osteochondrosis in pigs, associated with leg weakness. Journal of Veterinary Medicine Series A 42, 489-504

Komm K (2010) Fine mapping of quantitative trait loci (QTL) for osteochondrosis in Hanoverian warmblood horses. PhD Thesis, University of Veterinary Medicine Hannover

Kuroki K, Cook JL, Tomlinson JL, Kreeger JM (2002) In vitro characterization of chondrocytes isolated from naturally occurring osteochondrosis lesions of the humeral head of dogs. Am J Vet Res 63, 186-193

Lampe V (2009) Fine mapping of quantitative trait loci (QTL) for osteochondrosis in Hanoverian warmblood horses. PhD Thesis, University of Veterinary Medicne Hannover

Lee GJ, Archibald AL, Garth GB, Law AS, Nicholson D, Barr A, Haley CS (2003) Detection of quantitative trait loci for locomotion and osteochondrosis-related traits in Large White X Meishan pigs. Animal Science 76, 155-165

Lepeule J, Bareille N, Robert C, Ezanno P, Valette JP, Jacquet S, Blanchard G, Denoix JM, Seegers H (2009) Association of growth, feeding practices and exercise conditions with the prevalence of Developmental Orthopaedic Disease in limbs of French foals at weaning. Prev Vet Med 89, 167-177

Lillich JD, Bertone AL, Malemud CJ, Weisbrode SE, Ruggles AJ, Stevenson S (1997) Biochemical, histochemical, and immunohistochemical characterization of distal tibial osteochondrosis in horses. Am J Vet Res 58, 89-98

Lykkjen S, Dolvik NI, McCue ME, Rendahl AK, Mickelson JR, Roed KH (2010) Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. Anim. Genet. 41, 111-120

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9, 356-369

McIlwraith CW (1993) Inferences from referred clinical cases of osteochondritis dissecans. Equine Vet J 25, 27-30

McIlwraith CW (2002) Arthroscopic surgery for osteochondral chip fragments and other lesions not requiring internal fixation in the carpal and fetlock joints of the equine athlete: What have we learned in 20 years? Clin Tech Equine Pract 1, 200-210

McIlwraith CW (2004) Developmental orthopedic disease: problems of limbs in young horses. J Equine Vet Sci 24, 475-479

McIlwraith CW (2010) Recent Advances in Diagnosis of Equine Joint Disease. In: Proceedings of the 17th Kentucky Equine research Nutrition Conference. Lexington, KY, pp 23-33

McIlwraith CW (2011) Lameness in the Young Horse: Osteochondrosis. In: Baxter, GM (ed) Adams and Stashak's Lameness in Horses Iowa Sate University Press, Ames, pp 1155-1164

Oliver LJ, Baird DK, Baird AN, Moore GE (2008) Prevalence and distribution of radiographically evident lesions on repository films in the hock and stifle joints of yearling Thoroughbred horses in New Zealand. N Z Vet J 56, 202-209

Olivier A, Nurton JP, Guthrie AJ (1997) An epizoological study of wastage in Thoroughbred racehorses in Gauteng, South Africa. J S Afr Vet Assoc 68, 125-129

Philipsson J, Andréasson E, Sandgren B, Dalin G, Carlsten J (1993) Osteochondrosis in the tarsocrural joint and osteochondral fragments in the fetlock joints in Standardbred trotters. II. Heritability. Equine Vet J 25, 38-41

Pieramati C, Pepe M, Silvestrelli M, Bolla A (2003) Heritability estimation of osteochondrosis dissecans in Maremmano horses. Livestock Prod Sci 79, 249-255

R Development Core Team (2009) R: A Language and Environment for Statistical Computing (v2.10.0). R Foundation for Statistical Computing, Vienna. http://www.R-project.org

Rossdale PD, Hopes R, Digby NJ, Offord K (1985) Epidemiological study of wastage among racehorses 1982 and 1983. Vet Rec 116, 66-69

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70, 425-434

Schougaard H, Ronne JF, Phillipson J (1990) A radiographic survey of tibiotarsal osteochondrosis in a selected population of trotting horses in Denmark and its possible genetic significance. Equine Vet J 22, 288-289

Sue N, Jack BHA, Eaton SA, Pearson RCM, Funnell APW, Turner J, Czolij R, Denyer G, Bao S, Molero-Navajas JC, Perkins A, Fujiwara Y, Orkin SH, Bell-Anderson K, Crossley M (2008) Targeted Disruption of the Basic Kruppel-Like Factor Gene (Klf3) Reveals a Role in Adipogenesis. Mol Cell Biol 28, 3967-3978

Teyssèdre S, Dupuis MC, Elsen JM, Guérin G, Schibler L, Denoix JM, Ricard A (2010) Genome-wide SNP association study identifies region of interest associated

with osteochondrosis in French Trotters. In: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany.

Teyssèdre S, Dupuis MC, Guérin G, Schibler L, Denoix JM, Elsen JM, Ricard A (2011) Genome-wide association studies for osteochondrosis in French Trotters. J Anim Sci, DOI: 10.2527/jas.2011-4031

van Grevenhof EM, Ducro B, Weeren P, Tartwijk J, Belt A, Bijma P (2009a) Prevalence of various radiographic manifestations of osteochondrosis and their correlations between and within joints in Dutch Warmblood horses. Equine Vet J 41, 11-16

van Grevenhof EM, Schurink A, Ducro BJ, van Weeren PR, van Tartwijk JMFM, Bijma P, van Arendonk JAM (2009b) Genetic variables of various manifestations of osteochondrosis and their correlations between and within joints in Dutch warmblood horses. J Anim Sci 87, 1906-1912

van Weeren PR, Sloet vO-O, Barneveld A (1999) The influence of birth weight, rate of weight gain and final achieved height and sex on the development of osteochondrotic lesions in a population of genetically predisposed Warmblood foals. Equine Vet J 31, 26-30

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blocker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MCT, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, la Valle G, Fryc S, Guerin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Roed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvanen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC, Lander ES, Lindblad-Toh K,

Broad Inst Genome Sequencing P, Broad Inst Whole Genome A (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326, 865-867

Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6, 109-118

Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Diabetes Genetics I, Wellcome Trust Case Control C, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Cambridge GEMC, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM (2008) Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40, 575-583

Wittwer C, Hamann H, Rosenberger E, Distl O (2006) Prevalence of osteochondrosis in the limb joints of south German coldblood horses. Journal of Veterinary Medicine Series A-Physiology Pathology Clinical Medicine 53, 531-539

Wittwer C, Hamann H, Rosenberger E, Distl O (2007a) Genetic parameters for the prevalence of osteochondrosis in the limb joints of South German Coldblood horses. J Anim Breed Genet 124, 302-307

Wittwer C, Lohring K, Drogemuller C, Hamann H, Rosenberger E, Distl O (2007b) Mapping quantitative trait loci for osteochondrosis in fetlock and hock joints and palmar/plantar osseus fragments in fetlock joints of South German Coldblood horses. Anim Genet 38, 350-357

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 88, 76-82

Ytrehus B, Carlson CS, Ekman S (2007) Etiology and pathogenesis of osteochondrosis. Vet Pathol Online 44, 429-448

Zhang J, Yang C, Brey C, Rodriguez M, Oksov Y, Gaugler R, Dickstein E, Huang CH, Hashmi S (2009) Mutation in Caenorhabditis elegans Krüppel-like factor, KLF-3 results in fat accumulation and alters fatty acid composition. Exp Cell Res 315, 2568-2580

Tables

Table 1 A description of conditions (other than OC) for which horses were treated. For further information see Supplementary Material, Document S1.

	No. of affected		
Condition	**Cases**	**Controls**	**Total**
Angular limb deformity (ALD)	38	90	128
Fetlock chip(s)	36	71	107
Other chip(s)	3	3	6
Sesamoid fracture(s)	8	23	31
Other – bone related	4	1	5
Other – not bone related	7	3	10

Table 2 Genotype frequencies of BIEC2-799865 and results of chi-square tests for association with OCD

	Genotype frequency				
	C/C	C/T	T/T	Total No. of Samples	*p-value from X^2 test*[a]
Controls	0.26	0.55	0.19	168	
Cases[b]	0.44	0.46	0.10	162	0.002
Hock cases	0.42	0.46	0.12	89	0.034
Stifle cases	0.48	0.44	0.08	50	0.008
Fetlock cases	0.44	0.46	0.10	41	0.062

[a]The chi-square tests compare each case category with the controls.

[b]Note, the number of cases is not equal to the sum of the cases in each joint location because some horses were affected in multiple joint locations.

Figure Legends

Figure 1 Distribution of cases and controls across farms.

Figure 2 Distribution of genomic relationships between pairs of individuals.

Figure 3 Distribution of residuals from mixed model analysis.

Figure 4 A Manhattan plot showing association results for ECA3. The solid horizontal line represents the genome-wide significance level and the dashed line represents the significance level used to identify surrounding SNP with possible relevance.

Figure 5 LD plot (Barrett et al. 2005) of ECA3 region 1Mb either side of BIEC2-799865 (solid line, black circle). SNPs within the UGDH gene are indicated by a white circle. SNPs with a $p<0.001$ in the GWAS are indicated by a dashed line. Marked haplotypes calculated according to (Gabriel et al. 2002).

Supplementary Material

Table S1 – Radiographic surveys: 32 radiograph views as recommended by Keeneland Thoroughbred Racing and Sales, Lexington and based on guidelines provided by the American Association of Equine Practitioners (AAEP). Description taken from (Preston et al. 2010).

Table S2 – Quality control criteria implemented on genotype data and the number of SNP discarded at each step

Table S3 - The distribution of samples across contemporary groups fitted in animal model

Table S4 - Details of QTL regions tested.

Document S1 – A description of conditions (other than OC) commonly suffered by horses sampled