International Journal of Sports Physiology and Performance, 2024, 19, 661-669
https://doi.org/10.1123/ijspp.2023-0184

Human Kinetics
ORIGINAL INVESTIGATION

Predicting Injuries in Elite Female Football Players With Global-Positioning-System and Multiomics Data

Juan R. González,[1,2,3] Alejandro Cáceres,[1,2] Eva Ferrer,[4,5] Laura Balagué-Dobón,[1] Xavier Escribà-Montagut,[1] David Sarrat-González,[1] Guillermo Quintás,[6] and Gil Rodas[4,5,6]

[1]Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain; [2]CIBER in Epidemiology and Public Health (CIBERESP), Barcelona, Spain; [3]Department of Mathematics, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain; [4]Medical Department of Football Club Barcelona (FIFA Medical Center of Excellence) and Barça Innovation Hub of Football Club Barcelona, Barcelona, Spain; [5]Sports and Exercise Medicine Unit, Hospital Clinic and Sant Joan de Déu, Barcelona, Spain; [6]Leitat Technological Center, Terrassa, Spain

Purpose: Injury prevention is a crucial aspect of sports, particularly in high-performance settings such as elite female football. This study aimed to develop an injury prediction model that incorporates clinical, Global-Positioning-System (GPS), and multiomics (genomics and metabolomics) data to better understand the factors associated with injury in elite female football players. ***Methods:*** We designed a prospective cohort study over 2 seasons (2019–20 and 2021–22) of noncontact injuries in 24 elite female players in the Spanish Premiership competition. We used GPS data to determine external workload, genomic data to capture genetic susceptibility, and metabolomic data to measure internal workload. ***Results:*** Forty noncontact injuries were recorded, the most frequent of which were muscle (63%) and ligament (20%) injuries. The baseline risk model included fat mass and the random effect of the player. Six genetic polymorphisms located at the *DCN, ADAMTS5, ESRRB, VEGFA*, and *MMP1* genes were associated with injuries after adjusting for player load ($P < .05$). The genetic score created with these 6 variants determined groups of players with different profile risks ($P = 3.1 \times 10^{-4}$). Three metabolites (alanine, serotonin, and 5-hydroxy-tryptophan) correlated with injuries. The model comprising baseline variables, genetic score, and player load showed the best prediction capacity (C-index: .74). ***Conclusions:*** Our model could allow efficient, personalized interventions based on an athlete's vulnerability. However, we emphasize the necessity for further research in female athletes with an emphasis on validation studies involving other teams and individuals. By expanding the scope of our research and incorporating diverse populations, we can bolster the generalizability and robustness of our proposed model.

Keywords: injury, genomic, metabolomic, longitudinal data

Injuries in sports, particularly among professional athletes, can have significant physical and economic implications.[1] Prognostic factors with an established causal role in injury occurrence and susceptibility factors (ie, genetic background) can be used to develop innovative intervention approaches to mitigate injury risk.[2] In addition, workload monitoring has practical applications to sports performance and injuries and has become an integral part of an athlete management system.[3] External workload (the physical demands placed on an athlete's body during physical activity) and internal workload (the physiological strain on an athlete's body) play a crucial role in the risk of injury in sports.[4] The former is commonly measured with the Global Positioning System (GPS), whereas the most appropriate approach for measuring internal workload is the analysis of the metabolome (ie, changes in the levels of specific metabolites in response to exercise).[5]

"Omics" are high-throughput, data-driven, holistic, and top-down methodologies examining all the constituents in a specific state. For instance, genomics entails the study of all genes and metabolomics all metabolic processes. Omics data play an important role in sports since they provide a comprehensive understanding of the biological processes and systems underlying athletic performance.[6] In particular, omics data have the potential to provide new insights into the complex biological mechanisms that contribute to sports injuries and to identify novel targets for intervention.[7,8] However, injuries have multiple causes and risk factors, and the debate about which omics variables (and under which conditions) are the most important to predict injury risk is still unresolved. Previous studies have independently evaluated the risk associated with training load,[9] GPS,[10] genomics,[11] and metabolomics.[12] However, their models have not always been able to predict injuries accurately since they have been evaluated independently. In addition, the analysis of multifactorial conditions must be gender-specific, as there are substantial biological and injury rate differences between men and women.[13] Unfortunately, the most relevant omics variables for women have not been established yet, as existing models are largely geared toward men or a mixture of both genders.[14]

Therefore, we designed a study to obtain models exclusively tailored for female players, integrating precise measurements of external load, and incorporating the key features from omics data

Cáceres https://orcid.org/0000-0001-8551-6695
Ferrer https://orcid.org/0000-0002-9716-1577
Balagué-Dobón https://orcid.org/0000-0003-3869-9737
Escribà-Montagut https://orcid.org/0000-0003-2888-8948
D. Sarrat-González https://orcid.org/0000-0002-9064-3303
Quintás https://orcid.org/0000-0002-4240-9846
Rodas https://orcid.org/0000-0002-8060-3914
J.R. González (juanr.gonzalez@isglobal.org) is corresponding author, https://orcid.org/0000-0003-3267-2146

(genomics and metabolomics) for effective injury prevention. The main objective of this study was to identify the factors associated with injury risk from the monitoring of a professional women's football team over 2 seasons. While the study emphasizes the development and creation of models for female athletes, as it was trained with female data, it also acknowledges the potential applicability of the model to male athletes, thereby extending its relevance to both genders.

Material and Methods

Participants

We designed a prospective, longitudinal cohort study carried out by Fútbol Club Barcelona (FCB, Barcelona, Spain). We studied 24 elite professional women football players of the first team squad of FCB over 2 seasons, 2019–2020 and 2020–2021 (from September 15, 2019 to June 30, 2021). Usually, the players had around 9 hours of training and 1 or 2 competitive matches per week. We collected clinical and anthropometric data (age, gender, race, body mass index, and fat mass) for each season. The medical team that managed injuries, recorded them during the studied period, and entered them into a validated electronic medical record system (COR, version 2.0; FCB).

Ethics Approval and Informed Consent

The study was conducted according to the guidelines of the Declaration of Helsinki[15] and was approved by the local committee of Barça Innovation Hub and the Ethics Committee of Consell Català de l'Esport (code 012/CEICGC/2021). All participants were informed of the risks and benefits of the study and provided their written consent for genotyping. All personal information and results were anonymized to ensure data confidentiality.

Injury Diagnoses

Injury diagnoses were made by the same medical physician (the team's doctor) with the support of the FCB medical department during the evaluation period. Team doctors followed the same criteria in terms of diagnosis and return-to-play decision, according to the FCB guidelines.[16] Herein, we analyzed noncontact injuries observed in muscle, tendon, ligament, and cartilage.

Quantification of the External Load of Training and Competition by GPS Data

We captured the external workload during training and games by using GPS data collected with the WIMU PROTM device (Realtrack Systems, S.L.).[17] We preprocessed the collected data with the SPRO Software (version 927, Realtrack Systems, S.L.), which compiled data in RAW format and created the final variables. The variables related to volume included: total time (in minutes); total distance (in meters); high metabolic load distance (in meters; defined as 25.5 W/kg – distance covered above 5.5 m/s); decelerations; accelerations; player load (PL; load in arbitrary units); and high-speed running (in meters, speeds above 18 km/h). We also collected variables related to intensity: load (arbitrary units per minute); distance (in minutes); high metabolic load distance per minute; and high-speed running per minute. All these variables are described in detail elsewhere.[17]

We evaluated the quality of the data by performing paired-wise correlation plots. Concretely, we correlated the distance versus PL before and after removing outliers (defined as those values with an absolute value of the residual larger than 3) that were reassigned to the expected value using a linear model.

Genotyping

We complemented the results of a previous GWAS study[18] with a comprehensive literature review to select 108 single nucleotide polymorphisms (SNPs) associated with muscle, tendon, and ligament injuries. The list of SNPs, their genomic position, location within each candidate gene, and functions are described. The genotyping process is described elsewhere.[19] Quality control measures on SNP genotyping included Hardy–Weinberg equilibrium ($P < .01$), missing rate (>5%), and low minor allele frequency (5%). We removed 16 SNPs that failed a quality control test and analyzed 92.

Polygenic Score

The combined influence of the multiple SNPs was calculated using a polygenic score (PS).[20] A genotype score (GS) of 2 was assigned to the genotype "protective" for injuries, a GS of 1 to the heterozygous genotype, and a GS of 0 to the genotype "susceptible" to injuries. We then summed up the GS of each SNP used to create the PS.

Metabolomic Data

We quantified 69 urine metabolites by ultra-performance liquid chromatography–tandem mass spectrometry, as described elsewhere.[21] We conduct a total of 4 assessments per season. Starting in July, followed by 1 each quarter. Data were normalized after removing metabolites with 0 variability (5 metabolites) and a high percentage of missing values (3 metabolites). A total of 61 metabolites were finally chosen for downstream analyses. We measured each metabolite in 5 time periods and performed a linear interpolation of the metabolites.

Statistical Analysis

Model of Risk Injury and Covariables

We tested the association between the risk of injuries and (1) workload, (2) genomic features, and (3) metabolomic characteristics, with a frailty Cox model. The model included a random effect on the player and was able to account for the repeated nature of injuries.[22] The set of covariables for the models was selected from: (1) anthropometric variables, such as age and body mass index, which were registered 3 times per year (July, December, and March) and were introduced in the model as time-dependent variables; (2) number of previous injuries, also modeled as a time-dependent variable; and (3) period of the season, considered as a categorical covariate (July–October, November–February, and March–June). We used the likelihood ratio test to select the statistically significant covariables associated with injury risk.

Injury Risk Estimation Through Frailty Cox Model and Distributed Lag Nonlinear Models

We used the frailty Cox model with covariates to independently assess the association between injury risk and each variable, including training GPS features, genomic variables, and metabolomic characteristics. We modeled fat mass by using a second-order polynomial. We considered genetic polymorphisms and the PS in the model as fixed variables, and metabolites were included using natural splines to allow for nonlinear relationships. We introduced the GPS variables in the model by using the distributed lag nonlinear models (*dlnm*) to account for the cumulative effect of

workload. This model captures the underlying data structure since injuries are modeled as a recurrent event process while capturing the heterogeneity among players through the frailty term. Introducing the effect of workload through *dlnm* allows combining the magnitude of the training load and a function of the distance from day 0 up to lag make 27 (4 wk). This model can capture the nonlinear effects of workload on injury risk by using restricted cubic splines[23] and the lag function with a linear model. We determined statistically significant variables with the likelihood ratio test. We reported effects as hazard ratio (HR) and their 95% CI, except for metabolites that represented the effect (eg, log-HR), for which the risk was provided in log-scale per 1 SD change (we standardized the metabolite data in the preprocessing step). We used the Kaplan–Meier estimator to compare probability curves among genotypes.

We then selected significant risk factors by stepwise elimination to determine those to be included in the best-fitted model.[24] Predictive capacity was measured using C-index (or C-statistic).[25] Overfitting was controlled using the leave-one-out cross-validation procedure.

We performed all statistical analyses using R (version 4.2.2). We computed the cumulative effect of PL with the *crossbasis* function of the *dlnm* R package (version 2.4.7) and fitted frailty Cox models with the *coxph* function of the *survival* R package (version 3.4.0). We computed counting process data to accommodate time-dependent variables with the *exphist* and *survSplit* functions of *dlnm* and *survival* packages, respectively. The complete pipeline describing the reproducible R code is available at our GitHub: https://github.com/isglobal-brge/Supplementary-Material/blob/master/Gonzalez_2023/sportomics_sup_mat.Rmd.

Results

Participant Characteristics and Injuries

We studied 24 female players who were part of the first team of FCB during the seasons 2019–2020 and 2020–2021. Throughout these 2 seasons, several cases of COVID-19 were detected, and all followed the diagnosis, monitoring, and Return to Play protocol established by the club. In 1 case, the Return to Play was extended by 8 weeks because of respiratory symptoms. All players followed the nutritional recommendations of the nutritionist of the FCB medical department. Supplementation was based on isotonic drinks, postworkout and match carbohydrate and protein shakes, and vitamin D and omega-3 supplements.

Table 1 summarizes the main characteristics of the studied subjects. The most frequent type of injuries involved the muscles or the ligaments (Table 2). The injury incidence rate per 1000 player hours was 6.4 for games and 4.9 for training. During the 2019–2020/2020–2021 period, 19 players (79.2%) suffered at least 1 injury. Of these, 14 (73.7%) had 1 or 2 injuries and 3 players (15.8%) had 4 or more injuries (Figure 1).

Table 1 Baseline Characteristics of the Study Participants

Variable	Overall players (N = 24)
Ancestry, n (%)	
European	23 (95.8)
African	1 (4.2)
Position, n (%)	
Defender	8 (33.3)
Forward	7 (29.2)
Midfielder	6 (25.0)
Goalkeeper	3 (12.5)
Age, y	26.5 (17.2–31.9)
Height, cm	1.70 (1.63–1.80)
Weight, kg	63.1 (51.4–76.0)
Body mass index, kg/m^2	21.7 (19.0–25.7)
Fat mass, %	19.1 (14.4–31.1)

Note: Figures are median (and range) unless otherwise stated.

Table 2 Description of Injuries, n (%)

Variable	Overall injuries (N = 40)
Type of injury	
Muscle	25 (62.5)
Tendon	4 (10.0)
Muscle/tendon	2 (5.0)
Ligament	8 (20.0)
Cartilage	1 (2.5)
Side leg	
Left	23 (57.5)
Right	17 (42.5)
Activity	
Practice	26 (65.0)
Game	14 (35.0)
Number of injuries per player	
0	5 (20.8)
1	7 (29.2)
2	7 (29.2)
3	2 (8.3)
4	2 (8.3)
5	1 (4.2)

Associations Between Injury Risk and External Workload Variables (GPS)

To test whether GPS variables were associated with injury risk, the model with the GPS variable was compared with the baseline model that includes fat mass and a random effect for players (likelihood ratio test: $P = .035$). A statistically significant association between injury risk and the cumulative workload effect given by the daily PL feature (external workload) of GPS ($P = .033$) was observed (Figure 2). The HR of having an injury increased with PL intensity and day lag (Figure 2A) and decreased for a given day for any value of PL (effect of acute training improving physical condition; Figure 2B). However, the HR increased if the player kept the intensity over time (chronic load), becoming at risk after accumulating a load of 10 days (Figure 2B). After approximately 3 weeks of keeping a constant PL of 3, the risk became statistically significant (Figure 2C). A PL above 2 provided statistically significant protection against injuries for a given day (Figure 2D).

Associations Between Injury and Genomic Variables

The model comprising the random effect of the individual, the fat mass, and the cumulative effect of PL (adjusted) was considered the

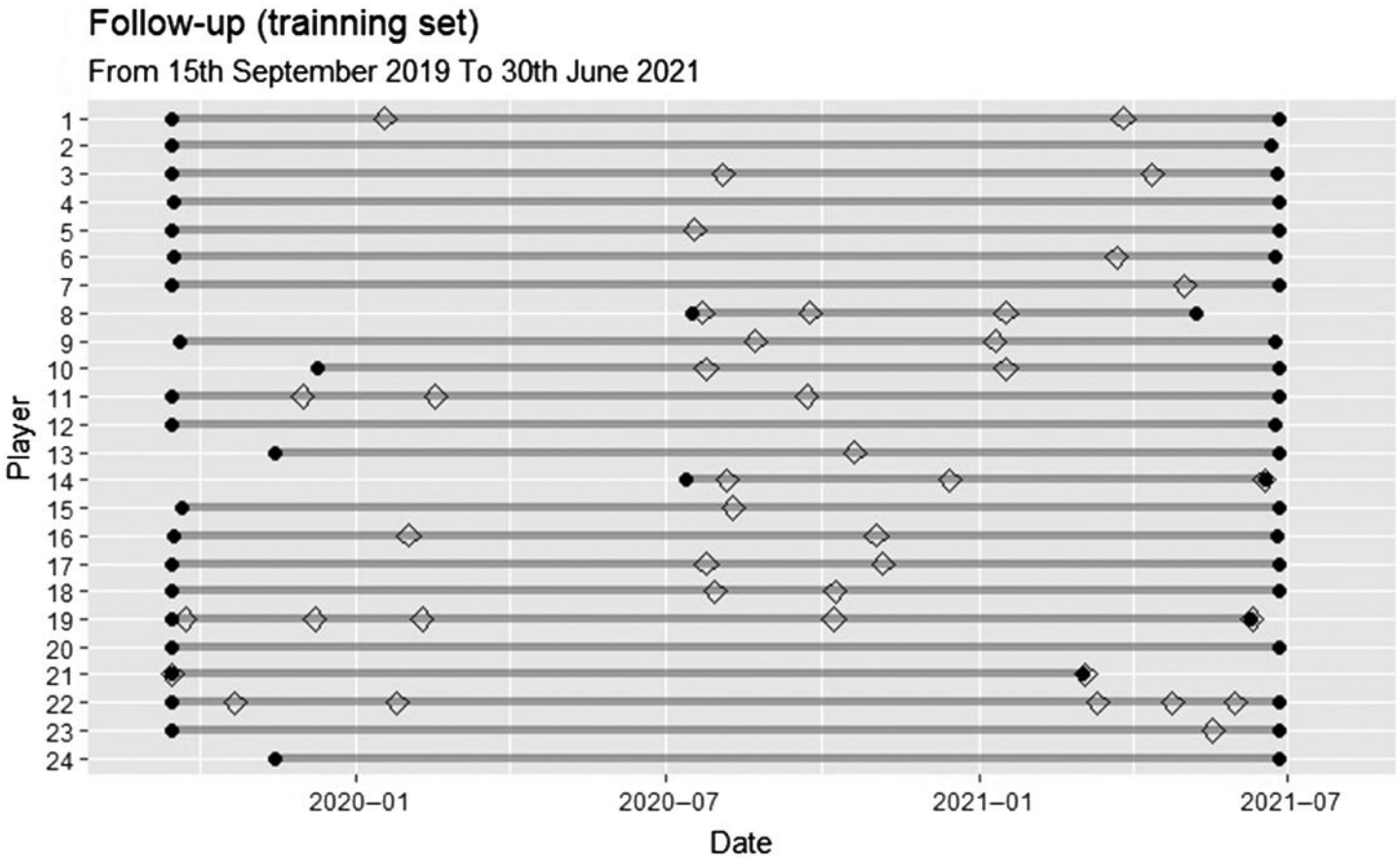

Figure 1 — Injury events over the observation period. Injury events (diamonds) and the follow-up period (gray line) are represented for each of the athletes during seasons 2019–20 and 2020–21.

reference model. The SNPs' major allele frequency mean was 72.6%, ranging from 50.0% to 93.8%. Six SNPs were significantly associated with injury risk at a 5% level (Table 3). Overall, a protective effect (ie, risk for the normal allele) was observed for the rare alleles of all polymorphisms (Table 3). For instance, the presence of the *MMP1* rs1799750 rare variant (denoted by −) was found to have a 2-fold decreased risk per allele (adjusted HR of 0.46; 95% CI, 0.29–0.82). The Kaplan–Meier curve showed that only one out of 4 individuals with the rare allele (−) of the variant rs1799750 suffered an injury during the observed period, while, 2 out of 3 individuals with the normal genotype (G) had an injury ($P = .008$; Figure 3).

A genetic score was established using the 6 aforementioned polymorphisms. This score allowed us to classify individuals into 3 groups based on the aggregate number of protective alleles (0–3, 4–5, and ≥ 6; Figure 4A). The number of injuries was lower in the category with the highest score. This score also exhibited substantial differences in the probability of having an injury ($P < .001$; Figure 4B).

Associations Between Injury Risk and Metabolomic Variables

This analysis was performed on 19 individuals tested at 4 timepoints during the 2 seasons. Our analyses revealed 3 metabolites significantly associated with injury risk after adjusting for the PL. Beta-alanine showed an inverse association with injury risk ($P = .045$), while serotonin ($P = .025$) and 5-hydroxy-tryptophan (5-HTF; $P = .037$) were directly associated with injury risk (Figure 4C).

Multivariate Model: Workload, Genomic, and Metabolomic Variables

A multivariate model was built by sequentially including all the variables significantly associated with injury risk (Figure 4D). The baseline model considering PL had a cross-validated C-index of 65.7. The inclusion of the 3 metabolites yielded a cross-validated C-index of 67.4 and showed a statistically significant improvement in performance. Even more strikingly, adding the genetic score to the PL baseline model (ie, genetic model) improved the cross-validated C-index to 74.4. Finally, the highest predictive capacity was reached with the full model, including both metabolites and genetic score to the baseline PL model (C-index = 75.7), although the difference in C-index between the full and the PL-genetic models was not significant.

Discussion

In our study, we observed a significant association between injury risk and cumulative workload in women football players. In addition, 6 SNPs and 3 metabolites were found to be associated with injury risk. Finally, our multivariate model, including workload, genomic variables, and metabolomic characteristics, provided a good predictive capacity of injury risk. Our model is based on the external workload accumulation measured by GPS technology. In agreement with our results, the accumulation of perceived workload across training sessions has been shown to be a strong predictor of injuries.[26] However, our assessment of load is

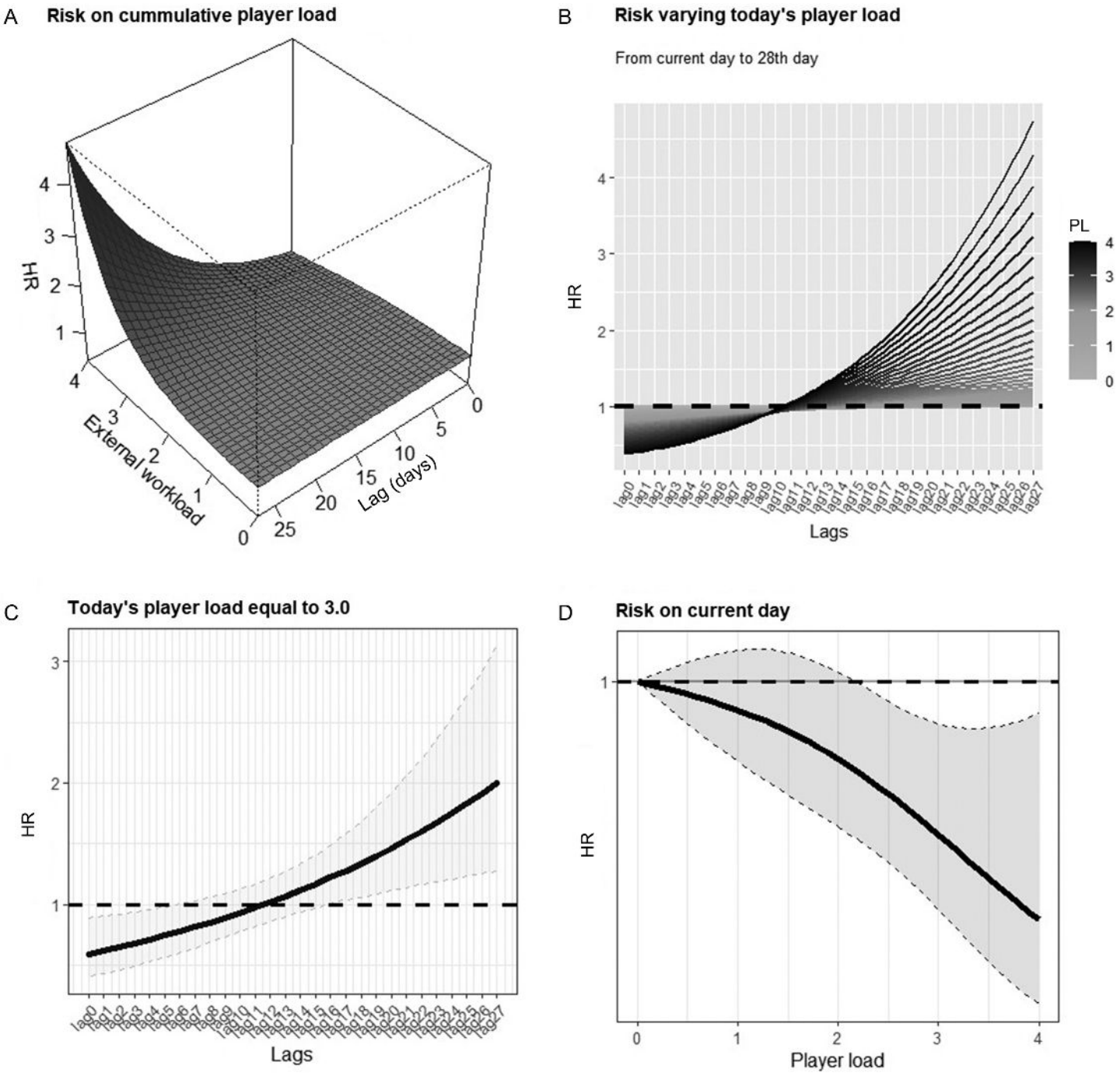

Figure 2 — Fitted function of accumulated PL, measured by Global Positioning System, on the risk of injuries for female elite football players. (A) Risk of injury as a function of PL levels and its lag, from 0 to 28 backward in time (cumulative effect of PL). (B) Given risk as a function of PL for a given day (eg, lag 0). (C) Projection of the risk function for different PLs as a function of the lags. (D) Given risk as a function of lags for PL = 2. HR indicates hazard ratio; PL, player load.

objective, so it lacks the internal state of fatigue that may be captured by a subjective measure. To compensate, our injury risk model contemplates the internal state of the athletes by assessing their omics features.

Regarding genetic susceptibility to injury, statistically significant associations were found with rs516115, rs162502, and rs4903399 within genes *DCN,ADAMTS5*, and *ESRRB,* previously associated with injury risk.[27] In particular, the GG genotype of rs516115 has been reported to have a protective effect against injuries in the anterior cruciate ligament (ACL). Also, the *VEGFA* rs699947 and the A allele have been found to protect against ACL injuries[28] and tendinopathies.[29] In addition, the variant rs4903399 in *ESRRB* has been reported as a risk factor for rotator cuff disease.[30] Finally, the variant rs1799750 in the *MMP1* gene has been described as a risk factor for noncontact muscle injuries since this gene is involved in maintaining and remodeling the connective tissue extracellular matrix surrounding the muscle cells and spindles.[27] Moreover, the PS described here could help

Table 3 Hazard Ratio of Injury Risk

SNP	n (%)[a]	Number of injuries, %[a]	Hazard ratio	95% CI	*P*
rs1799750			0.46	(0.29–0.82)	.0078
G/G	5 (21)	12 (30)			
G/–	14 (58)	26 (65)			
–/–	5 (21)	2 (5)			
rs699947			0.49	(0.25–0.94)	.0313
C/C	8 (33)	19 (48)			
C/A	9 (38)	12 (30)			
A/A	7 (29)	9 (22)			
rs9406328			0.46	(0.23–0.94)	.0337
G/G	8 (33)	25 (62)			
G/A	15 (63)	15 (38)			
A/A	1 (4)	0 (0)			
rs162502			0.38	(0.15–0.95)	.0377
G/G	18 (75)	37 (92)			
G/A	4 (17)	2 (5)			
A/A	2 (8)	1 (3)			
rs4903399			0.34	(0.14–0.81)	.0063
C/C	14 (58)	27 (68)			
T/C	8 (33)	12 (30)			
T/T	2 (8)	1 (3)			
rs516115			0.64	(0.09–0.68)	.0235
AA	11 (46)	27 (68)			
AG	8 (33)	7 (18)			
GG	5 (21)	6 (15)			

Abbreviation: SNP, single-nucleotide polymorphism. Adjusted effects for PL (from Global Positioning System) of statistically significant SNPs at 5% level. The models were adjusted for fat mass and individual as random effects. Hazard ratio provides the injury risk for the additive model. In the SNP column, the rare allele is indicated afrer "/".
[a]Total percentage might not add up to 100 because of rounding.

stratify women against injuries to be eligible for personalized training plans and exercise recommendations. In a previous study on the same athletes, we observed an increased risk of ACL injuries for variants in COL5A1 in relation to males.[19] In the present work, we study the effects of these variants, accounting for the cumulative workload in females only. Further work is needed to determine the impact of workload on sexual dimorphism in the genetic susceptibility of injury risk, emphasizing the differences in training plans between sexes.

In relation to metabolomic data, the most significant associations with injury risk were found for beta-alanine, serotonin, and 5-HTP. In agreement with ours, another study has suggested that a dietary supplement of beta-alanine may help reduce the risk of injury by improving muscle endurance and reducing muscle fatigue.[31] Also, beta-alanine supplementation can increase the concentration of carnosine in muscles, which can help buffer the accumulation of hydrogen ions during high-intensity exercise. This can delay the onset of muscle fatigue and improve muscle endurance, which may also help reduce the risk of injury.[31] Finally, in agreement with our results, high levels of 5-HTP, the precursor of serotonin, have been associated with an increased risk of injuries since it could increase the production of serotonin in the brain.[32] Moreover, high levels of 5-HTP can cause drowsiness and fatigue, reducing alertness and reaction time.[33]

In our study, the most accurate injury risk predictions were achieved with the multivariate model incorporating workload measures, various polymorphisms, and metabolites. However, according to the C-index, external workload and genetic predisposition were the most significant factors influencing injury risk. This finding has significant practical implications since collecting this information is relatively inexpensive and noninvasive, in comparison to obtaining metabolomic data. Additionally, workload can be routinely monitored using electronic devices, such as GPS, whereas genetic predisposition only needs to be measured once.

The main limitation of our study, and in general in this field of predictive models, is the small sample size. Nonetheless, our longitudinal data improved the statistical power of studies with small sample sizes, by enabling the examination of changes in the same individuals over time, which increased the amount of information that could be obtained from each participant (we analyzed more than 14,500 time points). This reduced the variability of the data and enabled a more accurate estimation of the effects of exposures being studied. Additionally, longitudinal data could help to account for individual differences and to identify patterns of change that may not be apparent in cross-sectional analyses. We highly recommend large collaborative projects that could increase the sample size and evaluate external validity. Future studies should consider incorporating information on menstrual status, oral contraceptive use, and hormone levels, adhering to established guidelines for research standards in women.

To address the multiple comparisons problem in omic studies, we selected features based on previous literature evidence of their relation to injury risk. This approach is suitable for targeted studies with

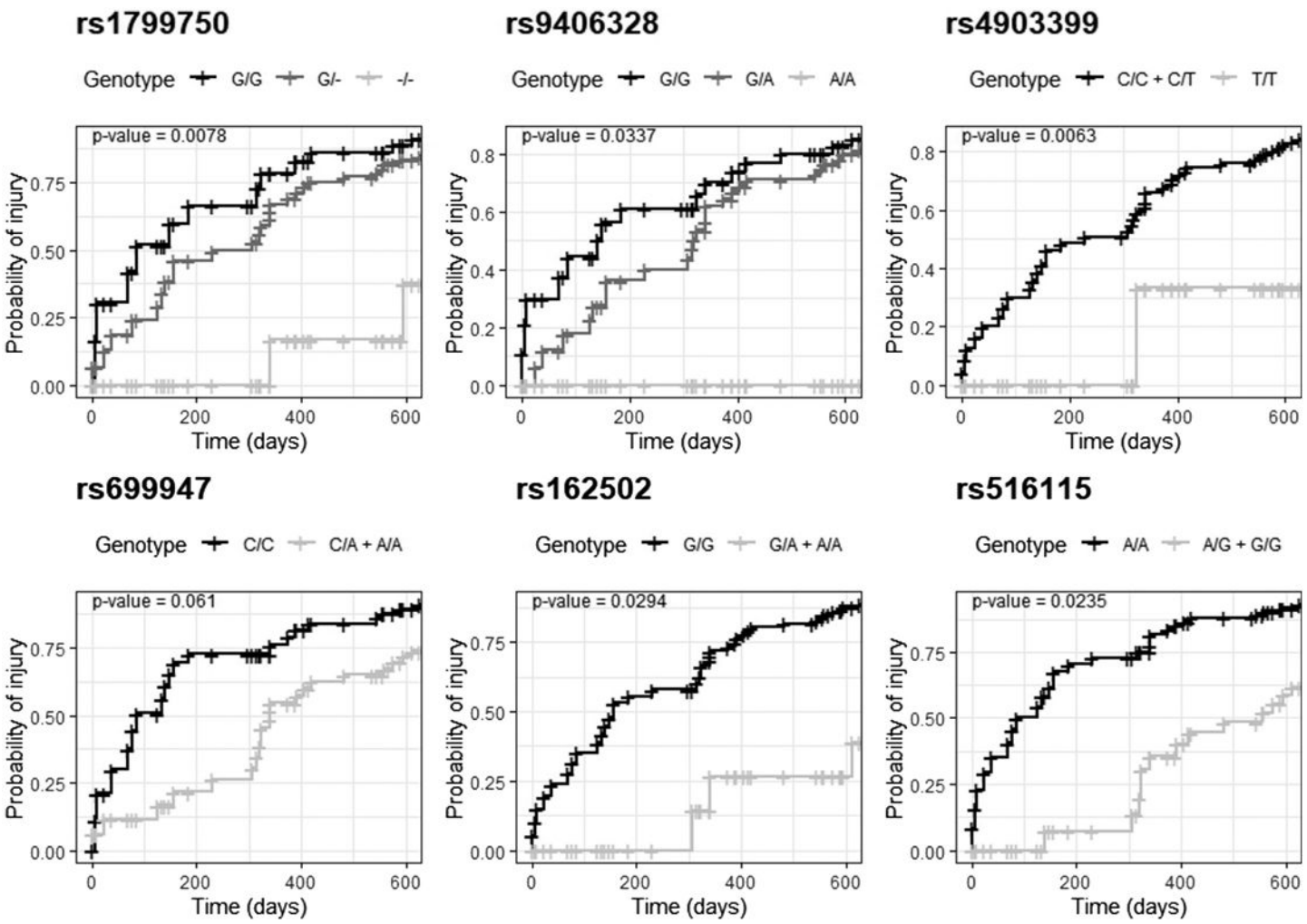

Figure 3 — Probability of injury estimated using Kaplan–Meier curves for the single-nucleotide polymorphisms that were significantly associated with injury in the frailty Cox models adjusted for fat mass and external workload. Some genotypes were collapsed (dominant model) for the single-nucleotide polymorphisms with low number of individuals (figures with 2 curves).

specific hypotheses or limited tests, such as our study.[34] Our goal is to identify relevant factors associated with injury risk, which can be facilitated by statistical models but may have limited predictive power. To improve prediction, we recommend using machine learning algorithms with large data sets and demonstrated omic data.

Practical Applications

Research in the field of sportomics (ie, the conjunction of sport with any omic layer, such as genomics and metabolomics) and, specifically, its use to predict sports injuries holds great potential for injury prevention for athletes at every level. To our knowledge, this is the first attempt to predict noncontact injury risk based on GPS, genomic, and metabolomic data using longitudinal data from female elite players who have been largely unstudied. Our paper also uses state-of-the-art data modeling approaches to properly deal with the complex nature of injury data and the cumulative effect of external workload. The integration of the proposed model in a larger cohort of players and its prospective evaluation will likely help to improve the injury risk prediction, ultimately benefitting injury management and prevention (the latter including in asymptomatic athletes).

Conclusions

This study demonstrates, for the first time, that injury risk prediction combining omics technologies with external workload measurements was superior to using external workload measurements alone. We present a comprehensive model that integrates workload, genomic variables, and metabolomic characteristics, demonstrating a promising predictive capacity for injury risk in elite female football players. However, we acknowledge that our findings represent only the initial step toward practical applications in real-world settings. Also, studies revolving around female players could also include other biological variables playing a role in the specific injury risk of females (eg, menstrual cycle).

By candidly discussing these challenges in our manuscript, we aim to encourage the scientific community to engage in ongoing dialogue and collaboration. This collaborative approach is essential for refining our injury prediction model and ultimately implementing it successfully in real-world sports environments. Through continued research and validation, we aspire to contribute to the advancement of injury prevention strategies, enhancing the well-being and performance of athletes across the globe.

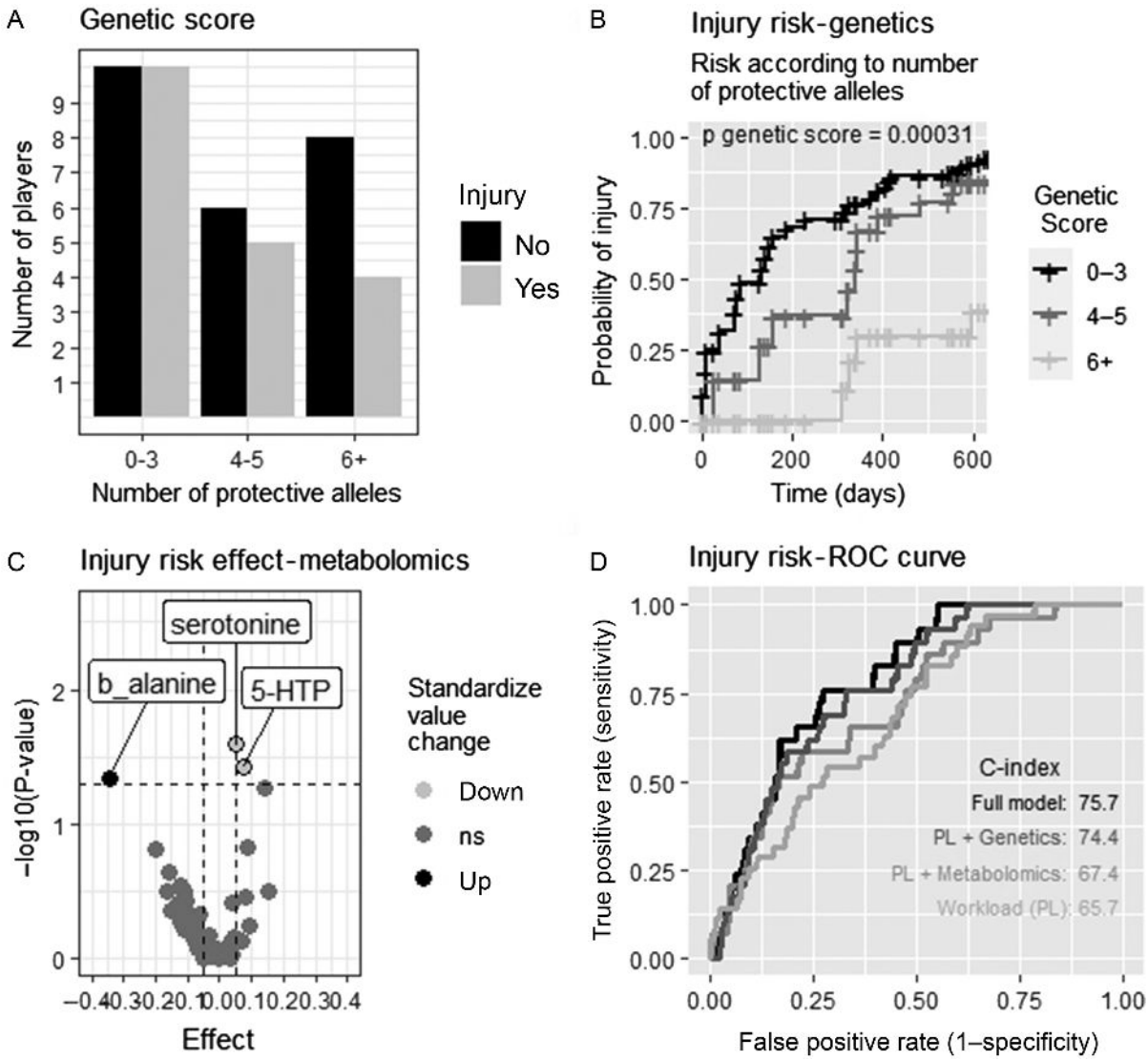

Figure 4 — Multivariate modeling. (A) Number of players having injuries or not according to the genetic score build counting the number of protective alleles (0–3, 4–5, ≥6). (B) Probability of injury over the 2 seasons split by the genetic scores. (C) Volcano plot showing the metabolites that were significantly associated with injury risk ($P < .05$) after adjusting for the baseline model (body fat and random effect of player) and player load. (D) Cross-validated C-index (ie, area under the ROC curve for survival data) describing the predictive capacity of the model including Global-Positioning-System and genomic data. 5-HTP indicates 5-hydroxytryptophan; ROC, receiver operating characteristic; PL, player load.

Acknowledgments

The first author would like to express gratitude toward Xavier Basagaña and Lena Kristin Bache-Mathiesen for their invaluable insights and discussions pertaining to the implementation of *dlnm* models in the current project. This project has received funding from the Ministry of Research and Universities, Government of Catalonia, under reference 2021 LLAV 00027—Operational Personal Sport-Injury WArning System—PSIWAS and AGAUR 2023 PROD 00020. The publication is also part of the research project (EXP_74937) in science and technology applied to physical activity beneficial for health (AFBS) and sports medicine, funded by the European funds from the Recovery, Transformation, and Resilience Plan, financed by the Ministry of Culture and Sport and the European Union's "NextGenerationEU"/PRTR. **Author Contributions:** Rodas designed the initial plan and conducted the study together with Ferrer. Gonzalez led statistical analyses and data modeling. Escribà-Montagut and Sarrat created R functions and packages to analyze GPS data. Balagué-Dobón performed genetic quality control and data analyses. Quintás generated metabolomic data and performed quality control and normalization analyses. Gonzalez and Cáceres wrote the first draft of the manuscript. All other authors were involved in interpreting findings and provided meaningful feedback in refining the final manuscript. Gonzalez and Rodas co-supervised the project. All authors have read and approved the final manuscript. **Competing Interests:** Ferrer and Rodas are employed by Barça Innovation Hub.

References

1. Eliakim E, Morgulev E, Lidor R, Meckel Y. Estimation of injury costs: financial damage of English premier league teams' underachievement due to injuries. *BMJ Open Sport Exerc Med*. 2020;6(1): e000675. doi:10.1136/BMJSEM-2019-000675
2. Schooling CM, Jones HE. Clarifying questions about "risk factors": predictors versus explanation. *Emerg Themes Epidemiol*. 2018;15(1): 80. doi:10.1186/S12982-018-0080-Z

3. Drew MK, Finch CF. The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports Med*. 2016;46(6):861–883. doi:10.1007/S40279-015-0459-8
4. Miranda-Comas G, Zaman N, Ramin J, Gluck A. The role of workload management in injury prevention, athletic, and academic performance in the student-athlete. *Curr Sports Med Rep*. 2022;21(6):186–191. doi:10.1249/JSR.0000000000000964
5. Khoramipour K, Sandbakk Ø, Keshteli AH, Gaeini AA, Wishart DS, Chamari K. Metabolomics in exercise and sports: a systematic review. *Sports Med*. 2022;52(3):547–583. doi:10.1007/S40279-021-01582-Y
6. Sellami M, Elrayess MA, Puce L, Bragazzi NL. Molecular big data in sports sciences: state-of-art and future prospects of OMICS-based sports sciences. *Front Mol Biosci*. 2022;8:410. doi:10.3389/FMOLB.2021.815410
7. Bongiovanni T, Lacome M, Fanos V, Martera G, Cione E, Cannataro R. Metabolomics in team-sport athletes: current knowledge, challenges, and future perspectives. *Proteomes*. 2022;10(3):27. doi:10.3390/PROTEOMES10030027
8. Maffulli N, Margiotti K, Longo UG, Loppini M, Fazio VM, Denaro V. The genetics of sports injuries and athletic performance. *Muscles Ligaments Tendons J*. 2013;3(3):173–189. doi:10.11138/mltj/2013.3.3.179
9. Verstappen S, van Rijn RM, Cost R, Stubbe JH. The association between training load and injury risk in elite youth soccer players: a systematic review and best evidence synthesis. *Sport Med-Open*. 2021;7(1):296. doi:10.1186/S40798-020-00296-1
10. Kupperman N, Hertel J. Global positioning system-derived workload metrics and injury risk in team-based field sports: a systematic review. *J Athl Train*. 2020;55(9):931–943. doi:10.4085/1062-6050-473-19
11. Maestro A, Del Coso J, Aguilar-Navarro M, et al. Genetic profile in genes associated with muscle injuries and injury etiology in professional soccer players. *Front Genet*. 2022;13:1035899. doi:10.3389/FGENE.2022.1035899
12. Quintas G, Reche X, Sanjuan-Herráez JD, et al. Urine metabolomic analysis for monitoring internal load in professional football players. *Metabolomics*. 2020;16(45):668. doi:10.1007/S11306-020-01668-0
13. Zech A, Hollander K, Junge A, et al. Sex differences in injury rates in team-sport athletes: a systematic review and meta-regression analysis. *J Sport Health Sci*. 2022;11(1):104–114. doi:10.1016/J.JSHS.2021.04.003
14. Winter SC, Gordon S, Brice SM, Lindsay D, Barrs S. A multifactorial approach to overuse running injuries: a 1-year prospective study. *Sports Health*. 2020;12(3):296–303. doi:10.1177/1941738119888504
15. WMA Declaration of Helsinki. Ethical principles for medical research involving human subjects—WMA—The World Medical Association. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. Accessed March 6, 2023.
16. Barça Innovation Hub. *Muscle Injury Guide—Preventing and Treating Muscle Injuries in Football*.
17. Guitart M, Casals M, Casamichana D, et al. Use of GPS to measure external load and estimate the incidence of muscle injuries in men's football: a novel descriptive study. *PLoS One*. 2022;17(2):e0263494. doi:10.1371/JOURNAL.PONE.0263494
18. Rodas G, Osaba L, Arteta D, Pruna R, Fernández D, Lucia A. Genomic prediction of tendinopathy risk in elite team sports. *Int J Sports Physiol Perform*. 2019;15(4):489–495. doi:10.1123/IJSPP.2019-0431
19. Rodas G, Cáceres A, Ferrer E, et al. Sex differences in the association between risk of anterior cruciate ligament rupture and COL5A1 polymorphisms in elite footballers. *Genes*. 2022;14(1):33. doi:10.3390/GENES14010033
20. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. *J Physiol*. 2008;586(1):113–121. doi:10.1113/JPHYSIOL.2007.141887
21. Rodas G, Ferrer E, Reche X, Sanjuan-Herráez JD, McCall A, Quintás G. A targeted metabolic analysis of football players and its association to player load: comparison between women and men profiles. *Front Physiol*. 2022;13:608. doi:10.3389/FPHYS.2022.923608
22. Therneau TM, Grambsch PM, Pankratz VS. Penalized survival models and frailty. *J Comput Graph Stat*. 2012;12(1):156–175. doi:10.1198/1061860031365
23. Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, Clarsen B, Fagerland MW. Not straightforward: modelling non-linearity in training load and injury research. *BMJ Open Sport Exerc Med*. 2021;7(3):e001119. doi:10.1136/BMJSEM-2021-001119
24. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health*. 2020;8(1):e000262. doi:10.1136/FMCH-2019-000262
25. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–1117. doi:10.1002/SIM.4154
26. Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, Clarsen B, Fagerland MW. Assessing the cumulative effect of long-term training load on the risk of injury in team sports. *BMJ Open Sport Exerc Med*. 2022;8:e001342. doi:10.1136/BMJSEM-2022-001342
27. Lim T, Santiago C, Pareja-Galeano H, et al. Genetic variations associated with non-contact muscle injuries in sport: a systematic review. *Scand J Med Sci Sports*. 2021;31(11):2014–2032. doi:10.1111/SMS.14020
28. Feldmann D. Whole genome sequencing approach to identifying genetic risk factors underlying anterior cruciate ligament injuries in a twin family study. 2022. https://hdl.handle.net/11427/36615
29. Salles JI, Duarte MEL, Guimarães JM, et al. Vascular endothelial growth factor receptor-2 polymorphisms have protective effect against the development of tendinopathy in volleyball athletes. *PLoS One*. 2016;11(12):717. doi:10.1371/JOURNAL.PONE.0167717
30. Motta GR, Amaral MV, Rezende E, et al. Evidence of genetic variations associated with rotator cuff disease. *J Shoulder Elb Surg*. 2014;23(2):227–235. doi:10.1016/J.JSE.2013.07.053
31. Furst T, Massaro A, Miller C, Williams BT, LaMacchia ZM, Horvath PJ. β-Alanine supplementation increased physical performance and improved executive function following endurance exercise in middle aged individuals. *J Int Soc Sports Nutr*. 2018;15(1):32. doi:10.1186/S12970-018-0238-7
32. Dalley JW, Roiser JP. Dopamine, serotonin and impulsivity. *Neuroscience*. 2012;215:42–58. doi:10.1016/J.NEUROSCIENCE.2012.03.065
33. Weinberg-Wolf HB, Fagan N, Monte OD, Chang SWC. Increasing central serotonin with 5-hydroxytryptophan disrupts the inhibition of social gaze in nonhuman primates. *J Neurosci*. 2022;42(4):670–681. doi:10.1523/JNEUROSCI.0413-21.2021
34. Daly AK, Day CP. Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol*. 2001;52(5):489–499. doi:10.1046/J.0306-5251.2001.01510.X