Genome-Wide Linkage Scan for Athlete Status in 700 British Female DZ Twin Pairs

Marleen H. M. De Moor,[1] Tim D. Spector,[2] Lynn F. Cherkas,[2] Mario Falchi,[2] Jouke Jan Hottenga,[1] Dorret I. Boomsma,[1] and Eco J. C. De Geus[1]

[1] *Department of Biological Psychology, Vrije Universiteit Amsterdam, the Netherlands*

[2] *Twin Research and Genetic Epidemiology Unit, St Thomas' Hospital, London, United Kingdom*

Association studies, comparing elite athletes with sedentary controls, have reported a number of genes that may be related to athlete status. The present study reports the first genome wide linkage scan for athlete status. Subjects were 4488 adult female twins from the TwinsUK Adult Twin Registry (793 monozygotic [MZ] and 1000 dizygotic [DZ] complete twin pairs, and single twins). Athlete status was measured by asking the twins whether they had ever competed in sports and what was the highest level obtained. Twins who had competed at the county or national level were considered elite athletes. Using structural equation modeling in Mx, the heritability of athlete status was estimated at 66%. Seven hundred DZ twin pairs that were successfully genotyped for 1946 markers (736 microsatellites and 1210 SNPs) were included in the linkage analysis. Identical-by-descent probabilities were estimated in Merlin for a 1 cM grid, taking into account the linkage disequilibrium of correlated SNPs. The linkage scan was carried out in Mx using the $\hat{\pi}$-approach. Suggestive linkages were found on chromosomes 3q22-q24 and 4q31-q34. Both areas converge with findings from previous studies using exercise phenotypes. The peak on 3q22-q24 was found at the SLC9A9 gene. The region 4q31-q34 overlaps with the region for which suggestive linkages were found in two previous linkage studies for physical fitness (FABP2 gene; Bouchard et al., 2000) and physical activity (UCP1 gene; Simonen et al., 2003). Future association studies should further clarify the possible role of these genes in athlete status.

Elite athletic performance is thought to be the result of the combination of a high genetic potential and the optimal environmental conditions, such as training and nutrition (MacArthur & North, 2005). It has long been recognized that physical performance is genetically determined (Bouchard & Malina, 1998), as evidenced by the significant heritabilities ranging from 31% to 85% for physical performance traits involving different aspects of cardiorespiratory fitness (Bouchard et al., 1998; Bouchard et al., 1999; Perusse et al., 2001) and skeletomuscular strength and performance (De Mars et al., 2007; Thomis et al., 1998).

The number of studies aiming to identify the actual genetic variants that account for the heritability of physical performance phenotypes is increasing and are summarized in the 'human gene map for performance and health-related exercise phenotypes', which is regularly updated (Perusse et al., 2003; Rankinen et al., 2001; Rankinen et al., 2002; Rankinen et al., 2004; Wolfarth et al., 2005). Besides linkage and association studies on different physiological parameters related to exercise, the human gene map also includes a number of association studies of candidate genes, in which elite athletes are compared with sedentary controls. Genes that have been related to elite athlete status are, for example, the angiotensin-converting enzyme (ACE) gene (Woods et al., 2000), and the bradykinin receptor (B2R) gene (Williams et al., 2004). These genes are thought to have an impact on physical performance through increasing cardiorespiratory fitness as a result of training, but possibly also through enhanced muscle efficiency (Williams et al., 2000). Other genes that have been related to elite athlete status are the α-actin-3 (ACTN3) gene (MacArthur & North, 2007; Paparini et al., 2007; Yang et al., 2003) and two adrenergic receptor genes (ADRA2A and ADRB2; Moore et al., 2001; Wolfarth et al., 2000).

To date, no linkage studies have been published that identified the genomic locations that may be related to athlete status. In this study, the results of a heritability analysis and genome-wide linkage scan are presented for data on athlete status in a sample of British female twin pairs. Athlete status was measured by asking the twins whether they ever participated in sports and what was the highest

Received 21 September, 2007; accepted 3 October, 2007.

Address for correspondence: Marleen H. M. de Moor, Vrije Universiteit, Department of Biological Psychology, van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. E-mail: mhm.de.moor@psy.vu.nl

level they ever competed at. Twins who competed at the national or county level were considered elite athletes.

Methods

Sample

Phenotype information came from a mailed survey sent out in 2000 to female twins registered at the TwinsUK Adult Twin Registry (Spector & Williams, 2006). There were 4527 twins with valid data on the level of participation in competitive sports. Twins with unknown zygosity were excluded (39 individuals), resulting in a total sample of 4488 individuals, from which 1793 complete twin pairs could be formed (i.e., both individuals had data on sports level). Table 1 gives an overview of the number of complete and incomplete twin pairs as a function of zygosity. Zygosity of the twin pairs was determined using a standardized questionnaire and DNA fingerprinting (Spector et al., 1996). The mean age of the individuals was 51.9 (standard deviation: 12.8), the minimum age was 20-years-old and the maximum 83-years-old.

Phenotyping

Athlete status was measured by asking the participants to indicate for a list of sports whether they had ever participated in each of these sports and what was the highest level at which they had ever competed in these sports. The levels were social only, in a school, club or university team, at the county level or at the national level. Table 2 lists the sports that were included in the survey. Yoga and walking were not included in the analyses, since these were low-impact sports (at less than 4 metabolic equivalents). Participants were classified into three categories: (1) never participated in sports in any organized form (no exercise and exercise at the social level), (2) ever participated in sports at the school, club or university level, and (3) ever participated in sports at the county or national level. Women who ever participated in sports at the county or national level are thought to have been elite athletes (MacArthur & North, 2005).

Genotyping

DNA of 2600 individuals was extracted from venous blood samples following standard protocols. Two different marker sets were used: a microsatellite marker set using standard ABI Prism genotyping methodologies (Applied Biosystems, Foster City, CA, USA) and a single nucleotide polymorphism (SNP) marker set (HuSNP GeneChips, Affymetrix), which are more fully described elsewhere (Wilson et al., 2003). Basic checks for sex and zygosity were carried out. Errors of Mendelian inheritance were detected using Pedstats (Abecasis et al., 2002). Unlikely recombinants were detected using Merlin (Abecasis et al., 2002). Unlikely or erroneous genotypes due to Mendelian or genotyping errors were set as unknown. SNPs with minor allele frequency lower than .05 were excluded. Genetic map positions of the markers were taken from the interpolated genetic map developed by Duffy (2006). For markers that were not present on this map, their physical position was taken from the National Center for Biotechnology Information (NCBI) build 35.1. Locally weighted linear regression of the genetic map positions on the physical positions was used to impute the missing genetic positions. Markers that could not be reliably located on the physical map (which was the case for some SNPs) were excluded from the dataset. This resulted in 1946 markers, of which 736 were microsatellites and 1210 SNPs. All genotyped dizygotic (DZ) twin pairs with complete data on athlete status were selected (705 pairs). Twin pairs for whom one of the individuals had less than 200 successfully genotyped markers were excluded (5 pairs). These were pairs with only microsatellites genotyped. There were 99 pairs with less than 200 successfully genotyped microsatellites, but with a large number of SNPs. These pairs were retained in the analyses. Thus, the final sample used for the linkage analysis consisted of 700 pairs. The mean age of these individuals was 54.6 (standard deviation: 11.5), the minimum age was 21 and the maximum 83 years old. Average spacing of all 1946 markers was 6.22 centiMorgan (cM) and the average heterozygosity 61.0%. For the statistical analyses the Haldane mapping function was used. All reported values are in Haldane cM.

Markers that were close together on the genetic map, as was the case for some of the SNPs in the dataset, may be in linkage disequilibrium. If linkage disequilibrium among markers is ignored in the linkage analysis, this might lead to an upward bias in linkage signals (Abecasis & Wigginton, 2005). Merlin accommodates a feature that identifies clusters of markers that are in linkage disequilibrium, and estimates the haplotype frequencies for the clustered markers. Using an r-squared of .30 as a cut-off value, clusters of markers and their haplotype frequencies were estimated. This resulted in 73 clusters, of which 62 contained two markers, 10 contained 3 markers, and 1 contained 4 markers. Three clusters contained both SNPs and microsatellites. For markers that were not clustered, the allele frequencies were determined by counting the alleles that occurred in the dataset. The obtained allele and

Table 1

Polychoric Twin Correlations and Number of Complete and Incomplete Female MZ and DZ Twin Pairs

	Complete twin pairs	Incomplete twin pairs	Total pairs (individuals)	Polychoric twin correlation (95% CI)
MZ	793	348	1141 (1934)	.66 (0.59–0.71)
DZ	1000	554	1554 (2554)	.32 (0.24–0.40)
Total	1793	902	2695 (4488)	

Note: 95% CI = 95% confidence interval; MZ = monozygotic; DZ = dizygotic

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

Table 2

Number of Women Who Never or Ever Participated in Sports at the School/Club/University or County/National Level for a List of Sports

	No sports	School/ club/ university level	County/ national level
	N (%)	*N* (%)	*N* (%)
1. Swimming	3849 (85.8)	567 (12.6)	72 (1.6)
2. Cycling	4463 (99.4)	23 (0.5)	2 (0.0)
3. Running	3656 (81.5)	742 (16.5)	90 (2.0)
4. Keep fit / Aerobics	4382 (97.6)	101 (2.3)	5 (0.1)
5. Gymnastics	4092 (91.2)	377 (8.4)	19 (0.4)
6. Tennis	3983 (88.7)	487 (10.9)	18 (0.4)
7. Badminton	4195 (93.5)	280 (6.2)	13 (0.3)
8. Squash	4413 (98.3)	68 (1.5)	7 (0.2)
9. Golf	4437 (98.9)	46 (1.0)	5 (0.1)
10. Skiing	4468 (99.6)	18 (0.4)	2 (0.0)
11. Ice skating	4480 (99.8)	8 (0.2)	0 (0.0)
12. Yoga	4472 (99.6)	13 (0.3)	3 (0.1)
13. Dancing	4351 (96.9)	122 (2.7)	15 (0.3)
14. Rugby / Football	4423 (98.6)	61 (1.4)	4 (0.1)
15. Cricket	4433 (98.8)	52 (1.2)	3 (0.1)
16. Martial arts	4457 (99.3)	26 (0.6)	5 (0.1)
17. Boxing	4485 (99.9)	2 (0.0)	1 (0.0)
18. Hill walking	4463 (99.4)	24 (0.5)	1 (0.0)
19. Walking	4444 (99.0)	40 (0.9)	4 (0.1)
20. Other	3765 (83.9)	597 (13.3)	126 (2.8)
Any sport	*2498 (55.7)*	*697 (37.4)*	*311 (6.9)*

Note: *N* = number of women; % = percentage of women
The numbers given for the individual sports do not add up to the numbers given for any sport, because some women had participated in more than one sport.

haplotype frequencies were used to estimate the identical-by-descent (IBD) probabilities, thereby assuming that within clusters the recombination fraction is zero and across clusters there is linkage equilibrium (Abecasis & Wigginton, 2005).

IBD Estimation

A sibling pair shares an allele at a specific genomic locus identical by descent (IBD) if the allele is inherited from the same parent. Thus, a sibling pair can share 0, 1, or 2 alleles IBD at a specific locus. If parental genotypes are not available, the probabilities of IBD status 0, 1, or 2 were estimated based on the allele frequencies as observed in the population. The IBD probabilities for each sibling pair were estimated using the Lander-Green algorithm in Merlin for a 1 cM grid (Abecasis et al., 2002). The IBD probabilities were used to estimate the proportion of alleles shared IBD for each sibling pair, using the formula (Sham, 1998):

$$\hat{\pi} = 0.5p(\text{IBD} = 1) + p(\text{IBD} = 2)$$

Where $p(\text{IBD} = 1)$ is the probability that IBD status is 1, and $p(\text{IBD} = 2)$ is the probability that IBD status is 2, and $\hat{\pi}$ can take values ranging between 0 and 1. Identical by descent estimates for all genomic positions for all sibling pairs were saved to files and later used for the linkage analysis in Mx.

Statistical analyses

Standard structural equation modeling methods were used in Mx (Neale et al., 2003; Neale & Cardon, 1992). Threshold models were fitted to the raw ordinal data on athlete status, including data from single twins, using full information maximum likelihood estimation. The threshold model assumes that a categorical variable has an underlying liability with a continuous and standard normal distribution. For athlete status, two thresholds divided the liability distribution into the three observed categories. In a saturated model, which was fitted to the data of MZ and DZ twin pairs, two thresholds were estimated for each twin in each zygosity group (8 thresholds in total). A cohort effect was allowed on each threshold, by modeling age as a definition variable on the thresholds. It was assumed that the effect was the same on all first thresholds and on all second thresholds (2 regression coefficients were estimated). Polychoric MZ and DZ twin correlations were estimated. Thus, the saturated model contained 12 free parameters. We tested for the significance of the cohort effects on the first and second thresholds, and we tested for birth order and zygosity effects in the thresholds. Based on the pattern of MZ-DZ correlations, an ADE model was fitted to the data and the proportions of variance in liability for athlete status due to A, D and E were estimated. We tested for the significance of D, and of A and D. The significance of these parameters was evaluated by the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (–2LL) between the full ADE model and the submodel in which a parameter was dropped is χ^2 distributed. The degrees of freedom of the c^2-test (Δdf) equal the difference in degrees of freedom between the two models. If the χ^2-test yielded a *p*-value higher than .05, the fit of the submodel was not significantly worse than the fit of the full model, and the submodel was kept as the most parsimonious and best fitting model.

In the sample of genotyped DZ twin pairs, the DZ twin correlation and the thresholds were computed to examine whether the genotyped sample deviated from the total sample with respect to the DZ twin correlation and the prevalence of sports participation at the various levels. Next, an AQE model was fitted to the data to test, at each locus, whether the proportion of alleles shared IBD in each DZ twin pair explained their resemblance in athlete status (Neale et al., 2003). Using the previously given formula, $\hat{\pi}$ was computed in Mx from the estimated IBD probabilities. At each locus, the

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

significance of the effect of Q was evaluated by use of the likelihood ratio test. From this test, the LOD score can be computed by dividing the obtained chi-square test statistic by 2ln10 (~ 4.6).

Empirical thresholds for suggestive and significant linkage were computed by randomly permuting the datasets 1000 times. Permutations were carried out by randomly assigning the IBD estimates to the sibling pairs, keeping the sibling pairs and IBD structure of the whole genome intact. A linkage scan was then performed on each permuted dataset. The empirical threshold for suggestive linkage was computed by obtaining the maximum LOD score for each chromosome out of the 1000 analyses, and determining what LOD score occurred a 1000 times out of 22,000. The threshold for significant linkage was computed by recording the maximum LOD score in each linkage scan on each permuted dataset, and then determining which LOD score occurred 50 out of 1000 times (see also Boomsma et al., 2006). The empirical threshold for suggestive linkage was 1.63 and for significant linkage 2.96.

Results

Prevalence of Sports Participation at the Various Levels

From 4488 women, there were 311 women who ever participated in sports at the county or national level (6.9%). Sixteen hundred and seventy-nine women ever participated in sports at the school, club or university level (37.4%). The remaining 2498 women never participated in any organized form of sports (55.7%). There were no differences in these prevalences between first and second-born twins (χ^2_4 =3.53, p = .47) nor between MZ and DZ twins (χ^2_4 =2.50, p = .64). Women from older birth cohorts less frequently participated in sports at the school, club, or university level, and the county or national level, as indicated by the significant and positive regression coefficients of age on both the first (χ^2_4 =100.64, p < .001) and the second threshold (χ^2_4 =33.63, p < .001).

Heritability of Athlete Status

The polychoric MZ twin correlation was .66 (95% Confidence Interval (CI): 0.59–0.71) and the polychoric DZ twin correlation .32 (95% CI: 0.24–0.40). We fitted an ADE model to the data. The thresholds were modeled according to the best fitting constrained saturated model, which contained two thresholds and a cohort effect on each threshold. The model fitting results of the heritability analyses of athlete status are given in Table 3. The AE-model described the data adequately. Of the total variance in liability for athlete status, 65.5% was explained by A (95% CI: 59.6%–70.6%) and 34.5% by E (95% CI: 29.3%–40.4%). Thus, a combination of additive genetic and nonshared environmental factors influences athlete status.

Genome-Wide Linkage Scan for Athlete Status

The prevalence of sports participation at the various levels in the genotyped DZ twins was highly similar to the percentages in the total sample: 6.8% ever participated in sports at the county or national level, 36.9% ever participated at the school, club, or university level and 56.3% did not ever participate in sports at the organized level. The twin correlation in the genotyped DZ twin pairs was .33. Thus, the genotyped sample was representative for the total sample in terms of both prevalence and the DZ twin correlation.

The results of the genome-wide linkage scan are depicted in Figure 1. Suggestive linkages were found on chromosomes 3q24 and 4q32.3. The peak on 3q24 has a maximum LOD score of 2.35 at 153 cM. The confidence interval of this peak, using the drop 1 LOD rule, is 3q22-q24 (145-160 cM). The markers at the peak are D3S1569 and the clustered markers rs2801 and rs2800, all residing within the sodium/hydrogen exchanger 9 (SLC9A9) gene. The second peak is on 4q32.3, has a maximum LOD score of 1.87, and is found at 168/169 cM near marker D4S1597. The confidence interval of the peak is 4q31-q34 (147-188 cM).

Table 3

Model Fitting Results for Heritability Analyses of Athlete Status

	Fit statistics						Parameter estimates						
Model	–2LL	*df*	Vs.	χ^2	Δ*df*	*p*	A	D	E	τ1	τ2	β1	β2
1. Saturated	7479.069	4476	–	–	–	–	–	–	–	–	–	–	–
2. ADE	7483.820	4482	1	4.751	6	.576	62.5	0.03	34.4	–0.720	0.797	0.017	0.014
3. AE	**7483.852**	**4483**	**1**	**4.783**	**7**	**.686**	**65.5**	**–**	**34.5**	**–0.719**	**0.797**	**0.017**	**0.014**
4. E	7767.048	4484	3	283.196	1	< .001	–	–	100	–0.673	0.799	0.016	0.014

Note: –2LL = –2 log-likelihood; *df* = degrees of freedom; vs. = model to which the fitted model is compared; χ^2 = chi-square value; Δ*df*=difference in degrees of freedom between fitted and comparison model; *p* = *p*-value; A = proportion of variance explained by additive genetic factors; D = proportion of variance explained by non-additive genetic factors; E = proportion of variance explained by nonshared environmental factors; τ1= first thresholds; τ2 = second threshold; β1 = regression coefficient of age on the first threshold; β2 = regression coefficient of age on the second threshold; Positive age regression coefficients indicate that athlete status was less frequent in older women.

The most parsimonious and best fitting model is shown in bold.

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

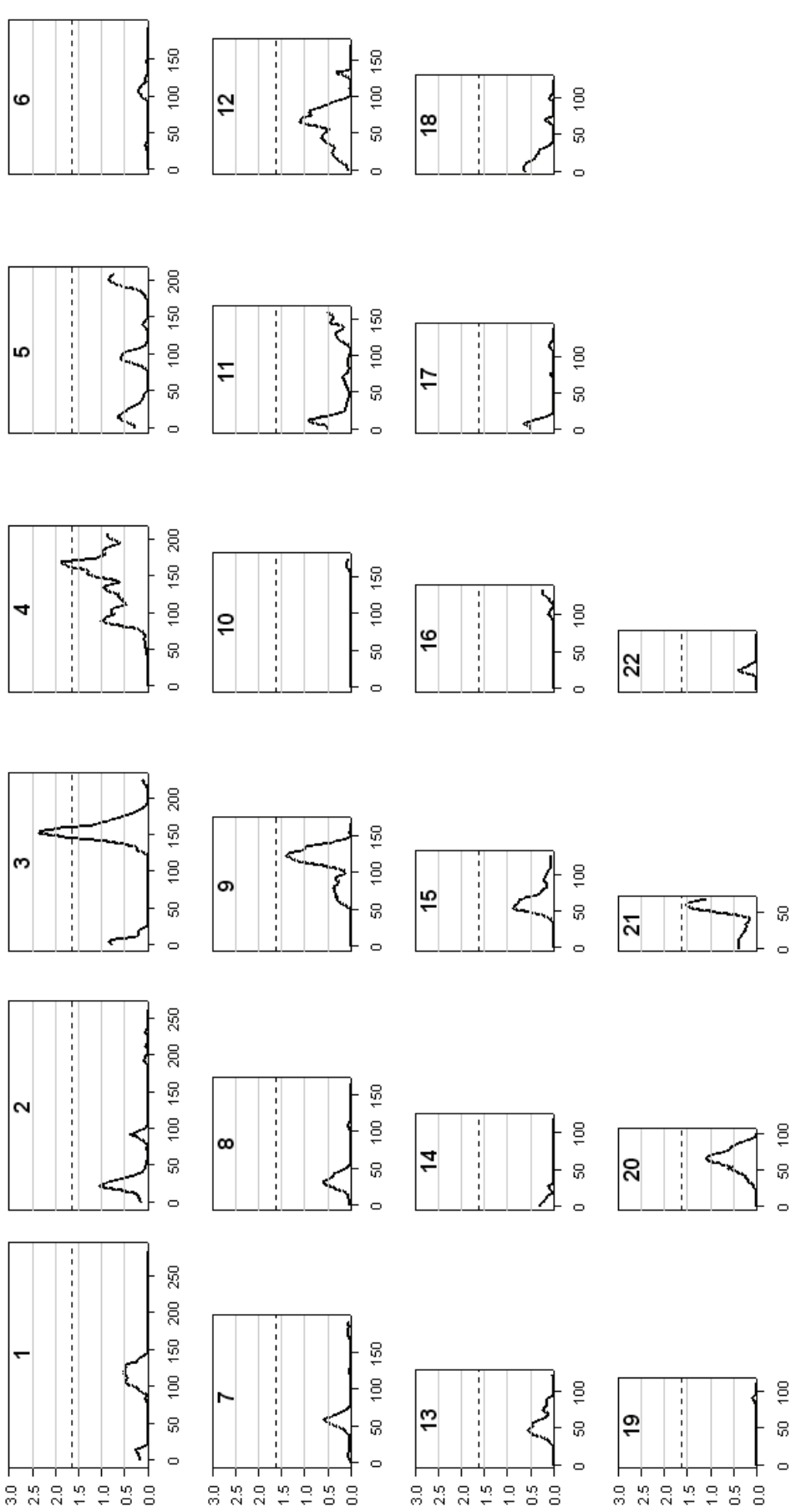

Figure 1
LOD scores of the genome-wide linkage scan for athlete status in the sample of genotyped DZ twin pairs.
Note: dashed line indicates the empirical threshold for suggestive linkage

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

Discussion

Seven per cent of the British adult women have ever competed in sports at the county or national level. Women from older cohorts have less frequently participated in sports in general and also at the county or national level. Athlete status in women is a heritable trait: around 66% of the variance in athlete status is explained by additive genetic factors. The remaining variance is due to nonshared environmental factors. Nonadditive genetic factors and shared environmental factors do not play a role in explaining individual differences in athlete status.

The heritability of athlete status is consistent with the significant heritabilities that have been reported for a range of measures of athletic abilities. For example, the heritabilities of different measures of oxygen uptake, such as maximal oxygen uptake in the sedentary state and in response to training and submaximal oxygen uptake at different power outputs, which are all indicators of aerobic capacity, ranged from 47% to 74% (Bouchard et al., 1998; Bouchard et al., 1999). Heritabilities of muscle strength phenotypes (muscle cross-sectional area and isometric, concentric, and eccentric strength) ranged from 31% to 85% (Thomis et al., 1998), with the majority of the genetic variance shared among the strength phenotypes (De Mars et al., 2007).

Suggestive linkage for athlete status was found on chromosome 3q24 (LOD = 2.35) at the SLC9A9 gene. This gene encodes a cell membrane transport protein that is part of the sodium/hydrogen exchange family and is expressed at high levels in heart and skeletal muscle. SLC9A9 has been related to attention deficit hyperactivity disorder (De Silva et al., 2003), but not to sports ability. We also examined whether there were genes listed in the 'human gene map' that were located at or nearby the peak on 3q24 (Wolfarth et al., 2005). The calcium-sensing receptor (CASR) gene is located just outside the confidence interval of the peak on 3q21. The CASR gene has been found to be associated with physical activity (hours per week during the last year) in adolescent girls, and is involved in calcium homeostasis (Lorentzon et al., 2001). It is well known that calcium is involved in muscle activity, bone formation, blood clotting, and nerve activity (Wilmore & Costill, 2004).

Suggestive linkage was also found on chromosome 4q32 (LOD = 1.87). There are two genes that are located in the region 4q28-31 that are listed in the 'human gene map'. The first is the fatty-acid binding protein 2 (FABP2) gene. In a linkage study carried out in 412 sibling pairs using 289 markers (Bouchard et al., 2000), suggestive linkage ($p < .01$) was found on 4q26 at marker FABP2 for maximal oxygen uptake in response to a 20 week exercise training program. Fatty-acid binding proteins are involved in the metabolism of long-chain fatty-acids. Through its involvement in metabolic processes, one could hypothesize that the FABP2 gene is related to sports performance and participation. The second gene located in the 4q28-31 region that is listed in the 'human gene map' is the uncoupling protein 1 (UCP1) gene. The UCP1 gene is involved in heat generation. In a sample of 395 offspring and 372 parents using 432 markers, suggestive linkage ($p < .01$) was found at the UCP1 gene for moderate to strenuous daily physical activity (Simonen et al., 2003). To conclude, the peaks on 3q and 4q reported in the present study are of potential interest because they are close to genes that have been previously related to aerobic fitness and physical activity.

A first limitation of this study is that we used a very general measure of athlete status. Because the grouping also reflected the contrast between regular exercisers and sedentary subjects, athlete status was partly confounded with voluntary exercise participation. Substantial heritability for leisure-time exercise participation (Simonen et al., 2002; Stubbe et al., 2006), has been found, ranging from 16 to 71%. Genes for athletic status may well constitute part of the heredity of voluntary regular exercise behaviour. Individuals with the genetic advantage to perform well at sports are more likely to continue participating in sports. They find that they are good at sports and they possibly also experience the physical need to be active, which together drives them to regularly participate in exercise. A G x E scenario may unfold where regular participation in exercise by those individuals 'who are good at it' provides the additional training that is needed to maximize actual performance, ultimately leading to elite athlete status (Brutsaert & Parra, 2006).

A second limitation is that the sports for which the twins could indicate their maximum level of competition were biased in favor of endurance sports, requiring mainly aerobic fitness, rather than strength or sprint capacity. The genes affecting endurance or power athlete status might be different, or different alleles of the same gene might be associated with endurance or power athlete status, the latter having been suggested for the ACE gene (Williams et al., 2000), and more recently also for the ACTN3 gene (MacArthur et al., 2007). It is increasingly acknowledged however that the distinction between endurance and power sports is not that sharp; many endurance sports also contain power elements and vice versa (Wilmore & Costill, 2004).

A third limitation of the study involves the inclusion of women from different ages in the sample. One could argue that the younger women (less than 30 years) might not have reached their maximum level of competition yet, which could have affected the linkage results. Still, the prevalence of competition at the county or national level is highest in the

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

younger adults, also in women aging between 20 and 25 years. The prevalence of competition at the county or national level is smallest in the oldest cohorts. The significant cohort effect on the prevalence of sporting ability can be explained by the fact that in the past it was less common for women to participate in sports at a competitive level. This cohort effect is unlikely to have biased the results of the genetic analyses however, since we regressed out the cohort effect in the analyses. It is possible, though, that the heritability of athlete status is higher in older cohorts since it was rarer, but we do not have the power to detect this.

A final limitation of this study relates to the sample size for the linkage scan. It is well known that in order to detect linkage signals, large sample sizes are needed, especially when categorical phenotypes are analysed (Neale et al., 1994). We therefore investigated the power to detect QTL effects of different magnitudes, given a sample size of 700 sibling pairs and a total heritability of 65%. The power to detect a QTL effect of 30% of the total variance is 0.91, and the power to detect a QTL effect of 25% of the variance is 0.79. For QTL effects explaining, respectively, 20%, 15%, and 10% of the variance, the power is 0.60, 0.38, and 0.20. Thus, the power to detect linkage signals of 25% or larger is good, and the power to detect smaller effects is moderate to low. The limited power of the present study underlines the need for replication studies.

To summarize, it was shown that athlete status in women is a heritable trait. A genome-wide linkage scan revealed suggestive linkages on chromosomes 3q22-q24 and 4q31-q34. The peak at 3q22-q24 is found at the SLC9A9 gene. The region 4q31-q34 overlaps with the region for which suggestive linkages were found in two previous linkage studies for physical fitness (FABP2 gene; Bouchard et al., 2000) and physical activity (UCP1 gene; Simonen et al., 2003). Future association studies should further clarify the possible role of these genes in athlete status.

Acknowledgements

This study was financially supported by the Netherlands Organization for Scientific Research (NWO; NWO-904-61-193 and NWO 480-05-003). The TwinsUK cohort is supported by the Wellcome trust and EU FP6 and FP7 grants. We thank Scott Wilson, Peter Reid and Gemini-Sequenom for genotyping. We thank Danielle Posthuma for her help with carrying out the statistical analyses on the Genetic Cluster Computer (http://www.geneticcluster.org).

References

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics, 30*, 97–101.

Abecasis, G. R. & Wigginton, J. E. (2005). Handling marker-marker linkage disequilibrium: Pedigree analysis with clustered markers. *American Journal of Human Genetics, 77*, 754–767.

Boomsma, D. I., Cacioppo, J. T., Slagboom, P. E., & Posthuma, D. (2006). Genetic linkage and association analysis for loneliness in Dutch twin and sibling pairs points to a region on chromosome 12q23–24. *Behavior Genetics, 36*, 137–146.

Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J., Perusse, L., Leon, A. S., & Rao, D. C. (1999). Familial aggregation of VO2max response to exercise training: Results from the HERITAGE Family Study. *Journal of Applied Physiology, 87*, 1003–1008.

Bouchard, C., Daw, E. W., Rice, T., Perusse, L., Gagnon, J., Province, M. A., Leon, A. S., Rao, D. C., Skinner, J. S., & Wilmore, J. H. (1998). Familial resemblance for VO2max in the sedentary state: The HERITAGE Family Study. *Medicine and Science in Sports and Exercise, 30*, 252–258.

Bouchard, C. & Malina, R. M. (1998). Genetics of physiological fitness and motor performance. Exercise and Sport Sciences Reviews, 11, 306–339.

Bouchard, C., Rankinen, T., Chagnon, Y. C., Rice, T., Perusse, L., Gagnon, J., Borecki, I., An, P., Leon, A. S., Skinner, J. S., Wilmore, J. H., Province, M., & Rao, D. C. (2000). Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE Family Study. *Journal of Applied Physiology, 88*, 551–559.

Brutsaert, T. D. & Parra, E. J. (2006). What makes a champion? Explaining variation in human athletic performance. *Respiratory Physiology and Neurobiology, 151*, 109–123.

De Mars, G., Thomis, M. A. I., Windelinckx, A., Van Leemputte, M., Maes, H. H., Blimkie, C. J., Claessens, A. L., Vlietinck, R., & Beunen, G. (2007). Covariance of isometric and dynamic arm contractions: Multivariate genetic analysis. *Twin Research and Human Genetics, 10*, 180–190.

De Silva, M. G., Elliott, K., Dahl, H. H., Fitzpatrick, E., Wilcox, S., Delatycki, M., Williamson, R., Efron, D., Lynch, M., & Forrest, S. (2003). Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an attention deficit hyperactivity disorder-like phenotype. *Journal of Medical Genetics, 40*, 733–740.

Duffy, D. L. (2006). An integrated genetic map for linkage analysis. Behavior Genetics, 36, 4–6.

Lorentzon, M., Lorentzon, R., Lerner, U. H., & Nordstrom, P. (2001). Calcium sensing receptor gene polymorphism, circulating calcium concentrations and bone mineral density in healthy adolescent girls. *European Journal of Endocrinology, 144*, 257–261.

MacArthur, D. & North, K. (2005). Genes and human elite athletic performance. *Human Genetics, 116,* 331–339.

MacArthur, D. G. & North, K. N. (2007). ACTN3: A genetic influence on muscle function and athletic performance. *Exercise and Sport Sciences Reviews, 35*, 30–34.

MacArthur, D. G., Seto, J. T., Raftery, J. M., Quinlan, K. G., Huttley, G. A., Hook, J. W., Lemckert, F. A., Kee, A. J., Edwards, M. R., Berman, Y., Hardeman, E. C., Gunning, P. W., Easteal, S., Yang, N., & North, K. N. (2007). Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nature Genetics,* Published online.

Moore, G. E., Shuldiner, A. R., Zmuda, J. M., Ferrell, R. E., Mccole, S. D., & Hagberg, J. M. (2001). Obesity gene variant and elite endurance performance. *Metabolism-Clinical and Experimental, 50*, 1391–1392.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling.* (6th ed.) Richmond, Virginia: Department of Psychiatry, Virginia Commonwealth University.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.

Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics, 24,* 239–258.

Paparini, A., Ripani, M., Giordano, G. D., Santoni, D., Pigozzi, F., & Romano-Spica, V. (2007). ACTN3 genotyping by real-time PCR in the Italian population and athletes. *Medicine and Science in Sports and Exercise, 39*, 810–815.

Perusse, L., Gagnon, J., Province, M. A., Rao, D. C., Wilmore, J. H., Leon, A. S., Bouchard, C., & Skinner, J. S. (2001). Familial aggregation of submaximal aerobic performance in the HERITAGE Family study. *Medicine and Science in Sports and Exercise, 33*, 597–604.

Perusse, L., Rankinen, T., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2003). The human gene map for performance and health-related fitness phenotypes: The 2002 update. *Medicine and Science in Sports and Exercise, 35*, 1248–1264.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2001). The human gene map for performance and health-related fitness phenotypes. *Medicine and Science in Sports and Exercise, 33*, 855–867.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2002). The human gene map for performance and health-related fitness phenotypes: The 2001 update. *Medicine and Science in Sports and Exercise, 34*, 1219–1233.

Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2004). The human gene map for performance and health-related fitness phenotypes: The 2003 update. *Medicine and Science in Sports and Exercise, 36*, 1451–1469.

Sham, P. (1998). Statistics in human genetics. London: Arnold Applications of Statistics.

Simonen, R. L., Perusse, L., Rankinen, T., Rice, T., Rao, D. C., & Bouchard, C. (2002). Familial aggregation of physical activity levels in the Quebec family study. *Medicine and Science in Sports and Exercise, 34*, 1137–1142.

Simonen, R. L., Rankinen, T., Perusse, L., Rice, T., Rao, D. C., Chagnon, Y., & Bouchard, C. (2003). Genome-wide linkage scan for physical activity levels in the Quebec family study. *Medicine and Science in Sports and Exercise, 35*, 1355–1359.

Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J., & Hart, D. (1996). Genetic influences on osteoarthritis in women: A twin study. *British Medical Journal, 312,* 940–943.

Spector, T. D. & Williams, F. M. K. (2006). The UK Adult Twin Registry (TwinsUK). *Twin Research and Human Genetics, 9*, 899–906.

Stubbe, J. H., Boomsma, D. I., Vink, J. M., Cornes, B. K., Martin, N. G., Skytthe, A., Kyvik, K. O., Rose, R. J., Kujala, U. M., Kaprio, J., Harris, J. R., Pedersen, N. L., Hunkin, J., Spector, T. D., & de Geus, E. J. C. (2006). Genetic influences on exercise participation: A comparative study in adult twin samples from seven countries. *PLoS ONE*, 1, e22. doi:10.1371/journal.pone.0000022.

Thomis, M. A. I., Beunen, G. P., Maes, H. H., Blimkie, C. J., Van Leemputte, M., Claessens, A. L., Marchal, G., WIllems, E., & Vlietinck, R. F. (1998). Strength training: Importance of genetic factors. *Medicine and Science in Sports and Exercise,* 30, 724-731.

Williams, A. G., Dhamrait, S. S., Wootton, P. T. E., Day, S. H., Hawe, E., Payne, J. R., Myerson, S. G., World, M., Budgett, R., Humphries, S. E., & Montgomery, H. E. (2004). Bradykinin receptor gene variant and human physical performance. *Journal of Applied Physiology, 96*, 938-942.

Williams, A. G., Rayson, M. P., Jubb, M., World, M., Woods, D. R., Hayward, M., Martin, J., Humphries, S. E., & Montgomery, H. E. (2000). Physiology: The ACE gene and muscle performance. *Nature, 403*, 614.

Wilmore, J. H. & Costill, D. L. (2004). P*hysiology of sport and exercise.* Champaign, IL: Human Kinetics.

Wilson, S. G., Reed, P. W., Bansal, A., Chiano, M., Lindersson, M., Langdown, M., Prince, R. L., Thompson, D., Thompson, E., Bailey, M., Kleyn, P. W., Sambrook, P., Shi, M. M., & Spector, T. D. (2003). Comparison of genome screens for two independent cohorts provides replication of suggestive linkage of bone mineral density to 3p21 and 1p36. *American Journal of Human Genetics, 72*, 144-155.

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press

Wolfarth, B., Bray, M. S., Hagberg, J. M., Perusse, L., Rauramaa, R., Rivera, M. A., Roth, S. M., Rankinen, T., & Bouchard, C. (2005). The human gene map for performance and health-related fitness phenotypes: The 2004 update. *Medicine and Science in Sports and Exercise, 37*, 881-903.

Wolfarth, B., Rivera, M. A., Oppert, J. M., Boulay, M. R., Dionne, F. T., Chagnon, M., Gagnon, J., Chagnon, Y., Perusse, L., Keul, J., & Bouchard, C. (2000). A polymorphism in the alpha(2a)-adrenoceptor gene and endurance athlete status. *Medicine and Science in Sports and Exercise, 32,* 1709-1712.

Woods, D. R., Humphries, S. E., & Montgomery, H. E. (2000). The ACE I/D polymorphism and human physical performance. *Trends in Endocrinology and Metabolism, 11,* 416-420.

Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *American Journal of Human Genetics, 73*, 627-631.

https://doi.org/10.1375/twin.10.6.812 Published online by Cambridge University Press