This is a non-final version of an article published in final form in LARRUSKAIN, JON; CELORRIO, DAVID; BARRIO, IRANTZU; ODRIOZOLA, ADRIAN; GIL, SUSANA M.; FERNANDEZ-LOPEZ, JUAN R.2; NOZAL, RAUL; ORTUZAR, ISUSKO; LEKUE, JOSE A.; AZNAR, JOSE M. *Genetic Variants and Hamstring Injury in Soccer: An Association and Validation Study*. **Medicine & Science in Sports & Exercise** 50(2):p 361-368, February 2018. | DOI: 10.1249/MSS.0000000000001434

Genetic Variants and Hamstring Injury in Soccer: an Association and Validation Study

Jon Larruskain [1], David Celorrio [2], Irantzu Barrio [3], Adrian Odriozola [1], Susana M. Gil [4], Juan Ramon Fernandez-Lopez [2,5], Raul Nozal [2], Isusko Ortuzar [4,6], Jose A. Lekue [4,6], Jose M. Aznar [2]

[1] Department of Genetics, Physical Anthropology and Animal Physiology, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Leioa, Spain

[2] Baigene, Vitoria-Gasteiz, Spain

[3] Department of Applied Mathematics, Statistics and Operations Research, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Leioa, Spain

[4] Department of Physiology, Faculty of Medicine and Nursing, University of the Basque Country (UPV/EHU), Leioa, Spain

[5] Kirolene Public Centre for Sports Education, Durango, Spain

[6] Medical Services, Athletic Club, Lezama, Spain

Short title: Genetics and hamstring injury in soccer

Corresponding author: Jose M. Aznar, PhD

Baigene, Lascaray Research Center

Avda. Miguel de Unamuno, 3, 01006 Vitoria-Gasteiz, Spain.

Telephone: 00-34-696 081 961

Email: josem.aznaroviedo@gmail.com

ABSTRACT

Purpose: To investigate the association of candidate single nucleotide polymorphisms (SNPs) with non-contact hamstring muscle injuries in elite soccer players, and to create and validate a model to assess the risk of hamstring injury.

Methods: 107 elite male outfield players were prospectively followed for 6 seasons. Players were genotyped for 37 SNPs previously investigated in relation to musculoskeletal injuries. The association of SNPs, previous injury, age, level of play, position and anthropometric data with 129 hamstring injuries (413 observations) was investigated in the discovery phase (2010-2015), and a multivariable Cox-frailty model was created using forward selection. The model's discriminative ability was tested in the validation phase (2015-2016, 31 injuries, 98 observations) using Harrell's C index.

Results: Five SNPs were found to be significantly associated with hamstring injury in a multivariable model, *MMP3* (Matrix metalloproteinase-3) rs679620 (A vs. G, hazard ratio (HR)=2.06, 95% confidence interval (CI)=1.51-2.81), *TNC* (Tenascin-C) rs2104772 (A vs. T, HR=1.65, 95% CI=1.17-2.32), *IL6* (Interleukin-6) rs1800795 (GG vs. GC+CC, HR=1.68, 95% CI=1.11-2.53), *NOS3* (Nitric oxide synthase-3) rs1799983 (G vs. T, HR=1.35, 95% CI=1.01-1.79), and *HIF1A* (Hypoxia-inducible factor-1α) rs11549465 (CC vs. CT, HR=2.08, 95% CI=1.00-4.29). Age also entered the model (≥24 vs. <24 years, HR=2.10, 95% CI=1.29-3.42). The model showed acceptable discrimination in the discovery phase (C index=0.74), but not in the validation phase (C index=0.52).

Conclusion: Genetic variants appear to be involved in the etiology of hamstring injuries, but were not found to have predictive value by themselves. Further research, increasing the number of genetic variants and including environmental factors in complex multifactorial risk models is necessary.

Key words

elite; risk; football; screening; prevention

INTRODUCTION

Soccer injuries affect team performance negatively, have high economic costs and might induce long-term health consequences (1,2). Hamstring muscle injury is the most frequent injury in elite male soccer (3), and identifying those at risk and preventing hamstring injuries is a priority. Risk factor studies have revealed previous hamstring injury to be highly associated with the occurrence of hamstring injuries in male soccer players. Other risk factors, such as, older age, reduced hamstring flexibility, decreased hamstring strength or strength imbalances, as well as fatigue show conflicting or limited evidence (4).

Previous research has also suggested that a genetic susceptibility may contribute to the interindividual variation in musculoskeletal injury risk. Several single nucleotide polymorphisms (SNPs) located in genes responsible for encoding the structural and regulatory proteins of musculoskeletal soft tissues have been associated in case-control retrospective studies with injuries, such as anterior cruciate ligament (ACL) rupture and chronic Achilles tendinopathy (5,6). In contrast, very few studies have investigated non-contact muscle injuries (7,8). In addition, variants associated with exercise-induced muscle damage have been pointed out as potential markers of muscle injury risk (9). These polymorphisms might contribute to interindividual variation in the structural and functional properties of muscle and tendon, and their response to mechanical loading, thus potentially being implicated in the susceptibility to hamstring injury (5).

However, there is no evidence regarding the influence of genetic variants on the risk of hamstring injury. Since statistically significant associations might not be enough to predict players at risk of injury, the predictive ability of any test needs to be validated in independent samples (10). Thus, the aims of this study were to investigate the association of candidate genetic variants with non-contact hamstring injuries in elite soccer players over several

seasons, and to create a model to estimate the risk of hamstring injury and test its validity.

METHODS

Participants and study design

This study was approved by the Clinical Research Ethics Committee of the Basque Country (PI2014215). 107 male outfield players from Athletic Club voluntarily agreed to participate after receiving oral and written details outlining the study. Informed consent was obtained from each participant. All players were Caucasian from the Basque Country in Spain. Players were recruited and saliva samples were collected at the beginning of the 2014-2015 season. 28 players belonged to the First team, 43 to the two Reserves teams, and 36 to the two U19 teams. All players from the First, Reserves and U19 teams had been prospectively followed from the 2010-2011 season to the 2015-2016 season, and injury records, exposure time, and anthropometric data were collected by the medical and coaching staff following common procedures. The study was divided in two phases (Figure 1): 1) the discovery phase, from the 2010-2011 season to the 2014-2015 season, when the association between risk factors and hamstring injuries was investigated; and 2) the validation phase in the 2015-2016 season, when the predictive ability of the risk model was assessed.

Injury, exposure time, and anthropometric data recording

Time-loss injuries resulting from soccer training or match play were recorded following the consensus on definitions and data collection procedures outlined by the International Federation of Association Football (FIFA) (11). Non-contact hamstring injuries were recorded when a player was unable to participate in a future training session or match due to an injury to the hamstring muscle group, and was considered injured until the medical staff

cleared the player for full participation in training and match play. Structural-mechanical injuries, such as total and partial muscle ruptures, and functional injuries, such as fatigue-induced or neurogenic muscle hardening (hypertonia), were included (3). Injuries were confirmed through a clinical examination by the team doctor, and if indicated, the diagnosis was supported by ultrasonography and magnetic resonance imaging. Injuries during national team duties were also registered. Hamstring injuries with a sudden, identifiable onset were defined as acute injuries, while those with a gradual onset as overuse injuries. According to the number of days of absence injury severity was recorded as minimal (1-3 days), mild (4-7 days), moderate (8-28 days) or severe (>28 days). Recurrent hamstring injuries were those occurring in the same leg and during the same season as an index hamstring injury.

Individual player exposure time in training and matches (friendly and competitive), including national team exposure, was daily recorded in minutes. Anthropometric data were collected every 2 months approximately by the team doctor. Height was measured using a stadiometer (Añó Sayol, Barcelona, Spain), and body mass was measured with a portable balance (Seca, Bonn, Germany). Skinfold thicknesses were measured at 6 sites (triceps, subscapular, abdominal, suprailiac, thigh and calf) using a skinfold caliper (Harpenden, West Sussex, England) and the sum of these 6 skinfolds was calculated in millimeters.

Genotyping

37 SNPs previously investigated in relation to musculoskeletal injuries (6-8) or exercise-induced muscle damage (9) were selected for the study. The full list of SNPs and associated injuries are presented in Supplemental Table 1 (see Table, Supplemental Digital Content 1, associated injuries, genotype frequencies and missing data of the selected candidate SNPs). For a more detailed information on these SNPs readers are referred to recent reviews (6,9).

Saliva samples were obtained using buccal swabs (4N6FLOQSwab, Life Technologies, Carlsbad, CA, USA). DNA was extracted via QIAmp DNA Mini kit (Qiagen, Hilden, Germany), and quantified by fluorometry using Qubit (Life Technologies, Carlsbad, CA, USA). DNA samples were genotyped using SNP type assays in the Biomark HD system (Fluidigm, South San Francisco, CA, USA).

Statistical analysis

The required sample size was calculated using the powerSurvEpi package in R version 3.2.3 (R Core Team 2015, R Foundation for Statistical Computing, Vienna, Austria). With 80% power, a two-sided significance of 0.05 and injuries occurring in 25% of observations, to detect a hazard ratio (HR) of 2, the minimum required number of injuries was 89. Injury incidences are presented as the number of injuries/1000 player hours with 95% confidence intervals (CI). Descriptive data are presented as mean ± standard deviation (SD).

The Cox proportional hazards model with a frailty extension was used to investigate the association between risk factors and hamstring injuries in the discovery phase, using the survival package in R (12). This model accounts for varying exposure times between players (measured as total hours of exposure in each season), and the frailty term allows for correlation between observations from the same player to be accounted for (13,14). Observations started at the beginning of each season. Some players had no occurrence of injury during the season and contributed censored survival times; whereas, other players sustained one or more hamstring injuries and had multiple observations.

First, potential risk factors (37 SNPs, age, height, body mass, sum of 6 skinfolds, level of play, position and previous hamstring injury during the preceding or same season) were individually analyzed adjusting for the players' match exposure ratio (match hours/total hours

of exposure) (15). Individual analyses were separately performed for all, acute, overuse, severe and recurrent hamstring injuries. For recurrent hamstring injuries each observation started when a player suffered an index hamstring injury. The analysis of previous injuries included only prospectively recorded injuries in the club, and hence, the players' first season in the club was not considered for the analysis (15). Continuous variables were categorized according to the optimal cut-off value using the CatPredi package in R (16). Each SNP was analysed under dominant (aa+Aa vs. AA), recessive (aa vs. AA+Aa), overdominant (Aa vs. AA+aa) and log-additive (aa=2, Aa=1, AA=0) modes of inheritance, and the best mode for each SNP was selected based on the minimum *P* value.

Subsequently, and only for all hamstring injuries, variables with $P \leq 0.25$ were entered in a multivariable Cox-frailty model using forward selection. At each step, variables with $P \leq 0.05$ were separately added to the model, and the model with the smallest Akaike Information Criterion value was retained until no variable showed $P \leq 0.05$. HRs and 95% CIs were calculated. The proportional hazard assumption was assessed using the cox.zph function in R. Kaplan-Meier survival curves were plotted to illustrate the probability of remaining injury-free during a season using GraphPad Prism v.6.0c (GraphPad Software, La Jolla, CA, USA). The significance level was set at $P \leq 0.05$.

Finally, a risk score for each player relative to the average player within the dataset was estimated from the model. The discriminative ability of the model was tested separately in the discovery and validation phases calculating Harrell's C index. Harrell's C (for concordance) index estimates the probability that, of two randomly chosen players, the player with the higher risk score will be more likely to sustain an injury compared to the player with the lower risk score. Values of C index near 0.5 indicate that the model is as good as a random guess, while a value of 1.0 indicates that the model always discriminates players with a higher risk (17).

RESULTS

107 players (20±4 years, 179±5 cm, 72±6 kg, 51±12 mm of skinfolds) were followed up for at least one season for a total of 356 player-seasons (3±1 seasons per player). Descriptive data on player exposure and hamstring injuries are presented in Table 1. The discovery phase consisted of 413 observations and 129 hamstring injuries (107 players), whereas 98 observations and 31 hamstring injuries (67 players) were included in the validation phase (Figure 1). Genotype frequencies are presented in Supplemental Table 1 (see Table, Supplemental Digital Content 1, associated injuries, genotype frequencies and missing data of the selected candidate SNPs). Two SNPs had >5% missing data, *COL1A1* rs1800012 (10%) and *COL5A1* rs12722 (10%).

Analysis of individual SNPs revealed 7 polymorphisms significantly associated with the risk of hamstring injury (Table 2). Age was the only non-genetic variable significantly associated with hamstring injuries, even after adjusting for the level of play (≥24 vs. <24 years, HR=3.33, 95% CI=1.38-8.02, *P*=0.01). *MMP3* (Matrix metalloproteinase-3) rs679620 remained statistically significant when acute, overuse, severe and recurrent hamstring injuries were separately analysed (Table 3; see full Tables, Supplemental Digital Content 2, association between acute, overuse, severe and recurrent hamstring injuries and genetic and non-genetic factors in elite soccer players). Previous hamstring injury was significantly associated only with acute hamstring injuries. In a multivariable model, 5 SNPs and age were significantly associated with hamstring injury (Table 4). Kaplan-Meier survival curves for these variables are shown in Figure 2. These results show a higher hamstring injury risk for players older than 24 years, and for players with the *MMP3* rs679620 AA, *TNC* (Tenascin-C) rs2104772 AA, *IL6* (Interleukin-6) rs1800795 GG, *NOS3* (Nitric oxide synthase-3) rs1799983 GG, and *HIF1A* (Hypoxia-inducible factor-1α) rs11549465 CC genotypes. All

significant variables met the proportional hazards assumption. Finally, the C index of the model was 0.74 in the discovery phase and 0.52 in the validation phase.

DISCUSSION

Five SNPs and age were associated with hamstring injury in a Cox-frailty model

The most strongly associated SNP was *MMP3* rs679620 G/A, with each copy of the A allele increasing the risk of hamstring injury twice compared to the GG genotype. It was also the only SNP significantly associated with acute, overuse, severe and recurrent hamstring injuries. Matrix metalloproteinase-3 plays an important role in the maintenance of myofiber functional integrity by breaking down components of the extracellular matrix, and regulating skeletal muscle cell migration, differentiation and regeneration (18). *MMP3* rs679620 is in linkage disequilibrium with *MMP3* rs3025058 5A/6A (19), a functional promoter polymorphism. The 5A allele, which is linked to the A allele of rs679620, has been shown to result in a higher *MMP3* expression compared to the 6A allele (20). Conversely, this SNP was not associated with non-contact skeletal muscle injuries in a previous study in elite soccer players, although there are large methodological differences with the present study in terms of player ethnicity, statistical analysis and injury definition (7). Moreover, the GG genotype was overrepresented in individuals with Achilles tendinopathy compared to asymptomatic controls (21), but these findings were not replicated in another cohort and no association was found with the risk of ACL rupture (19).

Among the other significant SNPs, each A allele of *TNC* rs2104772 A/T was associated with a 1.65 times higher risk of hamstring injury. Tenascin-C is a glycoprotein that regulates cell-matrix interactions, plays an important role in the muscle damage-repair cycle, and provides strength and elasticity to withstand mechanical forces. It is expressed in regenerating myofibers, and in response to mechanical loading in the myotendinous junction, the most vulnerable site to injury (22). The T>A substitution results in a leucine to isoleucine

amino acid change in the fibronectin type III-D domain region of TNC that could cause structural instability and alter the molecular elasticity of the domain (23). The A allele was previously associated with Achilles tendinopathy (24), but not with non-contact muscle injuries (8).

The GG genotype of *IL6* rs1800795 G/C was associated with a 1.68 times higher risk of hamstring injury compared to the GC and CC genotypes. The cytokine interleukin-6 is produced by skeletal muscle following exercise, and it also targets skeletal muscle, paradoxically, as both stimulator of hypertrophic muscle growth and myogenesis, and promoter of atrophy and muscle wasting (25). The G allele appears to increase *IL6* gene transcription and plasma levels in response to stress stimuli (26), and it has been previously associated with Achilles tendinopathy (27), lumbar disc degeneration (28) and power/strength athlete status (29). In contrast, the CC genotype was associated with higher creatine kinase levels after eccentric exercise in healthy individuals (9,30).

Each G allele of *NOS3* rs1799983 G/T was associated with a 1.35 times higher risk of hamstring injury. Nitric oxide synthase-3 is the rate limiting enzyme for nitric oxide production. Nitric oxide has many relevant biological functions, such as, regulation of blood flow, muscle contractility, mitochondrial respiration and skeletal muscle injury repair (31). NOS3 produced from the G allele seems to be less susceptible to proteolytic cleavage, which might result in increased NOS3 activity and higher NO production (32). This SNP was not previously associated with Achilles tendinopathy (33).

The risk of hamstring injury was twice as high in players with the *HIF1A* rs11549465 CC genotype in comparison to those with the CT genotype. Hypoxia-inducible factor-1α is a transcription factor regulating several genes in response to hypoxia, stimulating angiogenesis and glycolytic metabolism (34). It can also be induced by mechanical loading, and it is an important component of matrix remodeling and skeletal myogenesis (34,35). Previously, the

T allele was linked with higher transcriptional activity of *HIF1A* (36) and power/strength athlete status (29), but this SNP was not associated with ACL injury (37) or lumbar disc degeneration (38).

Collectively, these five variants, or other closely linked polymorphisms, might influence musculotendinous integrity and function, and its response to mechanical loading. Nonetheless, mechanistic studies are required to unravel the molecular mechanisms behind these associations (5).

Lastly, players older than 24 years had a two times higher risk of injury compared to younger players, and the association was independent of the level of play. This association was observed also in overuse, but not acute hamstring injuries. Previous studies show conflicting evidence with regards to the effect of age, which may be due to differences in mean age and level of play between study cohorts (4). Older players might be at a higher risk of injury due to age-related physical changes or a greater likelihood of having suffered a previous hamstring injury (4,15).

The model did not have predictive ability in a subsequent independent season

The multivariable Cox model can be used to estimate the risk of injury of each player relative to the average player within the dataset. This might be useful to classify players into risk groups or to create a risk profile of each player if the risk of various injuries could be estimated. Unfortunately, despite an acceptable internal concordance (C index=0.74), the model was as good as a random guess in a subsequent independent season (C index=0.52). This means that of two random players, the player with the higher risk score was the one that would get injured only half of the time (17). This result shows the importance of appropriate

validation studies, as statistically significant associations might not translate into accurate predictive tests (10).

The lack of predictive ability may be due to several reasons. Sample size was small in the validation phase, and replication in larger samples might be necessary. However, the accuracy of a screening test in any given season is relevant for soccer clubs. The candidate gene approach is also limited, and the number of genetic variants investigated needs to be increased in order to understand the influence of genetics on musculoskeletal injury risk (5,39). Most importantly, injuries are multifactorial disorders, and the use of genetic tests is very limited without considering other potential risk factors (e.g. training load, fatigue, hamstring activation, eccentric strength and fascicle length, fixture congestion, high intensity running, compliance with preventive training) (4,5). In this regard, preventive exercises were performed routinely by all players in the study, including Nordic hamstring exercises, core exercises and strength training. However, this information was not registered, and it is a limitation of the study. Lastly, accurately identifying players at risk is challenging, and there are currently no screening tests available to predict sports injuries (10). Therefore, to understand such a complex phenomenon a complex system approach may well be necessary, investigating the influence of genetic variants in interaction with other environmental risk factors (5,40).

In light of the present findings, the use of genetic testing for hamstring injuries in soccer seems premature. Due to the inaccuracy of current screening tests and the high frequency of hamstring injuries in elite soccer, all players should be included in hamstring injury prevention programs. Research in genetics, overcoming the limitations of the present study, still holds great potential for injury risk screening and prevention. Even though genetic testing will never be prognostic or predictive, it may provide information about the baseline injury risk of an individual. As an important piece of the multifactorial injury model, genetic

information might be used together with all other risk factors to identify those at high risk of injury and individualize preventive strategies (5,10).

Other methodological considerations

This is the first study to prospectively investigate genetic risk factors for hamstring injuries, in an ethnically homogeneous sample of elite soccer players, and using previously recommended statistical methods accounting for individual player exposure time and correlation between injuries (13,14). Nevertheless, the study has limited external validity as only players from one club were investigated, and findings remain to be replicated in other populations. Moreover, the study had adequate power to detect moderate HRs when including all hamstring injuries in the analysis, but sample size was insufficient for the analysis of specific types of hamstring injury. In this sense, the influence of genetics might be different for hamstring injuries with different mechanism, size and location, and hence, a larger sample of well-defined injuries is required. Finally, two important SNPs, *COL1A1* rs1800012 and *COL5A1* rs12722, had 10% missing genotype data due to problems with the genotyping, which might have influenced the results.

Conclusion

Five SNPs (*MMP3* rs679620, *TNC* rs2104772, *IL6* rs1800795, *NOS3* rs1799983 and *HIF1A* rs11549465) and older age were significantly associated with the risk of hamstring injury in a Cox-frailty model over 5 seasons in elite soccer players. *MMP3* rs679620 was the only variable individually associated with acute, overuse, severe and recurrent hamstring injuries. However, the model could not identify players at higher risk of injury in a subsequent

independent season, and genetic testing for hamstring injury risk seems premature at the moment. Further research in larger cohorts, increasing the number of genetic variants and including environmental risk factors would appear to be necessary to understand the influence of genetics on musculoskeletal injuries. Such evidence might be used in the future to assess the injury risk of a player and to make informed decisions about preventing hamstring injuries in soccer.

ACKNOWLEDGMENTS

The study was funded by Baigene. JL was supported by a PhD Studentship from the Vice-Chancellorship for Basque of the University of the Basque Country UPV/EHU (Euskararen arloko Errektoreordetza). IB acknowledges financial support from the Basque Government (IT620-13). SMG acknowledges financial support from the Basque Government (IT922-16) and the University of the Basque Country (PPG17/34). Genotyping was carried out by the Sequencing and Genotyping Unit of the University of the Basque Country (SGIker, UPV/EHU).

CONFLICTS OF INTEREST

The study was funded by the genetics company Baigene. DC, JRFL, RN and JMA are members of Baigene. For the remaining authors none were declared. The results of the present study do not constitute endorsement by ACSM. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. Ekstrand J. Keeping your top players on the pitch: the key to football medicine at a professional level. *Br J Sports Med.* 2013;47(12):723-4.
2. Turner AP, Barlow JH, Heathcote-Elliott C. Long term health impact of playing professional football in the United Kingdom. *Br J Sports Med.* 2000;34(5):332-6.
3. Ekstrand J, Hagglund M, Walden M. Epidemiology of muscle injuries in professional football (soccer). *Am J Sports Med.* 2011;39(6):1226-32.
4. van Beijsterveldt AM, van de Port IG, Vereijken AJ, Backx FJ. Risk factors for hamstring injuries in male soccer players: a systematic review of prospective studies. *Scand J Med Sci Sports.* 2013;23(3):253-62.
5. Collins M, September AV, Posthumus M. Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications. *Br J Sports Med.* 2015;49(23):1497-503.
6. Rahim M, Collins M, September A. Genes and Musculoskeletal Soft-Tissue Injuries. *Med Sport Sci.* 2016;61:68-91.
7. Pruna R, Artells R, Lundblad M, Maffulli N. Genetic biomarkers in non-contact muscle injuries in elite soccer players. *Knee Surg Sports Traumatol Arthrosc.* 2016. (Epub ahead of print).
8. Pruna R, Artells R, Ribas J, et al. Single nucleotide polymorphisms associated with non-contact soft tissue injuries in elite professional soccer players: influence on degree of injury and recovery time. *BMC Musculoskelet Disord.* 2013;14:221.
9. Baumert P, Lake MJ, Stewart CE, Drust B, Erskine RM. Genetic variation and exercise-induced muscle damage: implications for athletic performance, injury and ageing. *Eur J Appl Physiol.* 2016;116(9):1595-625.

10. Bahr R. Why screening tests to predict injury do not work-and probably never will...: a critical review. *Br J Sports Med.* 2016;50(13):776-80.
11. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med.* 2006;40(3):193-201.
12. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000.
13. Gabbett TJ, Ullah S, Finch CF. Identifying risk factors for contact injury in professional rugby league players - application of a frailty model for recurrent injury. *J Sci Med Sport.* 2012;15(6):496-504.
14. Ullah S, Gabbett TJ, Finch CF. Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med.* 2014;48(17):1287-93.
15. Hagglund M, Walden M, Ekstrand J. Risk factors for lower extremity muscle injury in professional soccer: the UEFA Injury Study. *Am J Sports Med.* 2013;41(2):327-35.
16. Barrio I, Rodríguez-Álvarez MX, Meira-Machado L, Esteban C, Arostegui I. Comparison of two discrimination indexes in the polycothomisation of continuous predictors in time-to-event studies. *SORT.* 2017;41:1-20.
17. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361-87.
18. Chen X, Li Y. Role of matrix metalloproteinases in skeletal muscle: migration, differentiation, regeneration and fibrosis. *Cell Adh Migr.* 2009;3(4):337-41.
19. Gibbon A, Hobbs H, van der Merwe W, et al. The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups. *J Sports Sci.* 2017;35(7):655-62.

20. Medley TL, Kingwell BA, Gatzka CD, Pillay P, Cole TJ. Matrix metalloproteinase-3 genotype contributes to age-related aortic stiffening through modulation of gene and protein expression. *Circ Res.* 2003;92(11):1254-61.
21. Raleigh SM, van der Merwe L, Ribbans WJ, Smith RK, Schwellnus MP, Collins M. Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. *Br J Sports Med.* 2009;43(7):514-20.
22. Fluck M, Mund SI, Schittny JC, Klossner S, Durieux AC, Giraud MN. Mechano-regulated tenascin-C orchestrates muscle repair. *Proc Natl Acad Sci U S A.* 2008;105(36):13662-7.
23. Matsuda A, Hirota T, Akahoshi M, et al. Coding SNP in tenascin-C Fn-III-D domain associates with adult asthma. *Hum Mol Genet.* 2005;14(19):2779-86.
24. Saunders CJ, van der Merwe L, Posthumus M, et al. Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *J Orthop Res.* 2013;31(4):632-7.
25. Muñoz-Canoves P, Scheele C, Pedersen BK, Serrano AL. Interleukin-6 myokine signaling in skeletal muscle: a double-edged sword? *FEBS J.* 2013;280(17):4131-48.
26. Fishman D, Faulds G, Jeffery R, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest.* 1998;102(7):1369-76.
27. September AV, Nell EM, O'Connell K, et al. A pathway-based approach investigating the genes encoding interleukin-1beta, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy. *Br J Sports Med.* 2011;45(13):1040-7.

28. Kelempisioti A, Eskola PJ, Okuloff A, et al. Genetic susceptibility of intervertebral disc degeneration among young Finnish adults. *BMC Med Genet.* 2011;12:153.

29. Ahmetov, II, Fedotovskaya ON. Current Progress in Sports Genomics. *Adv Clin Chem.* 2015;70:247-314.

30. Yamin C, Duarte JA, Oliveira JM, et al. IL6 (-174) and TNFA (-308) promoter polymorphisms are associated with systemic creatine kinase response to eccentric exercise. *Eur J Appl Physiol.* 2008;104(3):579-86.

31. Stamler JS, Meissner G. Physiology of nitric oxide in skeletal muscle. *Physiol Rev.* 2001;81(1):209-37.

32. Leeson CP, Hingorani AD, Mullen MJ, et al. Glu298Asp endothelial nitric oxide synthase gene polymorphism interacts with environmental and dietary factors to influence endothelial function. *Circ Res.* 2002;90(11):1153-8.

33. Nell EM, van der Merwe L, Cook J, Handley CJ, Collins M, September AV. The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. *J Orthop Res.* 2012;30(11):1719-24.

34. Lindholm ME, Rundqvist H. Skeletal muscle hypoxia-inducible factor-1 and exercise. *Exp Physiol.* 2016;101(1):28-32.

35. Petersen W, Varoga D, Zantop T, Hassenpflug J, Mentlein R, Pufe T. Cyclic strain influences the expression of the vascular endothelial growth factor (VEGF) and the hypoxia inducible factor 1 alpha (HIF-1alpha) in tendon fibroblasts. *J Orthop Res.* 2004;22(4):847-53.

36. Tanimoto K, Yoshiga K, Eguchi H, et al. Hypoxia-inducible factor-1alpha polymorphisms associated with enhanced transactivation capacity, implying clinical significance. *Carcinogenesis.* 2003;24(11):1779-83.

37. Rahim M, Gibbon A, Hobbs H, et al. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *J Orthop Res.* 2014;32(12):1612-8.

38. Lin WP, Wang XJ, Wang CR, et al. Polymorphism in the hypoxia-inducible factor 1alpha gene may confer susceptibility to LDD in Chinese cohort. *PLoS One.* 2013;8(8):e73158.

39. Bouchard C. Exercise genomics - a paradigm shift is needed: a commentary. *Br J Sports Med.* 2015;49(23):1492-6.

40. Bittencourt NF, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. *Br J Sports Med.* 2016;50:1309-14.

FIGURE CAPTIONS

Figure 1. Schematic diagram of the study design.

Figure 2. Kaplan-Meier survival curves illustrating the probability of remaining hamstring injury free during a season for the risk factors significantly associated with hamstring injury in a multivariable Cox-frailty model.

LIST OF SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 1.docx

Supplemental Digital Content 2.docx

Table 1. Descriptive data on player exposure and hamstring injuries.

Exposure hours	Total	Per player-season*
Total	97421	274 ± 81
Training	84068	236 ± 69
Match	13353	38 ± 19
Non-contact hamstring injuries	Number (%)	Incidence/1000 hours (95% CI)
Total	160	1.64 (1.41-1.92)
Training	60 (38)	0.71 (0.55-0.92)
Match	100 (62)	7.49 (6.16-9.11)
Mechanism		
Traumatic	67 (42)	0.69 (0.54-0.87)
Overuse	93 (58)	0.95 (0.78-1.17)
Severity		
Minimal	36 (23)	0.37 (0.27-0.51)
Mild	54 (34)	0.55 (0.42-0.72)
Moderate	55 (34)	0.56 (0.43-0.74)
Severe	15 (9)	0.15 (0.09-0.26)
Recurrence		
<2 months	24 (15)	0.25 (0.17-0.37)
Same season	40 (25)	0.41 (0.30-0.56)

*Values are mean ± SD.
CI: confidence interval.

Table 2. Individual analysis of genetic and non-genetic risk factors for hamstring injuries in elite soccer players.

Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
MMP3	rs679620	A		G	1.79 (1.27-2.51)	0.001
COL5A1	rs16399	DI		DD+II	1.83 (1.13-2.97)	0.01
MMP1	rs1799750	DD+DI		II	2.05 (1.13-3.74)	0.02
NOS3	rs1799983	G		T	1.43 (1.02-1.99)	0.04
DCN	rs516115	A		G	1.51 (1.02-2.22)	0.04
HIF1A	rs11549465	CC		CT	2.32 (1.03-5.20)	0.04
MMP12	rs2276109	A		G	1.62 (0.99-2.64)	0.05
CASP8	rs3834129	DD		DI+II	1.62 (0.98-2.67)	0.06
ADAM12	rs3740199	GG+GC		CC	1.79 (0.93-3.45)	0.08
SOX15	rs4227	TT+TG		GG	6.38 (0.79-51.67)	0.08
COL5A1	rs12722	TC+CC		TT	1.59 (0.90-2.80)	0.11
TNC	rs2104772	A		T	1.35 (0.92-1.99)	0.12
COL1A1	rs1107946	C		A	1.68 (0.80-3.50)	0.17
CCL2	rs2857656	GG+GC		CC	2.16 (0.71-6.63)	0.18
VEGFA	rs2010963	GG+GC		CC	4.48 (0.49-40.52)	0.18
ADAMTS5	rs226794	GA+AA		GG	1.44 (0.83-2.48)	0.20
ACTN3	rs1815739	CC		CT+TT	1.39 (0.83-2.32)	0.21
ACAN	rs1516797	TG+GG		TT	1.37 (0.82-2.28)	0.23
ADAMTS2	rs1054480	CT+TT		CC	1.35 (0.82-2.21)	0.24
IL6	rs1800795	GG		GC+CC	1.33 (0.81-2.18)	0.25
GDF5	rs143383	TT+CC		TC	1.33 (0.81-2.19)	0.26
ACE	rs1799752	D		I	1.25 (0.84-1.85)	0.27
COL12A1	rs970547	AA+AG		GG	2.11 (0.53-8.40)	0.29
SOD2	rs4880	T		C	1.19 (0.86-1.67)	0.29
MLCK	rs2700352	TT		CC+CT	1.61 (0.61-4.24)	0.34
TIMP2	rs4789932	CC+TT		CT	1.21 (0.73-1.99)	0.46
IL6R	rs2228145	AA+AC		CC	1.24 (0.69-2.24)	0.46
ADAMTS14	rs4747096	AG		AA	1.22 (0.71-2.08)	0.47
EMILIN1	rs2289360	GG+AA		GA	1.19 (0.71-1.97)	0.51
CASP8	rs1045485	GG+GC		CC	1.86 (0.20-17.16)	0.58
TTN	rs2742327	AA+GG		AG	1.15 (0.68-1.93)	0.61
IGF2	rs3213221	CG		CC+GG	1.13 (0.68-1.87)	0.63
COL1A1	rs1800012	G		T	1.09 (0.75-1.60)	0.64
TNF	rs1800629	G		A	1.12 (0.67-1.88)	0.66
IL1A	rs1800587	T		C	1.07 (0.73-1.56)	0.72
CCR2	rs768539	T		C	1.06 (0.68-1.64)	0.81
IL1B	rs1143634	CC+TT		CT	1.03 (0.61-1.76)	0.90
	Age (years)	≥ 24		< 24	2.33 (1.30-4.17)	0.004
	Height (cm)	≥ 177		< 177	1.60 (0.91-2.80)	0.10
	Body mass (kg)	≥ 75		< 75	1.44 (0.91-2.28)	0.12
	Sum 6 skinfolds (mm)	< 54		≥ 54	1.38 (0.84-2.27)	0.20
	Previous injury	Yes		No	1.19 (0.80-1.75)	0.34
	Level of play	First team		U19	1.43 (0.78-2.63)	0.25
		First team		Reserves	1.04 (0.60-1.82)	0.88
	Position	For		Def	1.30 (0.68-2.50)	0.42
		Def		Mid	1.04 (0.59-1.85)	0.89

*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.

HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion. Def: defender, Mid: midfielder, For: forward.

Table 3. Individual analysis of genetic and non-genetic risk factors for hamstring injury subtypes in elite soccer players.

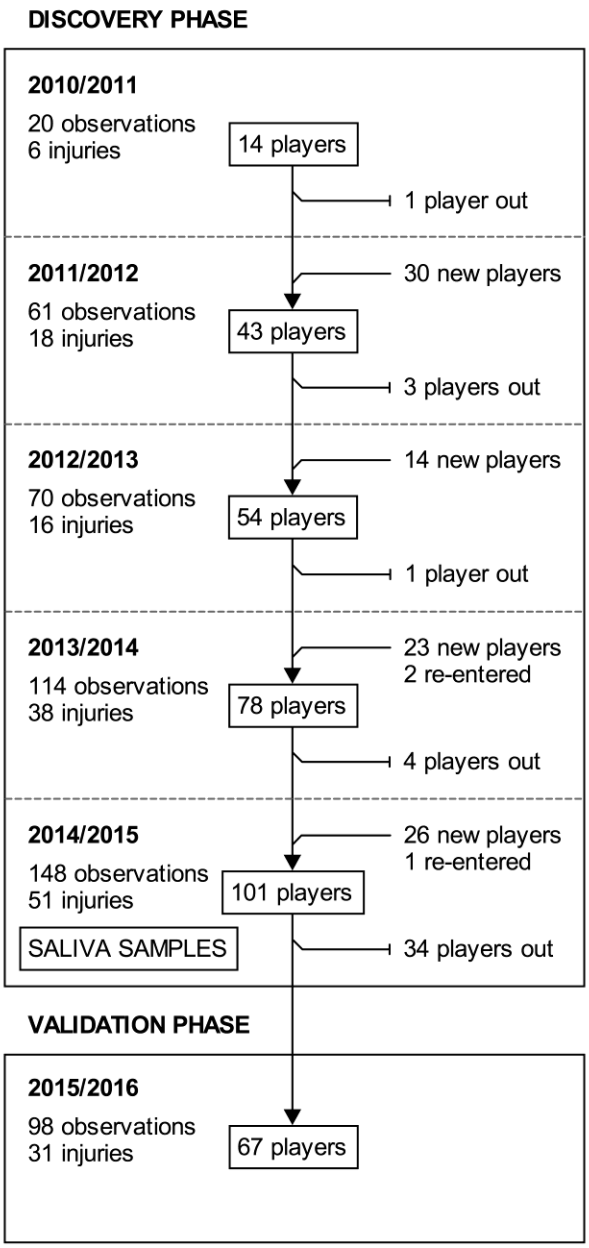
Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
Acute						
ACAN	rs1516797	GG		TT+TG	3.30 (1.46-7.44)	0.004
MMP3	rs679620	A		G	1.96 (1.18-3.24)	0.01
DCN	rs516115	A		G	2.20 (1.19-4.08)	0.01
MMP1	rs1799750	D		I	1.86 (1.13-3.06)	0.01
ADAMTS14	rs4747096	AG		AA	2.22 (1.09-4.53)	0.03
COL5A1	rs16399	DI		DD+II	2.03 (1.01-4.09)	0.05
	Previous injury	Yes		No	1.81 (1.00-3.28)	0.05
Overuse						
	Age (years)	≥ 24		< 24	2.95 (1.52-5.74)	0.001
MMP3	rs679620	A		G	1.66 (1.12-2.45)	0.01
Severe						
MLCK	rs2700352	TT		CC+CT	8.69 (2.42-31.18)	0.001
IL1A	rs1800587	CT		CC+TT	6.60 (1.74-25.02)	0.01
MMP3	rs679620	A		G	3.58 (1.33-9.66)	0.01
ADAMTS14	rs4747096	AG		AA	4.49 (1.18-17.15)	0.03
CASP8	rs3834129	DD+II		DI	4.36 (1.03-18.51)	0.05
Recurrent						
EMILIN1	rs2289360	GA		GG+AA	2.48 (1.00-6.14)	0.05
MMP3	rs679620	A		G	2.02 (0.99-4.13)	0.05

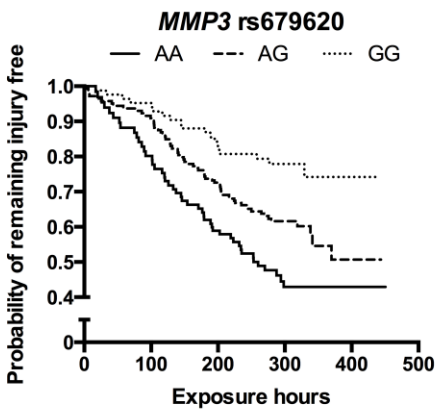
Only SNPs/variables with $P \leq 0.05$ are shown.
Sample size: Acute = 338 observations, 53 injuries; Overuse = 364 observations, 76 injuries; Severe = 300 observations, 13 injuries; Recurrent = 117 observations, 35 injuries.
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion.

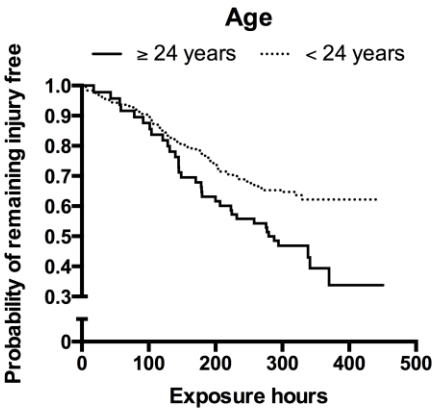
Table 4. Multivariable Cox-frailty model for the association between risk factors and hamstring injuries in elite soccer players.

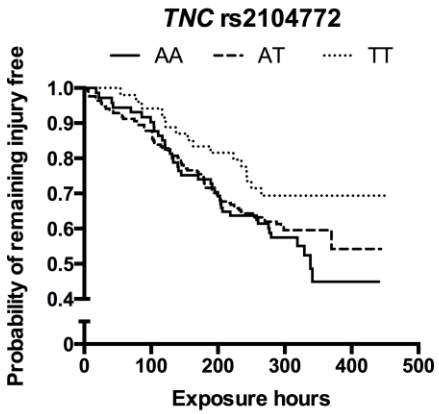
Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
MMP3	rs679620	A		G	2.06 (1.51-2.81)	6.2×10^{-6}
	Age	≥ 24		< 24	2.10 (1.29-3.42)	0.003
TNC	rs2104772	A		T	1.65 (1.17-2.32)	0.004
IL6	rs1800795	GG		GC+CC	1.68 (1.11-2.53)	0.01
NOS3	rs1799983	G		T	1.35 (1.01-1.79)	0.04
HIF1A	rs11549465	CC		CT	2.08 (1.00-4.29)	0.05

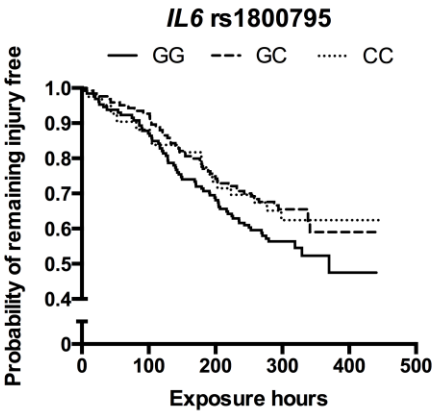
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury. CI: confidence interval.

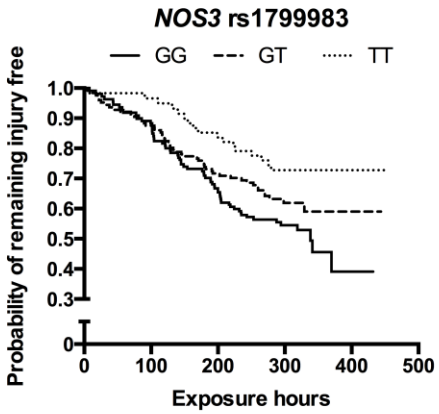

DISCOVERY PHASE
2010/2011
20 observations
6 injuries
14 players
1 player out
2011/2012
61 observations
18 injuries
30 new players
43 players
3 players out
2012/2013
70 observations
16 injuries
14 new players
54 players
1 player out
2013/2014
114 observations
38 injuries
23 new players
2 re-entered
78 players
4 players out
2014/2015
148 observations
51 injuries
26 new players
1 re-entered
101 players
SALIVA SAMPLES
34 players out
VALIDATION PHASE
2015/2016
98 observations
31 injuries
67 players

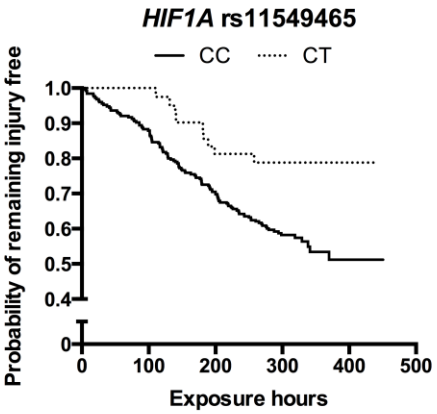

MMP3 rs679620
AA
AG
GG
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
0
100
200
300
400
500
Exposure hours


Age
≥ 24 years
< 24 years
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
0
100
200
300
400
500
Exposure hours


TNC rs2104772
AA
AT
TT
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
0
100
200
300
400
500
Exposure hours


IL6 rs1800795
GG
GC
CC
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
0
100
200
300
400
500
Exposure hours


NOS3 rs1799983
GG
GT
TT
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
0
100
200
300
400
500
Exposure hours


HIF1A rs11549465
CC
CT
Probability of remaining injury free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
0
100
200
300
400
500
Exposure hours

Supplemental Table 1. Associated injuries, genotype frequencies and missing data of the selected candidate SNPs.

Gene	Encoded protein	SNP	Injuries	Genotype frequencies, n (%)			% NA players	% NA observ.
ACAN	Aggrecan	rs1516797	ACL, LDD	TT 48 (45)	TG 46 (44)	GG 12 (11)	0.9	0.2
ACE	Angiotensin converting enzyme	rs1799752	EIMD	DD 25 (23)	DI 61 (57)	II 21 (20)	-	-
ACTN3	α-actinin-3	rs1815739	ANK, EIMD	CC 34 (32)	CT 55 (51)	TT 18 (17)	-	-
ADAM12	A disintegrin and metalloproteinase domain 12	rs3740199	AT*	GG 29 (27)	GC 57 (53)	CC 21 (20)	-	-
ADAMTS2	ADAM with thrombospondin type 1 motif, 2	rs1054480	AT*	CC 49 (46)	CT 50 (47)	TT 8 (7)	-	-
ADAMTS5	ADAM with thrombospondin type 1 motif, 5	rs226794	AT*	GG 81 (76)	GA 23 (21)	AA 3 (3)	-	-
ADAMTS14	ADAM with thrombospondin type 1 motif, 14	rs4747096	AT*	AA 75 (70)	AG 32 (30)	GG 0 (0)	-	-
CASP8	Caspase-8	rs1045485	AT	GG 75 (70)	GC 29 (27)	CC 3 (3)	-	-
CASP8	Caspase-8	rs3834129	AT	DD 39 (37)	DI 49 (47)	II 17 (16)	1.9	1.0
CCL2	Chemokine (C-C Motif) Ligand 2	rs2857656	MUS	GG 56 (53)	GC 41 (39)	CC 9 (8)	0.9	0.7
CCR2	Chemokine (C-C Motif) Receptor 2	rs768539	EIMD	CC 44 (44)	CT 53 (52)	TT 4 (4)	5.6	3.6
COL1A1	α1(I) collagen chain	rs1107946	ACL	CC 92 (87)	CA 13 (12)	AA 1 (1)	0.9	0.2
COL1A1	α1(I) collagen chain	rs1800012	ACL, LDD, SD	GG 54 (57)	GT 31 (33)	TT 10 (10)	10.9	10.4
COL5A1	α1(V) collagen chain	rs16399	AT	DD 53 (50)	DI 46 (43)	II 8 (7)	-	-
COL5A1	α1(V) collagen chain	rs12722	ACL, AT, CTS, MUS	TT 29 (30)	TC 51 (53)	CC 16 (17)	10.3	9.0
COL12A1	α1(XII) collagen chain	rs970547	ACL	AA 67 (62)	AG 34 (32)	GG 6 (6)	-	-
DCN	Decorin	rs516115	ACL	AA 56 (52)	AG 39 (37)	GG 12 (11)	-	-
EMILIN1	Elastin microfibril interfacer-1	rs2289360	LIG	GG 41 (39)	GA 48 (45)	AA 17 (16)	0.9	0.2
GDF5	Growth differentiation factor-5	rs143383	AT, LDD, MEN	TT 32 (30)	TC 50 (47)	CC 25 (23)	-	-
HIF1A	Hypoxia-inducible factor-1, α-subunit	rs11549465	LDD	CC 87 (82)	CT 19 (18)	TT 0 (0)	0.9	0.2
IGF2	Insulin-like growth factor-2	rs3213221	EIMD, MUS	CC 36 (34)	CG 57 (53)	GG 14 (13)	-	-
IL1A	Interleukin-1α	rs1800587	LDD	CC 53 (50)	CT 46 (43)	TT 8 (7)	-	-
IL1B	Interleukin-1β	rs1143634	AT, EIMD	CC 65 (61)	CT 37 (34)	TT 5 (5)	-	-
IL6	Interleukin-6	rs1800795	AT, LDD, EIMD	GG 47 (44)	GC 46 (43)	CC 14 (13)	-	-
IL6R	Interleukin-6 receptor	rs2228145	CTS	AA 31 (29)	AC 49 (46)	CC 26 (25)	0.9	0.2
MLCK	Myosin light-chain kinase	rs2700352	EIMD	CC 72 (67)	CT 29 (27)	TT 6 (6)	-	-
MMP1	Matrix metalloproteinase-1	rs1799750	ACL, LDD, PTT	II 30 (28)	ID 48 (45)	DD 29 (27)	-	-
MMP3	Matrix metalloproteinase-3	rs679620	ACL, AT	GG 30 (28)	GA 54 (51)	AA 22 (21)	0.9	0.5
MMP12	Matrix metalloproteinase-12	rs2276109	ACL	AA 67 (64)	AG 35 (33)	GG 3 (3)	1.9	0.7
NOS3	Nitric oxide synthase-3	rs1799983	AT*	GG 38 (35)	GT 48 (45)	TT 21 (20)	-	-
SOD2	Superoxide dismutase 2, mitochondrial	rs4880	EIMD	TT 32 (30)	TC 48 (46)	CC 25 (24)	1.9	0.7
SOX15	SRY-related HMG-box 15	rs4227	MUS	TT 64 (60)	TG 35 (33)	GG 8 (7)	-	-
TIMP2	Metalloproteinase inhibitor-2	rs4789932	AT	CC 40 (37)	CT 51 (48)	TT 16 (15)	-	-
TNC	Tenascin-C	rs2104772	AT	AA 25 (23)	AT 59 (56)	TT 22 (21)	0.9	0.5
TNF	Tumor necrosis factor	rs1800629	EIMD	GG 74 (69)	GA 31 (29)	AA 2 (2)	-	-
TTN	Titin	rs2742327	MUS*	AA 60 (56)	AG 41 (38)	GG 6 (6)	-	-
VEGFA	Vascular endothelial growth factor A	rs2010963	ACL	GG 46 (43)	GC 57 (53)	CC 4 (4)	-	-

SNP: single nucleotide polymorphism.. I: insertion, D: deletion. NA: missing values. Observ.: observations.
ACL: anterior cruciate ligament injury, ANK: ankle sprain, AT: Achilles tendinopathy, CTS: carpal tunnel syndrome, EIMD: exercise-induced muscle damage, LDD: lumbar disc disease, LIG: ligament injury, MEN: meniscus injury, MUS: muscle injury, SD: shoulder dislocation, PTT: tendinopathy of the posterior tibialis tendon.
*The association with the injury was not statistically significant.

Supplemental Table 2. Association between acute hamstring injuries and genetic and non-genetic factors in elite soccer players.

Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
ACAN	rs1516797	GG		TT+TG	3.30 (1.46-7.44)	0.004
MMP3	rs679620	A		G	1.96 (1.18-3.24)	0.01
DCN	rs516115	A		G	2.20 (1.19-4.08)	0.01
MMP1	rs1799750	D		I	1.86 (1.13-3.06)	0.01
ADAMTS14	rs4747096	AG		AA	2.22 (1.09-4.53)	0.03
COL5A1	rs16399	DI		DD+II	2.03 (1.01-4.09)	0.05
IL1A	rs1800587	T		C	1.64 (0.99-2.73)	0.06
HIF1A	rs11549465	CC		CT	4.40 (0.96-20.11)	0.06
NOS3	rs1799983	G		T	1.60 (0.96-2.67)	0.07
COL1A1	rs1800012	GG		GT+TT	1.91 (0.88-4.12)	0.10
ACE	rs1799752	DI		DD+II	1.88 (0.90-3.94)	0.10
ADAMTS5	rs226794	GA+AA		GG	1.84 (0.87-3.88)	0.11
ADAM12	rs3740199	GG+GC		CC	2.20 (0.80-6.04)	0.13
CASP8	rs3834129	DD		DI+II	1.74 (0.84-3.62)	0.14
TNC	rs2104772	AA		AT+TT	1.72 (0.80-3.71)	0.16
ADAMTS2	rs1054480	CT		CC+TT	1.63 (0.80-3.33)	0.18
IL6R	rs2228145	AA		AC+CC	1.63 (0.78-3.39)	0.19
GDF5	rs143383	CC		TT+TC	1.67 (0.76-3.65)	0.20
COL5A1	rs12722	TC+CC		TT	1.74 (0.74-4.12)	0.21
SOX15	rs4227	T		G	1.53 (0.78-2.99)	0.22
TNF	rs1800629	G		A	1.60 (0.73-3.51)	0.24
MMP12	rs2276109	AA		AG+GG	1.60 (0.73-3.49)	0.24
IL6	rs1800795	GG		GC+CC	1.50 (0.74-3.04)	0.26
CCR2	rs768539	CT		CC+TT	1.45 (0.70-3.00)	0.32
CCL2	rs2857656	GG+GC		CC	2.33 (0.44-12.47)	0.32
COL1A1	rs1107946	C		A	1.71 (0.57-5.11)	0.34
IL1B	rs1143634	T		C	1.31 (0.74-2.29)	0.35
VEGFA	rs2010963	GC		GG+CC	1.40 (0.68-2.89)	0.36
SOD2	rs4880	TT		TC+CC	1.40 (0.67-2.94)	0.37
MLCK	rs2700352	TT		CC+CT	1.85 (0.47-7.29)	0.38
TTN	rs2742327	AA+AG		GG	2.58 (0.28-24.01)	0.40
CASP8	rs1045485	GC		GG+CC	1.35 (0.61-2.96)	0.45
TIMP2	rs4789932	CC+CT		TT	1.35 (0.48-3.77)	0.57
COL12A1	rs970547	AG+GG		AA	1.22 (0.59-2.53)	0.59
IGF2	rs3213221	CC+CG		GG	1.33 (0.45-3.95)	0.61
ACTN3	rs1815739	C		T	1.08 (0.64-1.83)	0.77
EMILIN1	rs2289360	GA		GG+AA	1.03 (0.50-2.12)	0.95
	Previous injury	Yes		No	1.81 (1.00-3.28)	0.05
	Sum 6 skinfolds (mm)	< 54		≥ 54	1.64 (0.77-3.51)	0.20
	Height (cm)	≥ 177		< 177	1.68 (0.73-3.87)	0.23
	Body mass (kg)	≥ 75		< 75	1.39 (0.70-2.78)	0.35
	Age (years)	≥ 24		< 24	1.46 (0.61-3.51)	0.39
	Level of play	First team		U19	2.27 (0.83-6.25)	0.11
		Reserves		First team	1.55 (0.71-3.37)	0.27
	Position	For		Def	1.86 (0.79-4.39)	0.16
		Def		Mid	1.32 (0.56-3.03)	0.53

There were 338 observations and 53 acute hamstring injuries.
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion. Def: defender, Mid: midfielder, For: forward.

Supplemental Table 3. Association between overuse hamstring injuries and genetic and non-genetic factors in elite soccer players.

Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
MMP3	rs679620	A		G	1.66 (1.12-2.45)	0.01
MMP12	rs2276109	A		G	1.74 (0.97-3.11)	0.06
COL5A1	rs16399	DI+II		DD	1.65 (0.95-2.89)	0.08
NOS3	rs1799983	GG		GT+TT	1.65 (0.95-2.89)	0.08
ACE	rs1799752	DD		DI+II	1.65 (0.94-2.87)	0.08
ACTN3	rs1815739	CC+TT		CT	1.63 (0.93-2.87)	0.09
VEGFA	rs2010963	G		C	1.54 (0.91-2.62)	0.11
TNC	rs2104772	AA+AT		TT	1.94 (0.84-4.51)	0.12
TIMP2	rs4789932	CC+TT		CT	1.55 (0.87-2.75)	0.14
MMP1	rs1799750	DD+DI		II	1.67 (0.85-3.29)	0.14
CASP8	rs3834129	D		I	1.34 (0.90-2.00)	0.15
SOX15	rs4227	TT+TG		GG	3.74 (0.46-30.27)	0.22
ADAM12	rs3740199	GG+GC		CC	1.58 (0.75-3.32)	0.23
COL5A1	rs12722	TC+CC		TT	1.50 (0.77-2.91)	0.23
HIF1A	rs11549465	CC		CT	1.72 (0.70-4.21)	0.23
COL1A1	rs1107946	C		A	1.65 (0.71-3.86)	0.25
TTN	rs2742327	GG		AA+AG	1.85 (0.64-5.38)	0.26
CCL2	rs2857656	G		C	1.31 (0.82-2.09)	0.26
IL1A	rs1800587	C		T	1.28 (0.82-2.00)	0.28
EMILIN1	rs2289360	GG+AA		GA	1.37 (0.76-2.46)	0.29
ADAMTS2	rs1054480	T		C	1.25 (0.81-1.94)	0.31
GDF5	rs143383	TT		TC+CC	1.35 (0.75-2.43)	0.32
COL1A1	rs1800012	GT+TT		GG	1.32 (0.75-2.33)	0.34
IL1B	rs1143634	C		T	1.27 (0.78-2.08)	0.34
ADAMTS14	rs4747096	AA		AG	1.37 (0.72-2.60)	0.34
SOD2	rs4880	TT+TC		CC	1.38 (0.69-2.77)	0.36
ACAN	rs1516797	G		T	1.21 (0.80-1.83)	0.37
COL12A1	rs970547	A		G	1.26 (0.74-2.12)	0.40
TNF	rs1800629	GA		GG+AA	1.28 (0.71-2.31)	0.41
IL6	rs1800795	GG+CC		GC	1.26 (0.71-2.24)	0.43
CASP8	rs1045485	GG		GC+CC	1.30 (0.67-2.51)	0.44
DCN	rs516115	A		G	1.18 (0.77-1.83)	0.45
IL6R	rs2228145	AC		AA+CC	1.24 (0.70-2.19)	0.45
ADAMTS5	rs226794	A		G	1.20 (0.72-2.00)	0.48
MLCK	rs2700352	TT		CC+CT	1.46 (0.50-4.23)	0.49
IGF2	rs3213221	CG+GG		CC	1.24 (0.66-2.30)	0.51
CCR2	rs768539	CC+TT		CT	1.19 (0.67-2.11)	0.54
	Age (years)	≥ 24		< 24	2.95 (1.52-5.74)	0.001
	Body mass (kg)	≥ 75		< 75	1.51 (0.87-2.61)	0.15
	Height (cm)	≥ 177		< 177	1.52 (0.78-2.96)	0.21
	Sum 6 skinfolds (mm)	< 54		≥ 54	1.22 (0.68-2.19)	0.51
	Previous injury	Yes		No	1.14 (0.67-1.92)	0.64
	Level of play	First team		Reserves	1.47 (0.74-2.94)	0.28
		First team		U19	1.18 (0.60-2.33)	0.63
	Position	Def		For	1.10 (0.50-2.44)	0.81
		Mid		Def	1.06 (0.56-2.00)	0.87

There were 364 observations and 76 overuse hamstring injuries.
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion. Def: defender, Mid: midfielder, For: forward.

Supplemental Table 4. Association between severe hamstring injuries and genetic and non-genetic factors in elite soccer players.

Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
MLCK	rs2700352	TT		CC+CT	8.69 (2.42-31.18)	0.001
IL1A	rs1800587	CT		CC+TT	6.60 (1.74-25.02)	0.01
MMP3	rs679620	A		G	3.58 (1.33-9.66)	0.01
ADAMTS14	rs4747096	AG		AA	4.49 (1.18-17.15)	0.03
CASP8	rs3834129	DD+II		DI	4.36 (1.03-18.51)	0.05
IL1B	rs1143634	T		C	2.18 (0.94-5.08)	0.07
TTN	rs2742327	AG		AA+GG	3.29 (0.90-12.02)	0.07
ACAN	rs1516797	G		T	1.97 (0.85-4.56)	0.12
MMP1	rs1799750	DD		DI+II	2.76 (0.76-9.94)	0.12
MMP12	rs2276109	AA+GG		AG	3.78 (0.68-21.02)	0.13
SOD2	rs4880	TC		TT+CC	2.67 (0.75-9.55)	0.13
COL12A1	rs970547	GG		AA+AG	4.08 (0.63-26.52)	0.14
CASP8	rs1045485	GC		GG+CC	2.83 (0.70-11.49)	0.15
VEGFA	rs2010963	GC		GG+CC	2.67 (0.67-10.58)	0.16
IGF2	rs3213221	CC		CG+GG	2.52 (0.67-9.53)	0.17
ACE	rs1799752	DD		DI+II	2.48 (0.62-9.97)	0.20
CCR2	rs768539	T		C	1.96 (0.64-6.01)	0.24
IL6R	rs2228145	AA		AC+CC	2.06 (0.57-7.48)	0.27
EMILIN1	rs2289360	G		A	1.77 (0.63-4.99)	0.28
DCN	rs516115	AG		AA+GG	1.93 (0.57-6.52)	0.29
ACTN3	rs1815739	CT		CC+TT	2.05 (0.53-7.83)	0.30
NOS3	rs1799983	GG+GT		TT	2.97 (0.34-25.54)	0.32
COL5A1	rs12722	TC		TT+CC	1.91 (0.53-6.93)	0.32
TNF	rs1800629	G		A	2.15 (0.45-10.22)	0.34
ADAMTS2	rs1054480	CT+TT		CC	1.70 (0.47-6.16)	0.42
IL6	rs1800795	CC		GG+GC	1.79 (0.40-8.07)	0.45
COL5A1	rs16399	DI		DD+II	1.56 (0.44-5.55)	0.49
TNC	rs2104772	AA+AT		TT	2.08 (0.24-18.22)	0.51
SOX15	rs4227	T		G	1.36 (0.41-4.57)	0.62
TIMP2	rs4789932	C		T	1.25 (0.52-2.98)	0.62
HIF1A	rs11549465	CC		CT	1.73 (0.19-15.46)	0.63
GDF5	rs143383	TT+TC		CC	1.36 (0.25-7.30)	0.72
COL1A1	rs1800012	TT		GG+GT	1.38 (0.20-9.50)	0.74
ADAM12	rs3740199	GC		GG+CC	1.23 (0.35-4.30)	0.75
ADAMTS5	rs226794	GG		GA+AA	1.25 (0.22-7.08)	0.80
COL1A1	rs1107946	C		A	1.19 (0.24-5.96)	0.83
CCL2	rs2857656	GG+GC		CC	1.12 (0.11-11.33)	0.92
	Sum 6 skinfolds (mm)	< 54		≥ 54	5.00 (0.61-41.29)	0.13
	Previous injury	Yes		No	2.24 (0.72-6.93)	0.16
	Body mass (kg)	≥ 75		< 75	1.49 (0.42-5.28)	0.54
	Height (cm)	< 177		≥ 177	1.26 (0.32-4.90)	0.74
	Age (years)	≥ 24		< 24	1.09 (0.21-5.56)	0.92
	Level of play	First team		U19	1.03 (0.25-4.17)	0.39
		First team		Reserves	2.22 (0.36-14.29)	0.96
	Position	Def		For	1.23 (0.28-5.56)	0.52
		Mid		Def	1.68 (0.35-8.03)	0.78

There were 300 observations and 13 severe hamstring injuries.
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion. Def: defender, Mid: midfielder, For: forward.

Supplemental Table 5. Association between recurrent hamstring injuries and genetic and non-genetic factors in elite soccer players.

Gene	Variable	Higher risk*	vs.	Lower risk	HR (95% CI)	*P* value
EMILIN1	rs2289360	GA		GG+AA	2.48 (1.00-6.14)	0.05
MMP3	rs679620	A		G	2.02 (1.00-4.13)	0.05
ACTN3	rs1815739	CT+TT		CC	2.60 (0.97-6.97)	0.06
ADAM12	rs3740199	GG+CC		GC	2.42 (0.95-6.21)	0.06
MMP1	rs1799750	DD+DI		II	4.49 (0.85-23.73)	0.08
ACAN	rs1516797	TG+GG		TT	2.59 (0.88-7.65)	0.08
CCR2	rs768539	CT		CC+TT	2.22 (0.86-5.77)	0.10
TNF	rs1800629	G		A	1.55 (0.90-2.67)	0.12
MLCK	rs2700352	T		C	1.73 (0.86-3.48)	0.12
COL1A1	rs1107946	C		A	2.29 (0.76-6.93)	0.14
IGF2	rs3213221	C		G	1.72 (0.81-3.66)	0.16
SOD2	rs4880	TC+CC		TT	2.03 (0.74-5.58)	0.17
COL5A1	rs16399	DI		DD+II	1.83 (0.72-4.62)	0.20
TNC	rs2104772	AT		AA+TT	1.88 (0.68-5.15)	0.22
HIF1A	rs11549465	CC		CT	4.18 (0.41-42.35)	0.23
ACE	rs1799752	DI		DD+II	1.74 (0.68-4.44)	0.25
ADAMTS2	rs1054480	CC+CT		TT	3.82 (0.38-38.03)	0.25
GDF5	rs143383	TT+TC		CC	1.88 (0.61-5.80)	0.27
NOS3	rs1799983	G		T	1.43 (0.73-2.80)	0.30
COL5A1	rs12722	TC		TT+CC	1.64 (0.63-4.23)	0.31
TIMP2	rs4789932	C		T	1.43 (0.71-2.87)	0.32
CCL2	rs2857656	G		C	1.57 (0.63-3.93)	0.34
TTN	rs2742327	GG		AA+AG	2.38 (0.33-17.34)	0.39
MMP12	rs2276109	A		G	1.64 (0.51-5.24)	0.41
IL1B	rs1143634	C		T	1.43 (0.61-3.33)	0.41
IL6R	rs2228145	AA+AC		CC	1.61 (0.47-5.58)	0.45
IL1A	rs1800587	C		T	1.28 (0.63-2.60)	0.50
CASP8	rs3834129	DD		DI+II	1.38 (0.53-3.59)	0.51
ADAMTS5	rs226794	A		G	1.35 (0.55-3.34)	0.52
VEGFA	rs2010963	GC		GG+CC	1.32 (0.50-3.48)	0.57
DCN	rs516115	GG		AA+AG	1.79 (0.22-14.82)	0.59
COL1A1	rs1800012	G		T	1.23 (0.57-2.70)	0.60
COL12A1	rs970547	AG		AA+GG	1.24 (0.45-3.45)	0.68
CASP8	rs1045485	G		C	1.15 (0.42-3.18)	0.78
ADAMTS14	rs4747096	AG		AA	1.13 (0.40-3.19)	0.81
SOX15	rs4227	T		G	1.08 (0.42-2.77)	0.88
IL6	rs1800795	GG+GC		CC	1.04 (0.24-4.47)	0.96
	Age (years)	≥ 24		< 24	2.23 (0.78-6.39)	0.14
	Sum 6 skinfolds (mm)	< 54		≥ 54	1.94 (0.49-7.61)	0.34
	Body mass (kg)	≥ 75		< 75	1.40 (0.54-3.61)	0.49
	Height (cm)	≥ 177		< 177	1.23 (0.36-4.24)	0.75
	Level of play	First team		U19	2.70 (0.67-11.11)	0.16
		Reserves		First team	1.28 (0.44-3.69)	0.65
	Position	Def		Mid	1.49 (0.48-4.55)	0.49
		For		Def	1.11 (0.34-3.60)	0.87

There were 117 observations (starting after an index hamstring injury) and 35 recurrences within the same season.
*When alleles are compared the HR of the risk allele is additive (i.e., possessing two copies of the risk allele multiplies the HR by 2) compared to having two copies of the protective allele.
HR: hazard ratio of sustaining a hamstring injury adjusted for the players' match exposure ratio (match hours/total hours of exposure). CI: confidence interval. I: insertion, D: deletion. Def: defender, Mid: midfielder, For: forward.