Whole genome scan identifies several chromosomal regions linked to equine sarcoids

V. Jandova1, J. Klukowska-Rötzler2, G. Dolf3, J. Janda4, P. Roosje5, E. Marti6, C. Koch7, V. Gerber8, J. Swinburne9

1Equine Clinic, Institute of Genetics, Department of Clinical Research and Division of Clinical Dermatology, DermFocus, University of Berne, Centre for Preventive Medicine, Animal Health Trust, Newmarket, Suffolk, UK

Abstract

Despite the evidence for a genetic predisposition to develop equine sarcoids (ES), no whole genome scan for ES has been performed to date. The objective of this explorative study was to identify chromosome regions associated with ES. The studied population was comprised of two half-sibling sire families, involving a total of 222 horses. Twenty-six of these horses were affected with ES. All horses had been previously genotyped with 315 microsatellite markers. Quantitative trait locus (QTL) signals were suggested where the F statistic exceeded chromosome-wide significance at \( P < 0.05 \). The QTL analyses revealed significant signals reaching \( P < 0.05 \) on equine chromosome (ECA) 20, 23 and 25, suggesting a polygenic character for this trait. The candidate regions identified on ECA 20, 23 and 25 include genes regulating virus replication and host immune response. Further investigation of the chromosome regions associated with ES and of genes potentially responsible for the development of ES could form the basis for early identification of susceptible animals, breeding selection or the development of new therapeutic targets.

Keywords: horse, equine sarcoïd, genome scan, quantitative trait locus analysis, candidate gene

Eine genomweite Kopplungsstudie zeigte mehrere Chromosomenregionen auf, welche mit einer Empfänglichkeit für equine Sarkoide assoziiert sind


Schlüsselwörter: Pferd, equines Sarkoïd, genomweite Kopplungsstudie, quantitative Merkmalsgenorte
**Introduction**

Equine sarcoïds (ES) are the most common skin tumors in equids (Marti et al., 1993). ES are considered semi-malignant, as they can form locally aggressive fibroblastic tumors, which can occur as single or multiple lesions of variable size (Marti et al., 1993; Chambers et al., 2003). ES present a serious health problem, since they may compromise the use, value and the welfare of the animal, and may progress towards more aggressive tumors or recrudescence after treatment (Marti et al., 1993; Brandt et al., 2008).

The etiology of ES is not yet clear. Environmental and genetic factors have been implicated. Several studies suggest that bovine papilloma viruses (BPV) types 1 and 2 are important extrinsic factors in the development of ES (Marti et al., 1993; Bogaert et al., 2007; Brandt et al., 2008). However, even though ES-like tumors can be induced in horses using BPV, they usually spontaneously regress. This suggests that individual host factors must contribute to persistence of the tumor (Fadok, 1995). Previous studies have suggested a genetic predisposition for ES as well. Some authors observed an increased prevalence of ES cases in certain families (Ragland et al., 1966; Stannard and Pulley, 1978), whereas others did not (Studer et al., 2007). Breed predilection, with a high prevalence in Quarter horses and a low prevalence in Standardbreds (Angelos et al., 1988; Mohammed et al., 1992), and an association between ES and major histocompatibility complex (MHC) genes have been described (Lazary et al., 1985; Meredith et al., 1986; Brostrom et al., 1988; Lazary et al., 1994). While these results suggest a genetic component in the etiology of ES, to date, no genetic analysis of the whole equine genome has been performed to identify chromosomal regions and candidate genes associated with susceptibility to ES.

In the present study we used the whole genome scan approach (Swinburne et al., 2009) to identify chromosome regions in which genetic variants are located that contribute to inherent ES susceptibility. The scan was performed on horses belonging to two families of Swiss Warmbloods (SW) using a panel of microsatellite markers that spans the 31 horse autosomes with an average spacing of 10 Mb.

**Animals, Material and Methods**

**Sire families and phenotypes**

Genetic analysis was performed using the offspring of two stallions affected with recurrent airway obstruction (RAO). These two families have been described in detail elsewhere (Ramseyer et al., 2007). Briefly, 222 horses were included in the study, 26 of which suffered from ES. Owners of ES-affected horses completed a detailed questionnaire focusing on specific information about the history of the development of ES, age of onset, number of tumors, size and localization. Diagnosis of ES by the treating veterinarian was a requirement for inclusion. Horses with any questionable skin lesions were further examined by one of the authors (VJ).

The approximate area of the skin affected with ES was calculated based on measurements by VJ or, if the ES had been removed, descriptions by the owners. If horses had several ES, the areas of each ES were added to give an estimate of the total affected area. According to the total affected area, horses were divided into four groups (group 1 ranging from 0.1–2 cm²; group 2: 2.1–6.5 cm²; group 3: 7–20 cm²; group 4: 24–300 cm²).

Further, groups according to the number of ES were formed: group 1 (1 tumor), group 2 (2 tumors), group 3 (3–4 tumors), group 4 (5 and more tumors). A total severity score (ranging 2–8) was created by adding the area scoring (1–4) and number scoring (1–4). Finally, the subjects were divided into two groups: severity group 1 included horses with total score 2–4 (mildly affected) and severity group 2 included horses with total score 5–8 (severely affected).

**Statistical analysis**

Descriptive statistics of the horse population were performed using the NCSS program (http://www.ncss.com). Cross tabulation and \( \chi^2 \) tests were used to evaluate the effect of gender and coat color on the ES prevalence and extent of the disease (number of tumors, total affected area, number of ES localizations, severity group). Further, using cross tabulation and \( \chi^2 \) tests, the possible effects of contact to other ES-affected horses and contact to cattle at the time of ES-appearance on the extent of the disease were analyzed. The results were considered statistically significant at \( P < 0.05 \).

**Genotyping**

All 222 individuals had been previously genotyped with 315 microsatellite markers, as described in detail elsewhere (Swinburne et al., 2009). The location of each marker was identified by comparing the sequence of the unique region flanking the microsatellite with the second assembly of the horse genome sequence using BLAST (http://www.ensembl.org/Equus_caballus/Info/Index). For all markers a unique match to the genome sequence was observed. On ECA 20, additional 9 microsatellite markers were used, in order to pinpoint the significant region more accurately and thus, to investigate the relationship with MHC II more closely. Their positions (27,507,281; 33,246,700; 38,972,841; 40,097,621; 42,184,759; 44,762,621; 46,236,874; 49,675,570; 51,026,745) are marked with asterisk on Figure 1. These markers were identified from the horse genome sequence (EquCab2.0) and those that were polymorphic in one or both of the sires were genotyped.
Chromosomal regions linked to equine sarcoids

We observed that significantly fewer bay animals were affected than horses with other coat colors: nine (7%) out of 123 bay horses were affected vs. 12 (18%) out of 68 chestnut horses ($P = 0.01$). There was no difference in either prevalence of ES or severity score between male and female animals. Close contact to another affected animal at the time of ES appearance was associated with increased number of localizations ($p = 0.01$), but not with other parameters (diameter, size, number of sarcoids and total score). Horses with contact to cattle had tumors affecting a significantly greater total area of skin and had a higher severity score (both $p = 0.03$), and there was a trend towards a larger tumor diameter ($p = 0.07$). Contact to cattle had no effect on the number of sarcomas and number of localizations.

QTL analysis and identification of candidate genes

The signals that reached chromosome-wide significance at $P < 0.05$ were observed on ECA 20, 23 and 25, with the peak signal at 44.05, 34.25, and 10.98 Mb resp. For detailed information, see Table 1 and Figures 1a and 1b. The significant regions of ECA 20, 23 and 25 and flanking sequences were investigated for candidate genes. The candidate genes investigated on ECA 20, 23 and 25 include genes involved in anti-viral or anti-tumor immunity (for further details see Tab. 2).

Discussion

To our knowledge, this is the first study using a genome scan to search for candidate genes involved in ES development. Our identification of candidate gene regions supports the hypothesis that there is a genetic component in the etiology of ES. The significant signals were observed on ECA 20, 23 and 25, which suggests a polygenic inheritance. The candidate genes investigated on ECA 20, 23 and 25 include genes involved in anti-viral or anti-tumor immunity (for further details see Tab. 2).

Previous studies reported an association between ES susceptibility and MHC-class II alleles (Lazary et al., 1985; Meredith et al., 1986; Brostrom et al., 1988; Lazary et al., 1994). The genes coding for the equine MHC are located on ECA 20 at 34 Mb. In contrast, the peak signal on ECA 20 observed in the current study was at 44.05 Mb. However, the density of informative markers was relatively low.

**Table 1:** Significant QTL analysis results from families 1 and 2 analyzed together using graded phenotype definition.

<table>
<thead>
<tr>
<th>ECA</th>
<th>Position of peak signal (Mb)</th>
<th>Significant region (Mb)</th>
<th>F statistic</th>
<th>5% sig level – chromosome wide – 10 000 iterations</th>
<th>1% sig level – chromosome wide – 10 000 iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>44.05</td>
<td>44.05</td>
<td>4.94</td>
<td>4.94</td>
<td>6.94</td>
</tr>
<tr>
<td>23</td>
<td>34.25</td>
<td>24.25 – 47.25</td>
<td>4.62</td>
<td>4.06</td>
<td>5.88</td>
</tr>
<tr>
<td>25</td>
<td>10.98</td>
<td>8.98 – 12.98</td>
<td>4.16</td>
<td>3.82</td>
<td>5.93</td>
</tr>
</tbody>
</table>
in that region (Fig. 1) and a denser array of markers is now required to demonstrate whether the association is focused at the MHC. In humans, a relationship between (pre)malignant cervical carcinoma in women and human papilloma virus (HPV) infection has been identified (Walboomers et al., 1999) and different host genetic factors, mainly determining the ability of the immune system to clear an HPV infection, play an important role in the development of the tumor (Ivansson, 2009; Vissier et al., 2007). The heritability of cervical cancer is estimated to be 27%. The human leukocyte antigen (HLA) class II DQB1 and DRB1, p53 and the recently described thymic stromal co-transporter (TSCOT) are major associated genes (Engelmark et al., 2006). Equine TSCOT gene is located on the ECA 25 at 17.5 Mb. This study identified a significant signal on ECA 25 with a peak at 10.98 Mb. Further studies, using larger numbers of affected animals and a case-control design will have to confirm a possible association between TSCOT and ES. The equine homolog of p53 is located on ECA 11, but no significant signal associated with ES was found on this chromosome. The mode of inheritance of ES is unknown, though it is likely to be polygenic. In the families described here a dominant risk locus would likely be inherited via the dams, because the sires themselves were not ES-affected. On the other hand, recessive risk loci may well be inherited from the unaffected sires. Furthermore, the difference in ES prevalence between the two families (9 vs. 18%) is considerable and might be explained by differing risk alleles in each sire. The prevalence of ES among the mares was not known but was probably average for the SW population (7.5%, Baleri, 2008). Differences in environmental factors are an unlikely explanation for the discrepancy in prevalences between the families, since no significant differences were observed between the two families in regards to contact to other ES-affected horses, to livestock or the length of paddock/pasture stay at the time of sarcoid appearance. This study has utilized genomic data obtained in a previous study on RAO (Swinburne et al., 2009); consequently the prevalence of ES in the investigated population was lower than optimal for genetic analyses. Nonetheless, these preliminary data represent the first study using a whole-genome scan to investigate ES genetics and provide valuable information on which further experiments can be based. The replication of these findings should be tested in ideally independent case/control population with a larger number of affected animals. Additionally, detailed mapping using SNPs will be needed to more closely define the regions worthy of further scrutiny. Identification of mutations in candidate genes, together with the demonstration of varying expression in cases and controls, will define their role in the pathogenesis of ES more precisely. In all cases, the diagnosis of ES was made by a veterinarian and was based on a thorough clinical examination. Histopathological examination of biopsied tissue could have increased phenotype reliability. However, taking biopsies carries a significant risk of exacerbation (Ragland et al., 1978; Knottenbelt, 2009). Furthermore, a very close correlation (89–100%) between clinical diagnosis and histopathology has been reported in several studies (Lazary et al., 1983; Vanselow et al., 1988; Lazary et al., 1994) and limitations of histopathology as a gold standard have been pointed out (Martens et al., 2001). Therefore, clinical examination was appropriate for diagnosis in this study. In a future study, however, a prospective design with only one observer scoring all lesions and confirmation of questionable lesions by biopsy may improve the reliability of the phenotypic diagnosis. The mean age of ES onset (6 years) corresponds well with the time of sarcoid appearance. Histopathological examination of biopsied tissue could have increased phenotype reliability. However, taking biopsies carries a significant risk of exacerbation (Ragland et al., 1978; Knottenbelt, 2009). Furthermore, a very close correlation (89–100%) between clinical diagnosis and histopathology has been reported in several studies (Lazary et al., 1983; Vanselow et al., 1988; Lazary et al., 1994) and limitations of histopathology as a gold standard have been pointed out (Martens et al., 2001). Therefore, clinical examination was appropriate for diagnosis in this study. In a future study, however, a prospective design with only one observer scoring all lesions and confirmation of questionable lesions by biopsy may improve the reliability of the phenotypic diagnosis. The mean age of ES onset (6 years) corresponds well with observations of previous studies (3–6 years) (Mohamed et al., 1992; Marti et al., 1993). Ninety-five per cent of the horses (10 out of 220) were older than 10 years at the time of the first interview, therefore the risk of false negative phenotypes is small. The lower ES prevalence

Table 2: List of candidate genes located in the significant chromosomal regions of ECA 20, 23 and 25.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full annotation</th>
<th>Position (MB) ECA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECA 20 HSA 6</td>
<td>TAF13 TAF13 RNA polymerase II</td>
<td>43.6</td>
</tr>
<tr>
<td></td>
<td>RUNX2 Runt-related transcription factor 2</td>
<td>44.1</td>
</tr>
<tr>
<td></td>
<td>CLIC5 Chloride intracellular channel 5</td>
<td>44.5</td>
</tr>
<tr>
<td>ECA 23 HSA 9</td>
<td>JAK2 Janus Kinase 2</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>PDL1 Programmed death ligand 1</td>
<td>26.7</td>
</tr>
<tr>
<td></td>
<td>PDL2 Programmed death ligand 2</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>IFNB1 Interferon beta 1</td>
<td>40.1</td>
</tr>
<tr>
<td></td>
<td>IFNA1 Interferon alpha 1</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td>IFNB1 Interferon beta 1</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td>IFNK Interferon kappa</td>
<td>42.7</td>
</tr>
<tr>
<td></td>
<td>CCL21 CC-chemokine ligand 21</td>
<td>50.6</td>
</tr>
<tr>
<td>ECA 25 HSA 9</td>
<td>SMC2 Structural maintenance of chromosomes 2</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>FSD1L Fibronectin type III and SPRY domain containing 1-like</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>FKTN Fukutin</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>TAL2 T-cell acute lymphocytic leukemia 2</td>
<td>11.7</td>
</tr>
</tbody>
</table>

HSA Homo sapiens autosome.
Chromosomal regions linked to eqine sarcoids in brown horses compared to chestnut observed in the present study is in accordance with previous observations (Vanselow et al., 1988; Studer et al., 2007). Since there was no difference in coat color distribution between the families, this would not have influenced our results.

**Conclusion**

This study provides a first step in the elucidation of the genetic basis of ES. Candidate regions on ECA 20, 23 and 25 were identified, which may harbor genes responsible for the genetic basis of ES. Replication in additional pop-
ulations is required before further investigation of candidate genes identified in this explorative study is warranted, however. Ultimately, this approach could aid the prevention and treatment of this disease, which is a considerable economic, health, and welfare concern.

**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**

This study was supported by Vetsuisse and DKV grants, the Berne Equine Lung Research Group, the Swiss National Science Foundation grant number 310000116502. The Horse Trust, a UK horse charity committed to promoting education and welfare within the equine world, funded this project in the UK. We thank A. Ramseyer, U. Jost, N. Hasler, E. Laumen, P. Nussbaumer, and D. Burger and all horse owners for their help phenotyping the horses and collecting blood samples.

**Une étude d’association tenant compte de l’ensemble du génome démontre que plusieurs régions chromosomiques sont associées avec les sarcoïdes équins**

Bien que la prédisposition génétique pour la réceptivité des sarcoïdes équins soit depuis longtemps supposée, il n’a jusqu’à maintenant pas été réalisé d’étude d’association tenant compte de l’ensemble du génome avec le caractère «sarcoïde équin» (SE). Le but de la présente étude était d’identifier des régions de chromosomes associées avec le phénotype SE. La population utilisée dans le but se composait exclusivement de descendants directs de 2 étalons demi-sang suisse. Au total 222 individus issus des 2 populations ont été examinés, 26 d’entre eux étant affectés de SE. Les 222 chevaux ont tout d’abord été typisés au moyen de 315 marqueurs de microsatellites. On a défini comme loci de caractère quantitatif (QTL) pour le sarcoïde équin des régions des chromosomes pour lesquels la statistique F dépassait une signification chromosomique de P< 0.05. Au moyen de l’analyse QTL, on a identifié des signaux sur les chromosomes équins 20, 23 et 25, ce qui signale un mode d’héritabilité polygénique. Les régions candidates comprennent des gènes qui ont entre autre une influence directe sur la réplication virale et la réponse immunitaire de l’hôte. Des recherches plus poussées quant aux régions chromosomiques associées aux sarcoïdes équins et quant aux gènes qui y sont localisés, offrent pour l’avenir des bases en vue d’un diagnostic précoce des animaux sensibles, d’une meilleure hygiène de l’élevage ou du développement de nouvelles possibilité thérapeutique des sarcoïdes.

**Studio di genom-wide linkage su diverse regioni cromosomiche associate a sarcoïdi equini**

Anche se, da tempo, si ammette una certa predisposizione genetica al sarcoïde equino, finora non sono stati eseguiti studi di genome-wide linkage segnati con «Equine sarcoïdosi» (ES). Lo scopo di questo studio era di identificare le regioni cromosomiche associate al fenotipo ES. La popolazione equina che è stata utilizzata per questo studio è costituita esclusivamente da discendenti diretti di due stalloni a sangue caldo svizzero. Per lo studio sono stati impiegati un totale di 222 individui provenienti da entrambe le popolazioni di fratellastri fra cui ben 26 cavalli erano colpiti da ES. Tutti i 222 cavalli sono stati digitati in precedenza con 315 marcatori microsatelliti. Come tratto quantitativo (QTL - quantitative trait locus) per l’ES sono state definite le regioni cromosomiche per cui la statistica F supera un livello di significatività P< 0.05 della lunghezza di un cromosoma. L’analisi QTL ha identificato i segnali sui cromosomi equini (ECA) 20, 23 e 25, suggerendo una modalità d’ereditarietà poligenica. Le regioni candidate contengono geni che hanno, tra l’altro, un impatto diretto sulla replicazione del virus e sulla risposta immunitaria dell’host. Ulteriori indagini di regioni cromosomiche associate ES e dei geni localizzati in quella regione offrono in futuro una base per l’individuazione precoce dei soggetti suscettibili, una migliore igiene nell’allevamento o lo sviluppo di nuove terapie per l’ES.
References


Corresponding author

Vendula Jandova

Vetsuisse-Fakultät Universität Bern

Pferdeklinik

Länggass-Strasse 124

3001 Bern

Tel.: +41 (0)31 631 22 43

Fax: +41 (0)31 631 26 20

vendula.jandova@knp.unibe.ch

Received: 14 March 2011

Accepted: 16 June 2011